Structure and Dynamics of the Ovarian Tumour Domain Protease from the Crimean-Congo Hemorrhagic Fever Virus by Nuclear Magnetic Resonance Spectroscopy

by

Sagar Saran

A Thesis submitted to the University of Manitoba and the Faculty of Graduate Studies in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE (M.Sc.)

Department of Chemistry

The University of Manitoba

Winnipeg, Canada

Abstract

Crimean-Congo Hemorrhagic Fever Virus (CCHFV) is endemic to more than 30 countries and shows fatality rates among humans ranging from 30-50%. The CCHFV 169-residue L-segment deubiquitinase (DUB) is a member of the superfamily of Ovarian Tumour (OTU) ubiquitin thiolesterases that interfere with innate immune responses and hence is an attractive antiviral target. I report here the application of Nuclear Magnetic Resonance spectroscopy to probe the role of CCHFV OTU enzyme dynamics in the catalytic mechanism of the enzyme. $^{13}C/^{15}N$ triple-resonance experiments and an amino acid "*unlabelling*" scheme were used for backbone resonances assignments. NMR chemical shift analysis, NMR spin-relaxation experiments at two magnetic fields, *Lipari-Szabo* Model-free formalism, reduced spectral density mapping and *Carr-Purcell-Meiboom-Gill* Relaxation Dispersion experiments were done to obtain structure and dynamics data. The dynamics data suggested an unfolded C-terminus and a well-packed protein core. Relaxation dispersion measurements show that a significant number of residues undergo conformational exchange on the millisecond timescale. Some of these are near the active site and neighbouring segments and may represent a rate-limiting event along the proteolytic kinetic pathway. They may also play a role in determining the enzyme's broad substrate specificity.

Acknowledgments

Firstly, I would like to sincerely thank Dr. Joe O'Neil for giving me an immense opportunity to work on protein nuclear magnetic resonance spectroscopy research at the University of Manitoba. Dr. O'Neil has always been very supportive and kind throughout my M.Sc. degree training at the University of Manitoba. Dr. O'Neil was not only my research advisor, but mentor in the real-sense. Dr. O'Neil was always available to answer research questions and I can never recollect that I had to ever schedule an appointment with him. Dr. O'Neil always played pivotal role in helping me smoothly transition into the North American culture. I have not only learnt research skills, but also understood deep philosophy of life from his saying, "Life is a journey, not a destination otherwise it would be nirvana". I will always be fervently grateful to him for his unparalleled support to me in Canada and will follow his inspirational advice in my entire future endeavors.

Secondly, I would sincerely thank Dr. Brian Mark (Department of Microbiology at the University of Manitoba) for providing me with a opportunity to work closely with his group for purifying my protease. I also enjoyed working with Dr. Nigel Mcleish who initially mentored me in understanding the basic steps in the protein purification.

Undoubtedly after my advisor, the second most important person who's direct unparalleled contribution to my project has been Dr. Baldeep Khare. Thank you for supporting me through

thick and thin during difficult times in my research and making me realizing my potential to the fullest.

I would also thank my M.Sc. thesis committee- Dr. Pavil Dibrov and Dr. Sean McKenna for their valuable time in reviewing and providing suggestions for my thesis.

I am also highly obliged and thankful to Vu To for acquiring and processing all my NMR data. I will always remember how patient and calm temperament you had even when answering my most trivial questions about NMR. I hope that we will continue to be friends and hopefully work together again. I would like to thank Marikka Lee who closely worked with me during summer term to purify proteins and I was really fortunate to have worked with you. I would am also thankful to Mary Hernando for teaching me how to be organized at work place, sharing jokes, and making my stay enjoyable.

I would sincerely thank all my friends from the chemistry department and outside for their constant support, always being there and never making me feel "home-sick"- Mona (for always being caring and concerned about my well-being), Amit (for all the funny jokes and "OH YEAH", Rajarshi (for always accompanying me to shopping and watching movies at Malls, Steven (for introducing to Winnipeg culture, becoming my first friend from Winnipeg and making every moments forever to cherish in my life), Aditya (for sharing most memorable moments of cooking foods for me and jokes), Brijesh (your amiable and supportive nature has re-defined how a friend should be), Pavan (for always cracking funny jokes and always making

great efforts for friends reunion parties) , Ewan (for your sensitive, thoughtful, understanding, intelligent, yet witty nature that used to lighten my day) ,Sowmya (for matured advice in helping me set up in the lab), Balakishan (for cooking great meals), Gautam (for just being funny at times), Emmanuel (for discussing various topics that revolves our scientific life), Siddharat (for bringing new joy and laughter through your innocent acts), Inderjeet singh (uncle) (for making me feel home away from home).

In conclusion, I would have not been to achieve any success without unreciprocated constant love and impeccable support from my parents- dad, mother, sisters, and all my closed relatives for their constant support and encouragement.

Thank you god for making me believe in me!!!!

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ATP - Adenosine Tri-phosphate

AIC - Akaike information criterion

BAP1 - BRCA1 associated protein-1

BRCA1 - Breast Cancer 1

CREB - cAMP response element binding protein

CIN85 - Cbl-interacting protein 85

COSY - Correlation spectroscopy

CPMG - Carr-Purcell Meiboom Gill

CCHFV - Crimean-Congo Hemorrhagic Fever Virus

CSA - Chemical Shift Anisotropy

CYLD - Cylindromatosis (turban tumor syndrome)

DUB - Deubiquitylating

2D - 2-Dimensional

3D - 3-Dimensional

DTT- Dithiothreitol

E. coli - Escherichia Coli.

FPLC - Fast Protein Liquid Chromatography

g/L - grams/ Litre

g - Gram

GST - Glutathione-S-Transferase

G-proteins - Guanine nucleotide-binding proteins

HMQC - Heteronuclear multiple quantum coherence

Hrs - Hours

HRV3c - Human rhinovirus 3C

HSQC - Heteronuclear single quantum coherence

Hz - Hertz

IFN - Interferon

ISG15 - Interferon Stimulated Gene product-15

IPTG - Isopropyl β-D-1-thiolgalactopyranoside

Kb - Kilobase

kDa - KiloDaltons

KIX - Kinase-inducible domain

LS - *Lipari-Szabo*

LB - Luria broth

MBP - Maltose Binding Protein

MUP-I - Major urinary protein-I

MHz - Megahertz

mM - milliMolar

µL- Microlitre

µs - Microsecond.

ms - Millisecond

mL - Milliliters

mAU - Milli Absorbance Units

min - Minutes

M - Molar

NADPH - Nicotinamide adenine dinucleotide hydrogen phosphate

NADP - Nicotinamide adenine dinucleotide phosphate

Nedd8 - Neural precursor cell expressed, developmental down-regulated 8

nm - Nanometer

NOE - Nuclear Overhauser effect

ns - nanosecond

NMR - Nuclear Magnetic Resonance

ORF - Open Reading Frame

OTU protease - Ovarian Tumor Domain protease

pKID - Phosphorylated kinase-inducible-domain

ppm - Parts per million

ps - Picosecond

R_{ex} - Exchange rate

RD - Relaxation Dispersion

r.p.m. - Revolutions per minute

R_f – R factors

S. cerevisiae - *Saccharomyces cerevisiae*

SH3 – SRC Homology 3 Domain

s - Second

S^2 - Order parameters

ssRNA- Single stranded Ribonucleic acid

SW- Sweep Width

S75 - Superdex 75

Tim23 - Translocase of the inner membrane-23

TOCSY - Total correlation spectroscopy

T_1 - Longitudinal relaxation time

T_2 - Transverse relaxation tie

T - Temperature.

UCH-L1 - Ubiquitin carboxy-terminal hydrolase L1

UCH37 - Ubiquitin carboxy-terminal hydrolase 37

Ub- Ubiquitin

1 INTRODUCTION

1.1 Crimean-Congo Hemorrhagic Fever Virus (CCHFV)

The Crimean Congo Hemorrhagic Fever Virus causes fatal hemorrhagic fevers in humans predominantly spread across Africa, Asia, and the Middle East and Indian sub-continent with fatality rates between 30-50%[1]. The CCHFV infection is caused by a tick-borne virus (*Nairovirus* - family *Bunyaviridae*) and is strongly correlated to the geographical distribution of a tick *Hyalomma*[1] (Figure 1). Although the disease is asymptomatic in animals it is highly fatal in humans beings characterized by non-specific febrile symptoms. The tick is the main vector responsible for transmitting CCHFV diseases, along with percutaneous or mucosal exposure between humans. Some of the current preventive measures include prophylactic therapies along with regular infection-control strategies to curb spreading of this disease. The application of chemical pesticides are also undertaken to prevent spreading of the tick[1].

Figure 1. Hyalomma Tick[2].

1.2 The genome structure of the CCHFV

The genetic diversity of CCHFV pathogenesis and its potential use as a bioterrorism agent made it imperative to characterize the complete genome sequence of the virus[3]. The CCHFV viruses are negatively single-stranded RNA viruses (90-100 nm) and contain three genome segments (L, M, and S)[4]. The electron microscopy of extracted vRNA from the phlebovirus *Uukuniemi* showed non covalent closed circular RNA structures[5]. The characterization of 16 CCHFV genomes indicates a high degree of evolution and diversity because of polymerase error rates[3]. Many studies have reported massive rearrangements in the CCHFV genome around the M and S segments. Out of the 5 published CCHFV strain sequences, three of them are partial "S" segment sequences and the remaining two are "M" and "L" segments, respectively[3].

The first complete genome sequence of the *Kosova Hoti* virus sequence was isolated from a CCHFV-infected female patient from the Balkan region[3]. The genome size of the CCHFV was found to be 19.2 Kb in length consisting of L, M and S segments[3]. The Open Reading Frame (ORF) of the S (1672 nucleotides), M (5364 nucleotides) and L (12150 nucleotides) segments along with the tetra peptides RSKR251, RKLL523, and RKPL1043 were identified in *Kosova Hoti*. Figure 2 shows the cleavage sites for the GP38, Gn and Gc proteins in the M polyprotein. The *Kosova Hoti* RKLL523 tetra peptide is identical for all the strains within group V (Europe/Turkey). It slightly differs from the RKLL523 tetrapeptide in all the sequences of CCHFV and these tetrapeptides correspond to large cleavage recognition site for subtilase SKI-1[3]. Figure 2 also shows the presence of five transmembrane helices for the *Kosova Hoti* M segment polyprotein as black vertical bars. Figure 3 shows the organization of the *Kosova Hoti*

L-segment polyprotein containing an OTU-like protease domain. The identified sequences highlighted in colour are similar in all the L segments of the CCHFV strain[6].

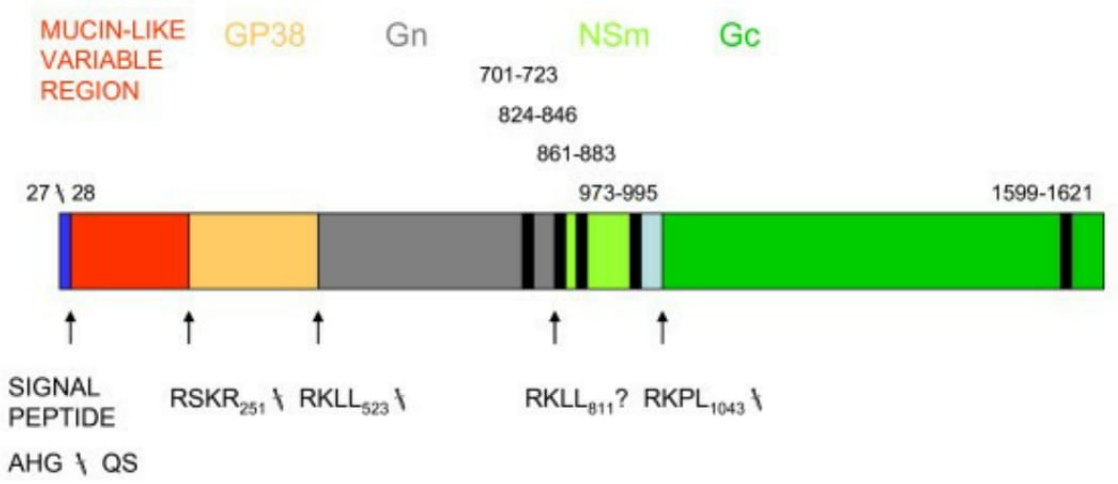


Figure 2. The structure of the M polyprotein of CCHFV *Kosova Hoti* strain[3]. The figure is from reference 3, Darja *et al*., 2008. Illustrated is the mucin-like variable region (red strip) containing the AHG/ QS tetrapeptide region between amino acids 27 and 28, the RSKR251/RKLL523 containing the cleavage site for the transmembrane glycoprotein-GP38 protein (yellow strip), the envelope glycoprotein- Gn (grey strip), the accessory protein- NSm (light green strip), and the envelope glycoprotein-Gc (dark green strip) at the extreme right.

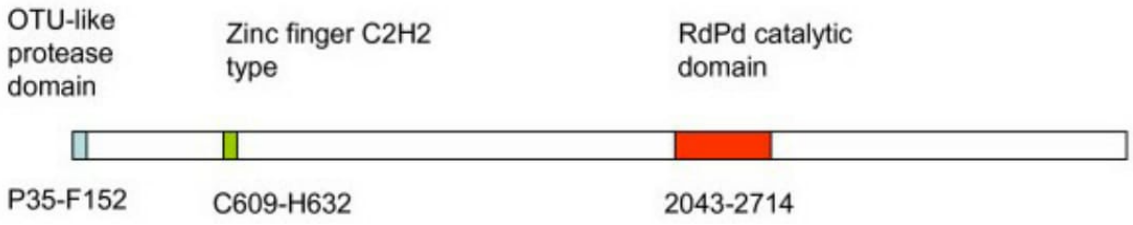


Figure 3. The structure of the L protein of the CCHFV *Kosova Hoti* strain[6]. The figure is from reference 6, Deyde *et al*., 2006. The figure shows the L protein encoded by the L segment of the *Kosova Hoti strain* showing the OTU-like protease (blue strip) from amino acids 35-152, a zinc finger C2H2 type (green strip) domain between residues 609-632, and an RNA- dependent RNA polymerase showing the catalytic domain (red) between amino acids 2043-2714.

1.3 CCHFV- Pathogenesis and Current Treatment

Disease manifestation in humans occurs in four distinct stages: namely the incubation period, the pre-hemorrhagic phase, the hemorrhagic phase, and eventually the final convalescent stage[7]. The incubation phase lasts from 3-7 days and is immediately followed by pre-hemorrhagic phase for 4-5 days. During this second phase, patients suffer from high fever, headache, myalgia, nausea, and non-bloody diarrhea. During the third, hemorrhagic, phase there is very rapid conjunctival hemorrhage, epistaxis, and melena with fatality rates as high as 60%. The final convalescent stage is seen among survivors who often display symptoms spanning from memory and hearing loss, feeble pulse and tachycardia after 10-20 days of severe illness[7].

Currently, there is no information in the literature that fully explains CCHFV pathogenesis in humans. CCHFV targets the antiviral response in human cells and replicates causing endothelium damage. The pre-inflammatory cytokines such as Interleukins (IL-6) and

Tumour Necrosis factor (TNF-α) are significantly higher among fatal CCHFV-infected patients as opposed to non-fatal infected patient[7].

The treatment of CCHFV-infected patients is usually contingent upon early diagnosis, severity, and fluid intake. In the case of hemorrhagic manifestation, the patients can be supported by replacing blood, platelets and plasma. Recently, randomized control studies revealed that ribavirin has no effect on the mortality of the patients[8]. Although the isolation of immunoglobulin CCHFV-Venin containing the antibodies against the CCHFV infection looks promising, currently there is no evidence to prove its efficacy against this fatal disease[8].

1.4 Ubiquitylation

The innate immune system plays a crucial role in combating infections in the human body such as the CCHFV virus[9]. During viral and bacterial infections, interferon (IFN) production stimulates the transcription of more than 300 IFN-regulated genes whose products deliver anti-viral activity. The up-regulation and post-translational modification of these type I IFN-dependent activities is frequently carried out by ubiquitin molecules through a process called Ubiquitylation. Ubiquitin is an 8.5 kDa regulatory protein found among eukaryotic organisms[9]. Figure 4 shows the secondary structure of the ubiquitin molecule and with its α-helices (blue), β-strands (green) and its 7 lysine side chains (orange coloured).

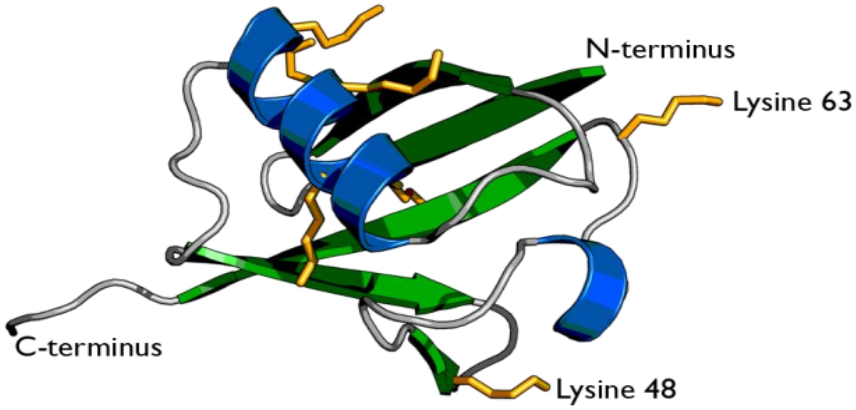


Figure 4. Structure of the Ubiquitin molecule. (https://en.wikipedia.org/wiki/Ubiquitin Rogerdodd (Own work), CC BY-SA 3.0, via Wikimedia Commons). The figure shows the secondary structure of the Ubiquitin molecule showing its alpha helices (blue colored) and beta-strands (green colored), and indicating the positions of its N and C terminal and Lys-48 and Lys-63.

Ubiquitylation is a post-translational modification in which the ubiquitin molecules are attached to a target protein and regulates the functions and properties of the modified protein. As shown in Figure 5, the ubiquitin activating enzyme (E1), conjugating enzyme (E2), and ubiquitin ligase (E3) play vital roles in catalyzing the ubiquitylation process[9]. These ubiquitin-activating enzymes activate ubiquitin by first forming a thiol ester bond between the C-terminus of the ubiquitin and the E1 active-site cysteine through an ATP dependent step. The ubiquitin eventually transfers to the cysteine active sites of the ubiquitin conjugating enzyme (E2) and, with the aid of the ubiquitin ligase (E3), is then transferred to a lysine amino acid of the target protein. This step is highly regulated and is substrate-specific[9]. Ubiquitylation can be a single ubiquitin molecule on a single lysine in a target protein or single ubiquitins can be added to

several lysines. Furthermore, chains of ubiquitins can be added and the chains can be linear or branched[10].

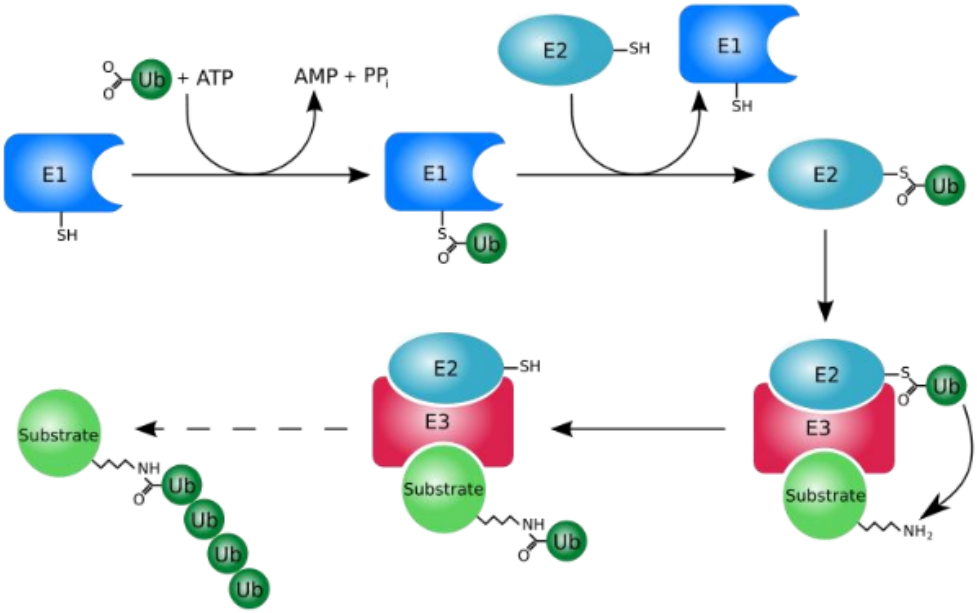


Figure 5. The Ubiquitylation process (https://en.wikipedia.org/wiki/Ubiquitin Roger Dodd at the English language Wikipedia (GFDL), CC-BY-SA-3.0 or, via Wikimedia Commons). The figure shows the ubiquitylation steps by which E1= ubiquitin activating enzyme (blue colored), E2 (blue colored oval) = Ubiquitin conjugating enzyme, and E3 (red colored rectangle) = Ubiquitin ligase enzyme assist in attaching Ubiquitin molecules (dark green colored) to its target substrate (light green colored).

1.5 Deubiquitylation (DUBs)

The enzymes that cleave ubiquitin from its target proteins or ubiquitin pro-proteins are referred to as the deubiquitylating enzymes (DUBs) and play an integral role in the ubiquitylation pathways[9,11]. The DUBs activate ubiquitin pro-proteins by removing the C-terminal ubiquitin

from polyubiquitin or removing it from ribosomal proteins. DUBs also are involved in recycling ubiquitin molecules trapped in thiol ester intermediates with small cellular nucelophiles. DUBs actively reverse the ubiquitylation modifications of target proteins similar to the function of phosphatases in kinase/phosphatase regulatory pathways[9]. Figure 6 shows the ubiquitylation and deubiquitylation of a substrate protein. In the case of CCHFV OTU protease the DUB activity is thought to prevent activation of innate immune pathways.

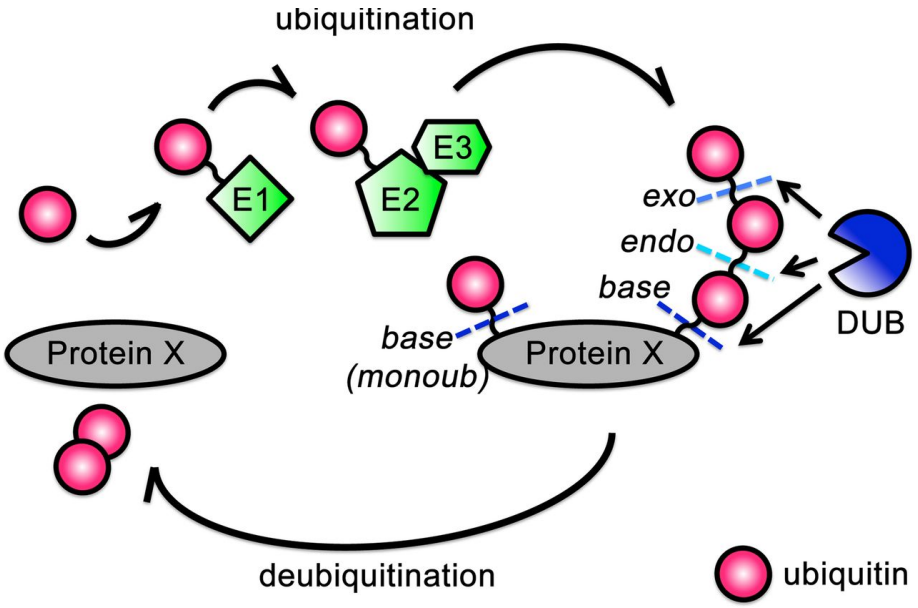


Figure 6. This figure is showing in more detail the ubiquitination and the deubiquitination processes. Continuous ubiquitin addition to one of the seven lysine residues present within ubiquitin results in polymeric ubiquitin chains of different polyubiquitin linkage types. DUBs hydrolyse the isopeptide bonds between the ubiquitins or between the ubiquitin and the target protein, thus deubiquitinating the substrate. DUBS can hydrolyse the chains either from the end (exo activity), at the base of the chain (including monoubiquitination), or from within a polyubiquitin chain (endo activity). Deubiquitination prevents the signaling that activates the innate immune response. This figure is from[11].

1.6 Significance of DUBs

There is still a paucity of information related to the mechanisms responsible for DUBs recognition and association with their substrates and the formation of a catalytically competent conformation with them[12 13]. The modular DUBs contain a ubiquitin-binding domain, a catalytic domain as well as protein-protein interaction domains. These domains assists DUBs in binding and associating with the target substrate molecules with different chain linkages, forming multiple complexes and thus making them more substrate-specific. There are many studies in the literature that focus on the substrate-specificity for ubiquitin-like proteins of DUBs from pathogens[12].

1.7 Families of DUBs: substrate-specificity

The human genome encodes more than 100 putative DUBs that are classified into five different families with four of them belonging to papain-like proteases including ubiquitin C-terminal hydrolase (UCH), ubiquitin-specific protease (USP/UBP), the ovarian tumour domain (OTU) protease, and the Josephin domain[9]. The fifth class is the zinc-dependent metalloprotease family JAB1/MPN/Mov34 metalloenzyme (JAMM). The bacterial and viral DUBs that cleave ubiquitin from the Interferon-stimulated gene product-15 (ISG-15) belong to the small family of Adenain cysteine proteases such as Ubiquitin-like protease (ULP), specific for SUMO (small-ubiquitin like modifier) and papain-like proteases such as the ovarian tumour domain protease[9].

1.7.1 The Ubiquitin C-terminal hydrolase (UCH) domain

Humans have four and *S. cerevisiae* has one UCH domain enzyme composed of a two hundred and thirty amino acid catalytic core domain e.g. UCH-L1 and -L3 isozymes[14]. Figure 7 is an

illustration of a human UCH-L1 DUB domain. These DUBs form a core catalytic fold geometry similar to papain-like proteases[14]. These proteases are involved in processing small substrates with disordered sequences and associating with the C-terminus of ubiquitin. The UCHL-1 and –L3 group of proteases are speculated to be involved in pro-ubiquitin protein processing and degrading trapped ubiquitin. Two human UCH's contain additional C-terminal extensions: UCH37 (100 amino acids) and BAP1 (500 amino acids). The former is involved in cleaving polyubiquitin from target proteins whereas the latter contains a region for a nuclear localization signal and site for binding with the BRCA1 N-terminal ring finger[14].

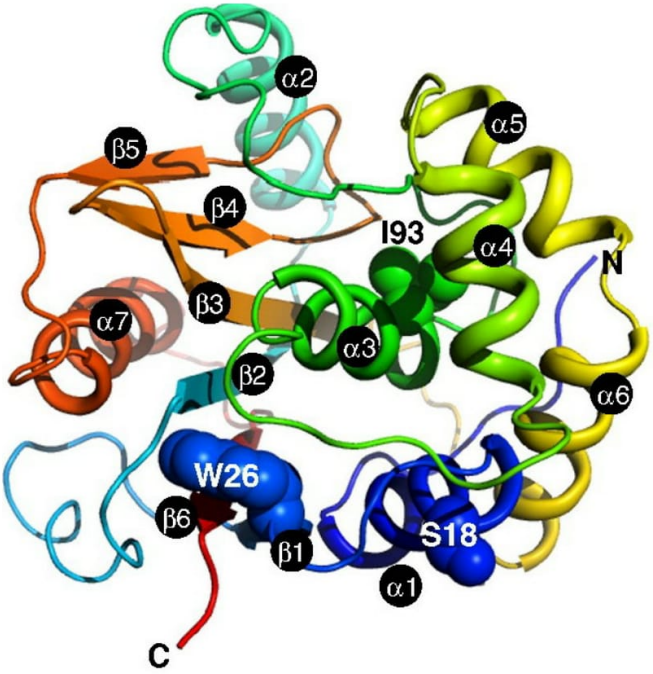


Figure 7. The structure of a human UCH-L1 domain[15]. The figure shows backbone structure of a human UCH-L1 domain protease showing its 7 alpha helices and 5 beta-sheets along with its side chain residues such as Serine (S) 18, Isoleucine (Ile) 93, and Tryptophan (Trp) 26. PDB ID: 1UCH.

1.7.2 Ubiquitin specific protease domain (USP)

The USP's are the largest class of DUBs[16]. Yeast are reported to contain 16 family USP DUBs whereas humans are reported to encode more than 50 USP DUBs. The six structures of these DUBs have a highly conserved domain, low sequence specificity and contain three sub-domains such as the finger, palm and thumb, resembling a right hand (Figure 8). The human benign tumour syndrome cylindromatosis is caused by truncation of the CYLD gene product resulting in a USP that lacks the finger-sub-domain. The majority of these USP's contain a core catalytic domain with terminal extensions and fingers which constitute an extended protein-binding domain[16].

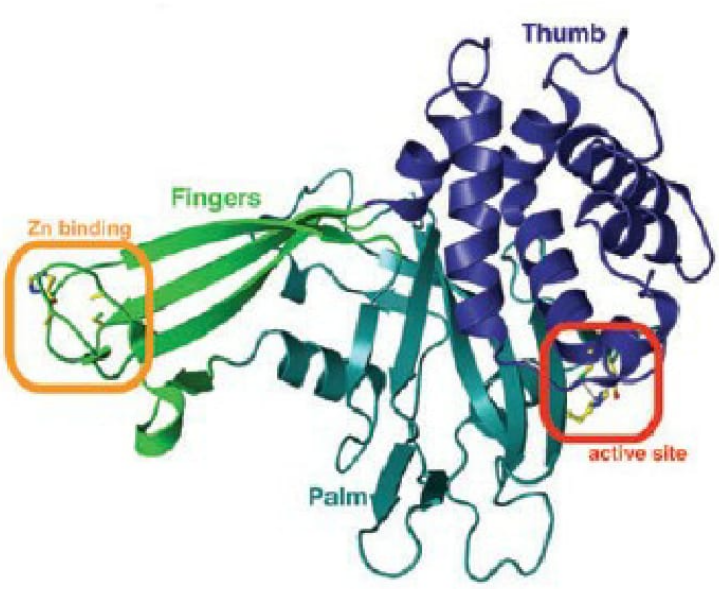


Figure 8. The ubiquitin-specific protease domain. The figure shows the secondary structure of the Ubiquitin-specific protease domain showing Thumb, Palm, and Finger, and Zinc-binding subdomains and the active site containing catalytic triad residues. PDB ID: 2AYN (Simon Caulton (Own work) CC BY-SA 3.0, via Wikimedia Commons. https://en.wikipedia.org/wiki/Deubiquitylating_enzyme).

1.7.3 The Josephin domain

Among the four human members of the Josephin domain DUB family, the best-studied is Ataxin-3 a protein responsible for the Machado-Joseph disease, a neurodegenerative disorder also known as spinocerebellar ataxia type 3 (Figure 9)[17,18]. The underlying cause of this disease is an unstable CAG repeat that results in that addition of a polyglutamine segment, protein misfolding and accumulation. This protein structure has been solved by NMR spectroscopy and correlated with UCH domain ubiquitylating enzymes[18]. The active sites of this protein edit K64-linked chains specifically and are found in a catalytically competent conformation.

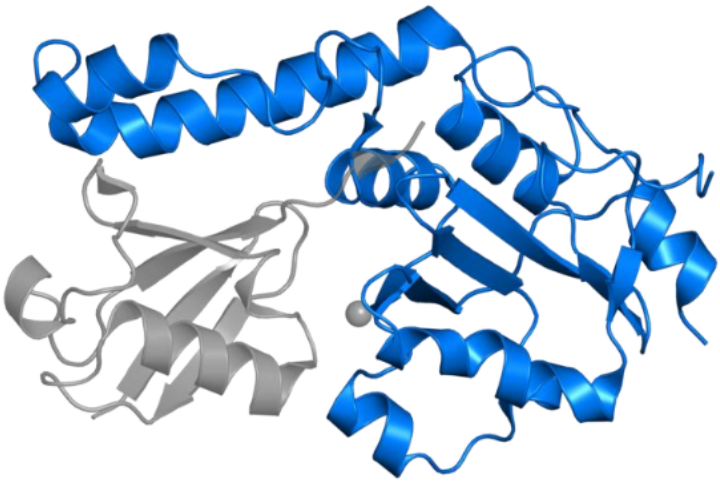


Figure 9. The Josephin domain DUB. The figure shows the crystal structure of the Josephin domain (blue colored) bound to Ubiquitin (grey colored)[19]. PDB ID: 3O65

1.7.4 JAB1/MPN/Mov34 metalloenzyme (JAMM) domain

Out of the four different JAMM family domains, three of them are known to be involved in processing ubiquitylated substrates and the remaining one acts on proteins with the ubiquitin-like

modifier Nedd8[20]. The structure of the AMSH-like (Associated Molecule with the SH3 domain of STAM) DUB protease (Figure 10) involved in vesicle trafficking was solved by biophysical techniques[21]. This structure was the first DUB structure bound to polyubiquitin that explains its role in catalysis. There are other JAMM domain proteins that lack the AMSH-like domain specificity towards polyubiquitin[21].

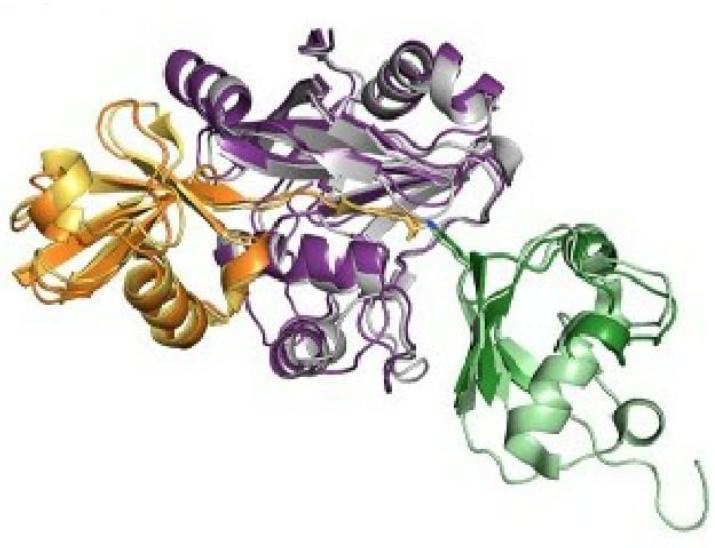


Figure 10. The structure of an AMSH-like DUB protease attached to K-63 kinked diubiquitin[22]. (http://www.rcsb.org/pdb/explore.do?structureId=4NQL). The figure shows the superimposition of the secondary structures of Sst2 (purple) and AMSH-LP (grey) bound to diubiquitin. The proximal Ubiquitin is bound to Sst2 (dark green) and AMSH-LP (light green) whereas the distal ends of diubiquitin are bound to Sst2 (orange colored) and to AMSH-LP (yellow colored). PDB ID 4NQL.

1.7.5 The Ovarian Tumour (OTU) domain protease

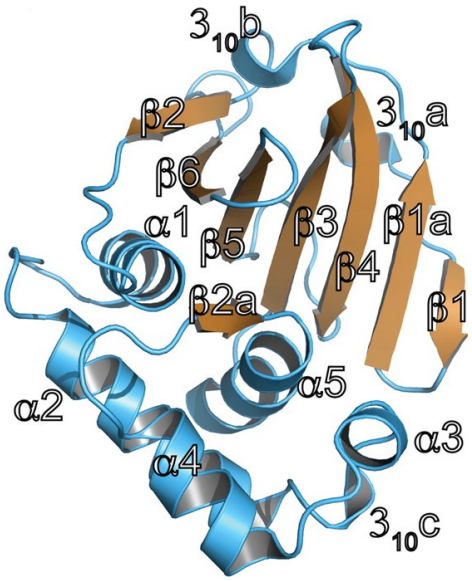
The OTU domain DUBs were named based on their sequence homology with the ovarian tumour gene involved in the development of fruit fly ovaries[23]. The human genome encodes fourteen

OTU domains whereas the yeast encodes only two OTU domains[24]. Many studies have reported the deubiquitylating activity of OTU domains but not all of them are involved in deconjugating Ub or ISG-15 from their target cellular proteins. This is strongly supported by the *Drosophila* ovarian tumour domain protease which has a non-active serine in place of cysteine in the active site[25].

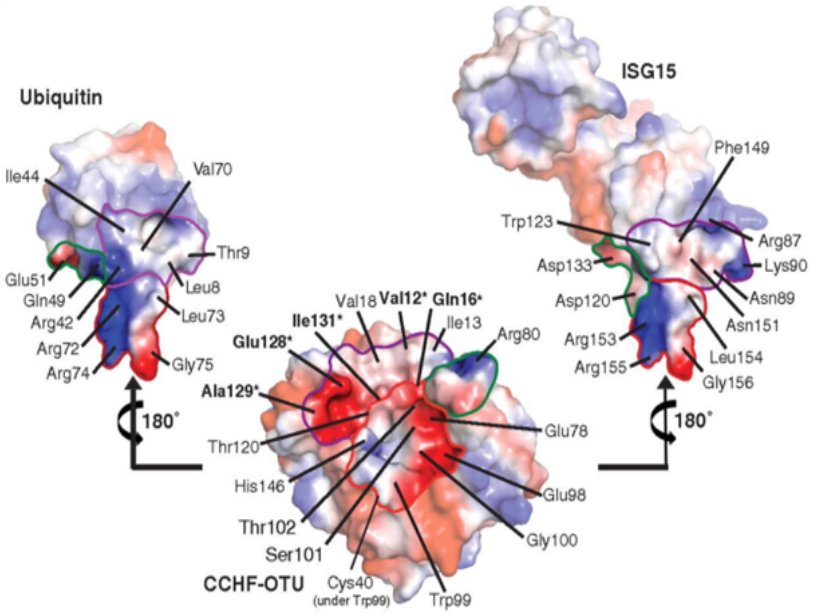
The CCHFV OTU protease is located within the L-protein of the RNA-dependent RNA polymerase[24]. The CCHFV (negative-stranded virus) can replicate inside the host without needing a protease to cleave the viral polypeptide unlike many positive-sense single-stranded RNA viruses[24]. The studies in the literature also claim that the RNA-dependent RNA polymerase activity is independent of the OTU protease domain. The OTU domains can remove both ubiquitin and ISG-15-like molecules from target proteins[24]. To prove OTU domain dual-specificity activity, Pegan *et al.*[26] obtained the crystal structures of the CCHFV OTU bound to Ub, ISG-15, mono-Ub and NEDD8 and measured OTU activity against K-48 and K63-linked poly-Ub[26]. The CCHFV OTU cleaved Ub and ISG-15 molecules with similar kinetics which experimentally proved dual-specific OTU DUB activity.

Recently, the structures of DUB OTU domains were reported for the Nairovirus CCHFV (Figure 11), Dugbe-virus (DUGV), papain-like protease (PRO), and yellow-mosaic fever virus[27,26,28].

Structural biology studies have shown that the OTU core domain contains five β-strands sandwiched between α-helical domains with myriad variations in the OTU family[24] (Figure 11A).



A



B

Figure 11. A) The monomer vOTU domain protease secondary structure: α helices (blue), β strands (orange), and loops (light blue).

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3067871/figure/f3/. PDB ID: 3PSE

B) The structure of the CCHFV OTU protease bound with the ubiquitin (left) and ISG-15 (right)[26]. PDB ID: 3PRM

The crystal structures of the CCHFV OTU bound to Ubiquitin (Figure 11B left) and ISG-15 (Figure 11B right) revealed how the viral protease can bind both proteins[28] whereas the eukaryotic proteases bind only Ub. CCHFV OTU binds the β-grasp fold of Ub and ISG-15 but the molecules are rotated by 75 degrees compared to Ub bound to yeast OTU protease. The unique orientation of the CCHFV OTU N-terminal domain was observed to form a complex with Ile 44 region of ubiquitin forming a completely different orientation compared to eukaryotic OTU[26].

Studies have shown that enzymes are active on various time-scales spanning from milliseconds to microseconds[29]. Conformational selection theory suggests that enzymes alone sample many different conformations that offer potential binding regions for their target substrates during the enzyme turnover[29]. However, there was paucity of information on enzyme dynamics on the CCHFV OTU domain in solution[29]. Recently, the first dynamic information pertaining to CCHFV OTU and its complex with Ub/ISG-15 was reported by Pegan *et al*. (2015)[29]. The backbone dynamics of the OTU and OTU/Ub complex were measured to study the conformational flexibility of the OTU domain protease in solution[29]. The study revealed no significant conformational changes in the CCHFV OTU protease upon binding with Ub[29]. However, subtle conformational changes on the ms time-scales but no significant changes during the microsecond time-scales[29] were detected. This study confirmed that CCHFV OTU domain protease is highly flexible and dynamic in its catalytic region[29].

1.8 NMR spectroscopy and protein dynamics

There is still a lack of detailed understanding of the relationship between molecular dynamics and protein function[30]. Structural biology focuses on the low-energy, ground-state "static" three-

dimensional structures of protein molecules[30]. On the other hand, a complete understanding of a protein's function requires knowledge of its conformational changes on various time-scales[30]. Early hydrogen-deuterium (HD) exchange experiments showed that proteins undergo significant conformational changes at equilibrium. More recently, NMR and photo-dissociation studies showed that proteins undergo massive conformational changes on various time-scales suggesting a wide range of motions with significant amplitudes[30]. Molecular dynamics simulations revealed that the "static" structures of proteins exhibit fluid-like atomic motions on the picosecond time-scale[30]. The work described in this thesis is motivated by the possibility of obtaining a deeper understanding of protein function by the measurement of protein dynamics on a wide range of timescales.

1.9 The Free Energy Landscape

The protein free energy landscape highlights protein motional time-scales and their relations to motional amplitude and protein conformational states[31]. In Figure 12, NMR and X-ray diffraction structures are representative of the ground-state conformations that are present at the bottom of the free energy diagram and are separated from each other by kinetic barriers. The barriers can be overcome by thermal energy-inducing dynamics on the picosecond to nanosecond time-scales, including bond vibrations, unhindered dihedral angle rotations, and possibly chain displacements[31]. Experimentally, in some cases the temperature factors from X-ray crystal structures or NMR relaxation data can reveal picosecond (Tier-2) to nanosecond (Tier-1) dynamics information[31]. The higher free energy states are considered as excited states[31]. The microsecond and millisecond dynamics permit the interconversion between the ground and the excited state conformers and are often associated with protein folding, ligand binding and

enzyme catalysis[31]. In Figure 12 conformation A might represent an apoprotein conformational ensemble whereas conformation B might represent a ligand-bound holoprotein conformational ensemble. The aim is thus, understanding these structural conformers and their interconversions by using biophysical methods to elucidate biological function[31].

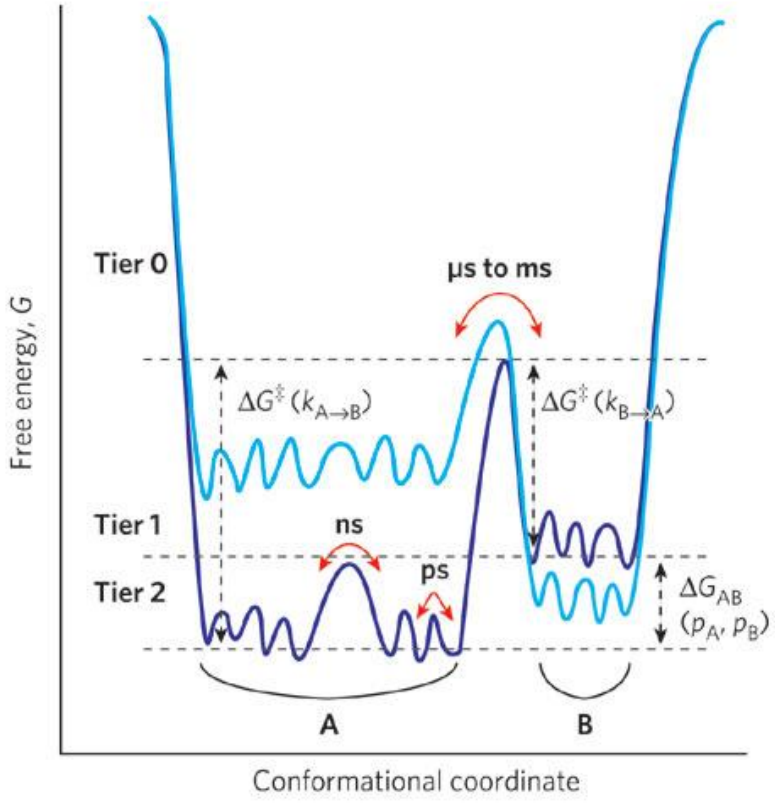


Figure 12. A one-dimensional free energy landscape depicting the protein dynamics hierarchy and possible energy barriers[32]. The figure depicts different timescales of protein motions and the rates of interconversion between the Tier 0, Tier 1 and Tier 2 energy sub-states.

However, there are a couple of issues that need to be addressed in studying excited state conformer dynamics[30]. Firstly, these excited state conformers exist for only a short period of

time. Secondly, the fraction of the excited state conformers may be below 1% compared to those of the ground state and thus highly sensitive methods are needed to capture and study these excited state conformers[30]. Many tools cannot detect these excited conformers because their signals are extremely weak and are not visible[30]. However, NMR experiments such as the *Carr-Purcell Meiboom Gill* relaxation dispersion experiments (CPMG RD), paramagnetic relaxation enhancement (PRE), and HD exchange measurements have advanced our understanding of these invisible excited-state conformers that exist during different biological processes such as protein-ligand interactions, enzyme catalysis and protein folding[30]. In this thesis I apply CPMG relaxation dispersion experiments to search for excited states of apoOTU protease.

1.9.1 Dynamics View of Enzyme Catalysis

Many techniques have been developed to probe conformational changes that occur during protein-substrate binding reactions[33]. For example, the *CPMG* RD and PRE approaches have made progress in elucidating the presence of binding intermediates during the catalytic reactions of some enzymes[34]. These can be explained by using two-binding mechanisms called "conformational selection" and "induced fit" [34,35].

In the conformational selection mechanism the enzyme is thought to exist in a conformational ensemble that includes all of the possible conformations that exist during catalytic turnover[36]. A substrate molecule binds preferentially to a sub-population of enzyme molecules adopting the substrate-bound conformation, shifting the equilibrium of the ensemble toward the substrate-bound state (See Figure 13). In the simplest version of the conformational selection model, the substrate-bound conformation of the enzyme would have all of the enzyme's catalytic groups in place and poised for reaction. In the

induced-fit mechanism, the substrate binds weakly to the apoenzyme in the lowest free-energy conformation and induces a change in conformation to the substrate-bound form[37]. Thus energy from the binding of substrate is used to change the conformational landscape of the enzyme. Experimental support for conformational selection has been obtained for several enzymes that have been shown to sample several conformations that exist during catalytic turnover in the absence of substrate and product[37] In dihydrofolate reductase, product release is proposed to occur from a high-energy sparsely populated state suggesting that conformational selection also explains product release[38]. Finally, it's worth mentioning that enzymes may use aspects of both conformational selection and induced-fit to catalyze reactions.

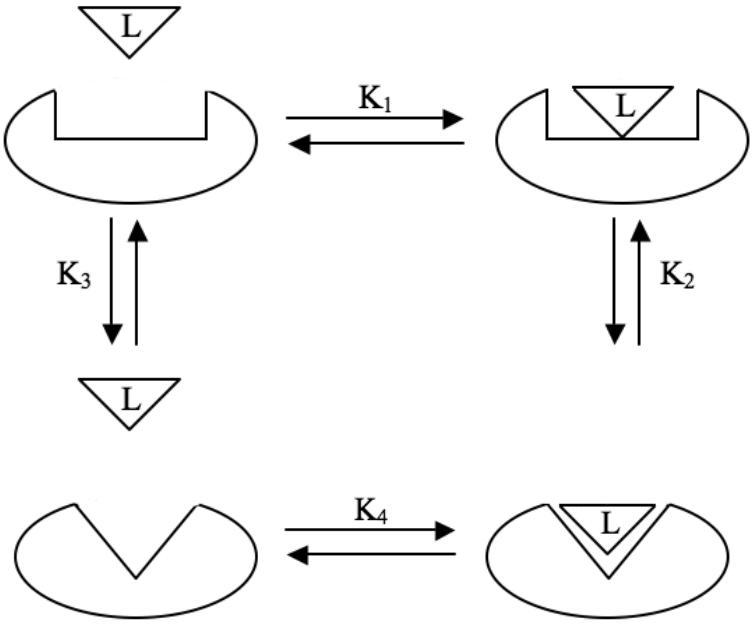


Figure 13. Ligand binding by conformational selection is illustrated by equilibria K_3 and K_4 whereas induced fit is illustrated by equilibria K_1 and K_2. Figure taken from reference[37].

1.10 Reverse-specific or "*Unlabelling*" Experiments

Protein structure determination by NMR spectroscopy requires sequence-specific assignment of all the amino acids in a polypeptide chain[39] and is usually accomplished using triple-resonance approaches. This involves complete backbone assignment of ^{1}H, ^{13}C and ^{15}N nuclei and may include side-chain assignments. The sequential correlation of resonances between all the neighbouring residues of amino acids in the polypeptide chain is accomplished by mapping resonances onto the primary sequence[40,41]. When the spectral data are of high quality automatic assignment procedures can be used to accelerate assignment[41].

When larger molecular weight proteins are being studied or when spectral overlap is severe different approaches are available[41]. These include: (a) Selective labelling, (b) designing specific NMR experiments, and (c) reverse or un-labelling experiments. In the selective labelling approach, the host organism, usually *E. coli*, is grown overnight in minimal medium with a selected isotopically-labelled amino acid whereas the rest of the amino acids are supplemented in the unlabelled form[42]. However, there are 3 principle drawbacks of this method: 1) the employment of expensive $^{13}C/^{15}N$ enriched amino acids, 2) the protein yield from such growing cultures can be very low, and 3) the method only aids in elucidating information pertaining to $^{1}HN/^{15}N/^{13}C$ chemical shifts for the labelled amino acids and fails in correlating a link between $^{1}HN/^{15}N/^{13}C$ resonances of the labelled residues and the unlabelled residues along the polypeptide backbone in the protein molecule[41]. The method is very useful for assigning resonances to amino acid types and can be a powerful way to reduce spectral overlap.

A second approach revolves around developing NMR experiments that can harness the differences in structure among the amino acid side-chains. The desired amino acids are selectively chosen for detection by tuning the magnetization transfer pathway[43]. However, the

main disadvantage of this type of experiment is the rapid loss of sensitivity with higher molecular weight proteins owing to pulse sequences having longer pulse trains resulting in weaker signals being detected owing to relaxation during the pulse trains[41]. Furthermore, deuteration, which is commonly used to lengthen relaxation times in larger proteins, cannot be used because the ^{1}H magnetization is required for the implementation of the pulse sequence[41].

The third approach is the protein "*unlabelling*" or "*reverse-labelling*" method. This is an inexpensive approach involving *unlabelling* the desired amino acid against the ^{13}C or ^{15}N uniform labelled background[44]. Unlabelling is carried out by growing the host organism with $^{15}NH_4Cl$ and/or ^{13}C-D-glucose as the principle nitrogen and carbon sources and with the desired amino acid in the unlabelled form (^{14}N and ^{12}C). Thus, the cross-peaks of these unlabelled amino acids are not observed in the spectrum. They are identified by comparison with a spectrum of uniformly $^{15}N/^{13}C$-labelled protein[44]. This method has the merit of being inexpensive since only unlabelled amino acids with $^{15}NH_4Cl$ and ^{13}C-D-glucose are needed. Another benefit is that it can be used in the deuterated protein sample preparation methods[41]. This method was applied in the current work and proved invaluable to the assignment of the backbone resonances of OTU protease.

Isotope site-specific or inverse labelling experiments are very valuable in solving spectral overlap issues often encountered in NMR experiments especially with larger proteins[45]. However, with isotope-labelling and unlabelling approaches there are number of issues including isotope scrambling of other amino acids in the spectrum due to cross-metabolism that must be dealt with. The most obvious consequence of isotope scrambling is complication in the cross-peaks of amino acids in the spectrum. An amino acid such as glutamic acid cannot be selectively labelled because it is a metabolic precursor to many other amino acids. The alpha amino acids

such as Leu, Val, and Ile undergo isotope scrambling at its alpha-amino acid site to other branched amino acids by bacterial transaminase. An approach to overcoming this problem is to use bacterial cell strains that lack the enzyme alpha-amino transaminase. On the other hand, using these transaminase-deficient strains can have significant negative impact on the cell growth and over-expression of heterologous proteins[45].

1.10.1 Three-dimensional NMR Experiments

Triple-resonance NMR experiments using correlations among ^{15}N, ^{13}C and ^{1}H backbone and side-chain spins provide the most reliable method for resonance assignment of amino acids in the polypeptide chain of proteins[46]. The sequential assignment method depends upon the strong scalar couplings across the peptide bond between nitrogen and carbon atoms to assign the spin systems in the main chain assignment. The large C-C coupling can also be utilized to transmit magnetization using TOCSY sequences (isotropic mixing) throughout the side-chains for side-chain assignment[46].

The heteronuclear scalar couplings are depicted in Figure 1. The one-bond coupling constants are situated adjacent to the bond that links the coupled spins. All of the one-bond couplings are completely independent of the secondary structure. On the other hand, the two-bond coupling between the ^{15}N and $^{13}C^{\alpha}$ is contingent upon the ψ dihedral angle[47].

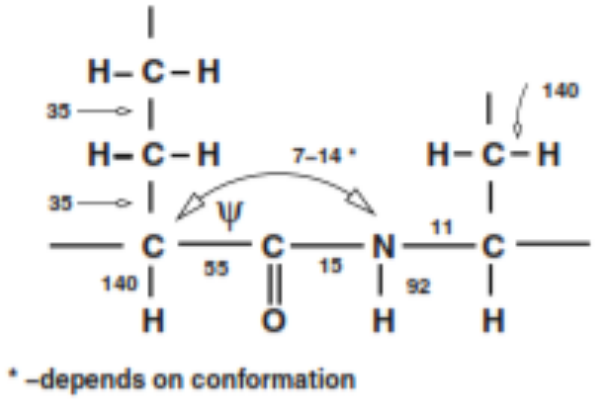


Figure 14. The heteronuclear scalar couplings in proteins[46].

1.10.2 Theory of Backbone Protein Assignment

Figures 15 - 19 below show the magnetization transfer pathways in some important triple-resonance experiments. The atoms in the Figures that are shown in circles indicate the atoms whose resonances are recorded. The atoms indicated with a dotted circle indicate atoms through which magnetization flows but their chemical shifts are not recorded[47]. The magnetization transfers are shown by arrows. All these triple-resonance experiments commence by transfer of the amide proton polarization to the amide nitrogen. The intra-residue shifts are indicated on the right hand sides of the Figures and give rise to one cross-peak per amide group. The matching atom is the carbonyl atom for HN(CA)CO (can generate both inter- with low intensity) and intra-residue carbonyl shifts) and HNCO (generates chemical shifts of the preceding carbonyl residue). The HNCA gives intra- and inter-α-carbon chemical shifts. The HN(CA)CB experiment generates shifts for both inter- and intra- residue beta-carbons with two cross-peaks per amide group. The HN(CO)CA experiment generates only one cross-peak per residue with the inter-

residue alpha-carbon and the HN(COCA)CB experiment generates one cross-peak per residue with the inter-residue C^{β} shift[47].

1.10.3 HN(CA)CO

The HN(CA)CO experiment provides intraresidue crosspeaks between a residue's ^{1}H, ^{15}N and C' using only one-bond correlations as indicated in Figure 15. In this experiment, magnetization is passed through the CA but the CA chemical shift is not recorded. Since the $^{1}J_{NC\alpha}$ is relatively small (8-12 Hz) this experiment is relatively insensitive[46]. Owing to the significance of this experiment during the assignment process, many successful attempts have been made in the past to improve its sensitivity[46].

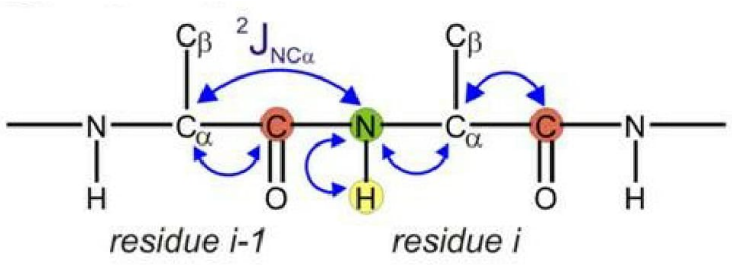


Figure 15. The triple-resonance HN(CA)CO protein backbone coherence transfer pathway[46].

1.10.4 HNCO

The HNCO (Figure 16) is the one of the very first triple-resonance experiments to be developed and is widely used in the assignment process because of its high sensitivity[48].

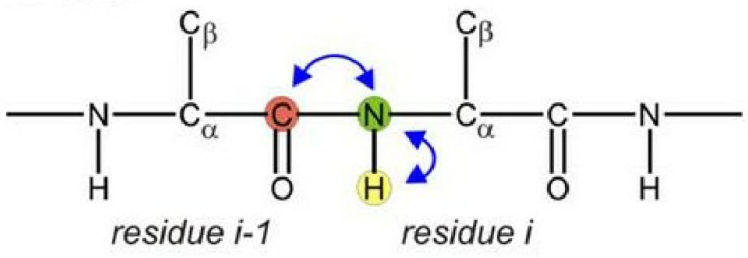


Figure 16. The triple-resonance HNCO protein backbone magnetization transfer pathway[46].

This experiment yields a strong inter-residue cross peak between one residue's ^{1}H and ^{15}N resonances and and the preceding C' using one-bond couplings. This correlation across the peptide bond is important in the sequential assignment process. A second weaker peak is usually also observed correlating the ^{1}H and ^{15}N of a residue with its own C' using the $^{2}J_{NHCO}$. This peak should also have been observed in the HN(CA)CO experiment and this is useful in distinguishing between the intra-residue and inter-residue cross peaks. Therefore, these two experiments are usually utilized together in pairs during the assignment process[49].

1.10.5 CBCA(CO)NH

The CBCA(CO)NH experiment (Figure 17) provides correlations between the ^{15}N and ^{1}H amide resonances of a residue and the $^{13}C^{\beta}$ and $^{13}C^{\alpha}$ resonances of the preceding amino acid in the polypeptide chain[50]. It provides an important connection between the side-chain C^{β} resonances and the backbone resonances.

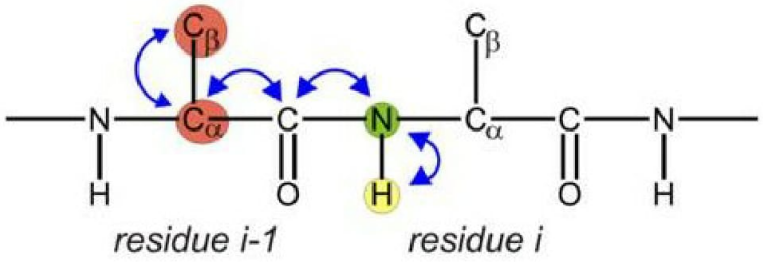


Figure 17. The triple-resonance CBCA(CO)NH coherence transfer pathway[51].

This experiment is also important in the sequential assignment by providing a connection between two neighbouring residues[52].

1.10.6 HNCACB

The HNCACB experiment (Figure 18) provides the same information as the CBCA(CO)NH experiment but also includes connections to a residue's own $^{13}C^{\beta}$ and $^{13}C^{\alpha}$ resonances. The determination of which resonances are intra-residual and which are inter-residual is provided by the complementary CBCA(CO)NH experiment. Usually the intra-residual peaks are stronger than the inter-residual peaks.

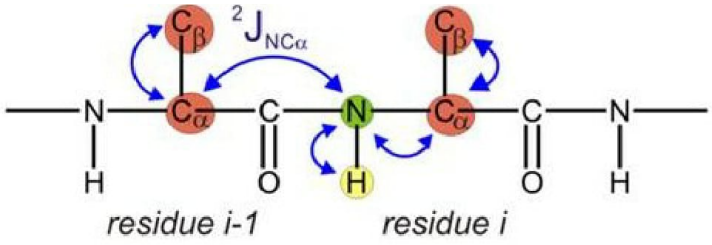


Figure 18. The triple-resonance HN(CA)CB coherence transfer pathway[46].

The CBCA(CO)NH and HN(CA)CB experiments are important in providing partial side-chain assignment of the C^{β} resonances.

1.10.7 HN(CO)CA

The HN(CO)CA (Figure 19) experiment provides intraresidue crosspeaks between a residue's ^{1}H and ^{15}N resonance and the $^{13}C^{\alpha}$ resonances of the preceding residue by passing magnetization through the $^{13}C'$ of the preceding residue. It is complementary to the HNCA experiment that provides a connection to both its own and the preceding $^{13}C^{\alpha}$.

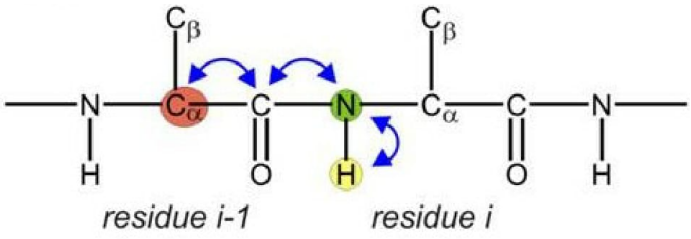


Figure 19. The triple-resonance HN(CO)CA coherence transfer pathway[46].

1.10.8 HNHA

The HNHA experiment is important because it permits assignment of the H^{α} atom to be added to the N, HN, C^{α} and C' assignments completing the assignment of the backbone resonances. Because of the coherence transfer pathway that is followed H^{α} resonances are unambiguously differentiated from H^{β} resonances. This is an advantage over the TOCSY type experiments discussed below. Furthermore, the intensity of the crosspeaks depends on the magnitude of the coupling constant between the HN and H^{α}. The ratio of the crosspeaks to the self peaks can be used to determine the size of the coupling constant and determine the secondary structure at each residue.[47]

1.10.8 Triple-resonance experiments correlating side-chain nuclei.

The triple-resonance HCCH-COSY and HCCH-TOCSY experiments are widely used for studying side-chain resonance assignments. Compared with their homonuclear ^{1}H 2D counterparts, COSY and TOCSY, resonance overlap is greatly reduced because of the third dimension that correlates ^{13}C nuclear shifts to the diagonal protons[53]. The HCCH-TOCSY experiment is often used to overcome the problems associated with the employment of the above mentioned experiments on the protein samples dissolved in D_2O[54]. This experiment connects the, C^{α} and H^{α} to the carbon-attached hydrogens in the side-chain of the same residue[55]. However, the sensitivity is often lowered because of the small scalar couplings. Connections to the amide group by transfer through the carbonyl carbon are preferred because they reduce spectral crowding as only inter-residue peaks are observed. Examples of this approach include the CCH(CO)NH-TOCSY and HCC(CO)NH-TOCSY experiments that use the larger $^{1}JC^{\alpha}C'$ and $^{1}JNC'$-couplings[56].

1.11 NMR Spin Relaxation Measurements

Solution NMR is the most powerful technique for elucidating information about the broad time-scale of protein motions and yields quantitative information at an atomic level[57]. NMR spin relaxation is most commonly employed for providing insights about amide groups in proteins on the ps-ns timescale but is increasingly being applied to probe quantitative information on the microsecond-millisecond timescales in solution. Although there has been enormous progress with regards to obtaining information about protein structure and function using X-ray diffraction and NMR spectroscopy, static structures alone cannot decipher information about the dynamic, motional properties of proteins. Proteins undergo substantial dynamic motion at

various time-scales although they may appear static in the 2D display of 3D structures. Figure 20 clearly shows the hierarchy of protein motions separated by orders of magnitude as well as NMR methods that can probe each of the timescales. The fast time-scale motions include processes such as H transfer and H-bonding formation and breakage on the ps timescale and side-chain rotations on the ns time-scale. On the other hand, the slower-scale motions include ligand binding and release, allosteric regulation and folding/unfolding behaviour of proteins as well as the catalytic rate-limiting steps of enzymes that span the microsecond to seconds to minutes timescales[57]. Comments on measuring these dynamics follow below.

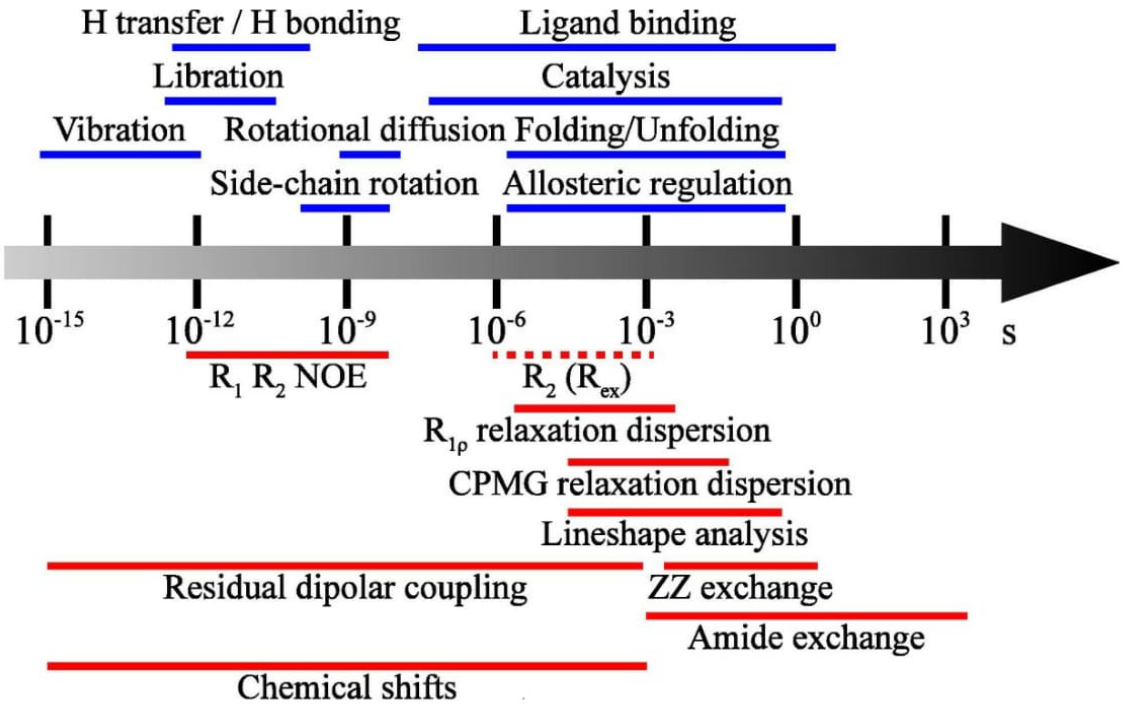


Figure 20. The protein dynamics time-scales are shown on the upper side of the arrow (underlined in blue) and the time-sensitivity of various NMR experiments (including the R_1, R_2, and *NOE*) are shown below the arrow (underlined in red)[57].

1.12 Studying the Dynamical or Flexibility Properties of Proteins using NMR Spectroscopy

By far the most common NMR techniques provide information about the motions of N-H groups of the amide backbones of proteins. ^{15}N spin relaxation is widely used in the NMR analysis of protein motions because of its usefulness in harnessing information about proteins in both fast-scale and slow-scale motions. The dynamic information at the N-H group includes degrees of freedom such as differences in the ϕ and ψ dihedral angles[57].

The laboratory-frame ^{15}N spin relaxation experiments are usually comprised of three measurements: longitudinal relaxation, transverse relaxation and the steady-state heteronuclear NOE. The magnetic field dependencies of the above mentioned spin relaxation parameters allows the elucidation of local and global protein dynamics to extract information about protein molecules on the ps-ns and microsecond-ms time-scales[57]. The local and global dynamics yield important insights into enzyme catalysis such as the rate-limiting steps, allostery, binding free energy, and conformational entropy that are influenced by ps-ms structural motions in proteins[57].

1.12.1 Longitudinal Relaxation (T_1 or $\frac{1}{R_1}$):

Longitudinal relaxation can be understood in two ways. It is the rate at which atomic nuclei become polarized upon introduction to an external magnetic field as for example when a sample is placed in the bore of an NMR magnet. In an NMR experiment, longitudinal relaxation is also the rate at which the thermal equilibrium polarization of the sample is re-established after population disturbance as for example by population equalization by a 90^o RF pulse or population inversion by a 180^o RF pulse. The relaxation rate depends upon the availability of magnetic field fluctuations in the transverse plane at a frequency matching one of the nuclear

resonance frequencies e.g. ω_H, ω_N *etc*. The transverse plane is defined as the plane perpendicular to the applied magnetic field. The magnitude of fluctuations available at any frequency are indicated by the spectral density, *J(ω)*. Equation 1 below shows the amide ^{15}N longitudinal relaxation rate as a function of the spectral density of motions[58].

$R_1 = \mathrm{d}\,[\mathrm{J}\,(\omega_H - \omega_N) + 3J\,(\omega_N) + 6J(\omega_H + \omega_N)] + cJ(\omega_N)$ ..... (1)

where d, the dipolar constant is

$$d = \frac{1}{4}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{(\hbar\gamma_N\gamma_H)^2}{10*r^6_{N-H}} \quad \text{..... (2)}$$

and the chemical shift anisotropy can also be explained by the following equation[58].

$$\mathrm{c} = \frac{(\omega_N \triangle\sigma)^2}{3} \text{..... (3)}$$

$J(\omega)$ = spectral density values at distinct frequencies.

$\hbar$ = Planck's constant divided by 2π.

μ_0 is the permittivity of free space.

γ_N and γ_H are the gyromagnetic ratios of ^{15}N and ^{1}H.

$r_{N\text{-}H}$ = N-H bond length.

$\Delta\sigma$ = Chemical Shift anisotropy where σ is the chemical shielding symbol;

$\omega_N = 2\pi\nu_N = \gamma_N B_0$;

where ν_N = Larmor frequency and B_0 = magnetic field strength.

T_1 (longitudinal relaxation time) is the reciprocal of R_1 ($T_1 = 1/R_1$)[58].

The measurement of R_1 in a 1D experiment is carried out by an inverse-recovery approach[59]. Figure 21 depicts the time-dependence of the magnetisation in such experiments. The experiment commences by placing the magnetisation along the –z axis with a 180° pulse[60]. Then a 90° pulse places the magnetisation into the xy plane for measurement after a delay to

permit spin relaxation. This is shown in Figure 22. The return to Boltzmann equilibrium is depicted in the following relation[57,60]:

$A_t = A_0(1 - 2e^{-R1t})$……..(4)

Where, A_t = the signal amplitude after an inter-pulse delay t and A_0= the amplitude at time 0.

On the other hand, the R_1 measurement during a 2D double INEPT-type pulse sequence to overcome the issues of overlapping peaks is given in the following equation[60]:

$A_t = A_o e^{-R1t}$…….. (5)

A comparison of the signal behaviour in the two experiments is given in Figure 21[61].

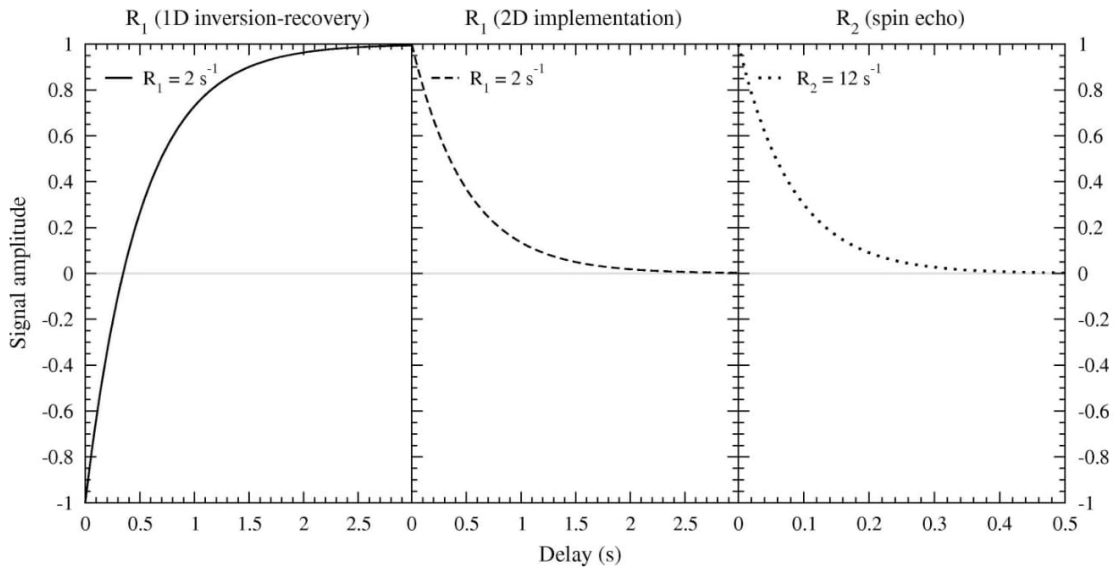


Figure 21. Relaxation rates measurements measured from signal amplitudes as a function of relaxation delay[58]. The left panel shows the 1D-inverse recovery scheme for longitudinal relaxation (equation-4), and the middle panel shows the 2D-implementation scheme. The right panel shows the behaviour of spins during the measurement of transverse relaxation R_2 in a spin echo experiment (equation-5). The figure is from[57].

R_1 also has myriad applications apart from measuring dynamics in protein studies. It has a significant impact on acquisition of NMR results[60]. Thus, in a multi-acquisition experiments if

full relaxation does not occur following each signal acquisition the full intensity of the signals will not be measureable. For nuclei that relax slowly it is advisable to adjust the relaxation delay and pulse repetition rate to optimize signal measurement. For example, near complete recovery of magnetisation can be achieved after a delay of $3T_1$. However, the repetition rate is usually adjusted to $1\text{-}1.5T_1$ for most experiments because nearly complete relaxation has been achieved by this time and the sensitivity per unit time is maximized[57].

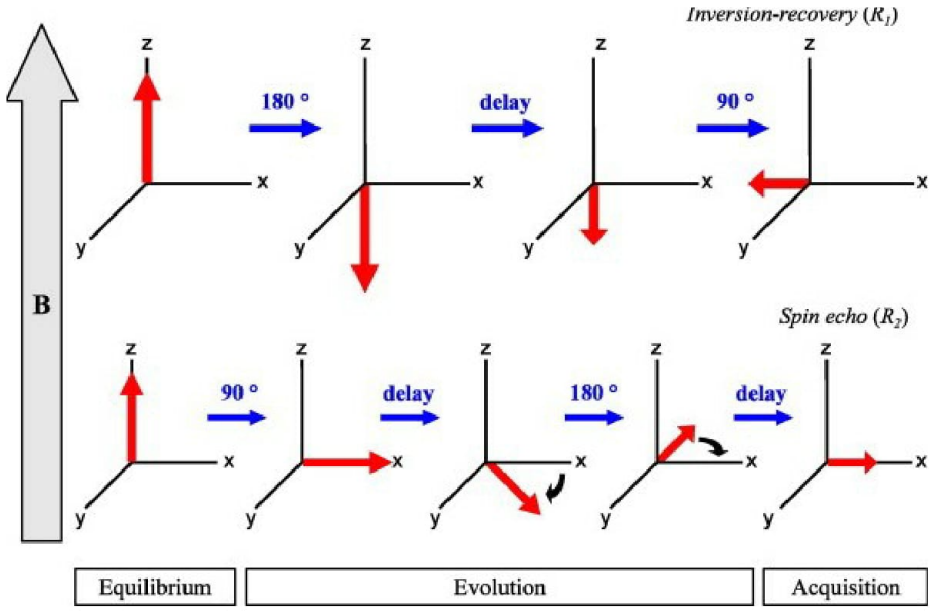


Figure 22. Pulse sequences for the measurement of R_1 and R_2. The red arrow represents the bulk magnetisation (M_0). The signal amplitudes of all the relaxation measurements is contingent upon varying the delays between the pulses as shown in the Figure. The delays are of the same duration in case of the spin echo experiment. The figure only considers the R_1 and R_2 experiments. The figure is from[57].

1.12.2 Transverse Relaxation (T_2 *or* $\frac{1}{R_2}$):

The transverse relaxation rate or the spin-spin relaxation rate (R_2) is the rate at which nuclear spins lose coherence in the transverse plane following the application of an RF pulse. Fluctuations in magnetic fields in the transverse plane that cause longitudinal relaxation also cause transverse relaxation. In addition, magnetic field fluctuations perpendicular to the transverse plane can also cause changes in nuclear precession frequency and thus a loss of coherence among the spins[58]. Equation 6 below shows the amide ^{15}N transverse relaxation rate as a function of the spectral density of motions[58]:

$$R_2{}^0 = \frac{d}{2}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c}{6}[4J(0) + 3J(\omega_N)]$$

in which, $R_2 = R^0{}_2 + R_{ex}$ …… (6)

R_{ex} = the contribution to R_2 resulting from slow conformational exchange in the time-scale range of microseconds to milliseconds. For example for exchange between two states A and B the following equation can be written:

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B \qquad \text{…… (7)}$$

where k_1 and k_{-1} are apparent exchange rates between the two states and $k_{ex} = k_1 + k_{-1}$. The populations of the two states are usually represented by p_A and p_B and $p_A + p_B = 1$. Each state is assumed to have a distinct chemical shift ω_A and ω_B and $\Delta\omega = |\omega_A - \omega_B|$. The rate of chemical or conformational change can influence the appearance of the NMR resonances altering both their width and position in the spectrum depending on the exchange regime. For fast exchange where $(k_{ex} / \Delta\omega) < 1$ a population-wieghted single resonance appears in the spectrum. For intermediate

exchange where (k_{ex} / $\Delta\omega$) ~ *1* a single broad peak may or may not be observed. In the slow exchange regime where (k_{ex} / $\Delta\omega$) > *1* two narrow lines are observed for each conformation. In the intermediate fast exchange regime where k_{ex} is slightly greater than $\Delta\omega$, R_{ex} is then given by[62]

$$R_{ex} \sim (p_A)(p_B)\,\Delta\omega^2/k_{ex}. \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

The exchange regime can be obtained from calculating the parameter alpha:

$$\propto = \frac{2\,(\frac{k_{ex}}{\Delta\omega})^2}{1+(\frac{k_{ex}}{\Delta\omega})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)^{57}$$

When α is between the range of 0-1, then the exchange can be considered as slow; when α is approximately 1 then exchange is intermediate and when α is between 1-2 then the exchange can be regarded as fast[63,62].

To measure R_2, the spin-echo approach is employed in a 2D ^{1}H-^{15}N double INEPT pulse sequence as shown in Figure 28[64,60]. The 90^o excitation pulse is followed by a series of spin echoes that serve to refocus magnetic field inhomogeneity contributions to coherence loss. Spin echoes consist of a 180^o refocusing pulse between two delays of equal length. By refocusing coherence losses owing to magnetic field inhomogeneity the intrinsic rate of coherence loss (R_2) can be determined.

1.12.3 Steady-state Heteronuclear Nuclear Overhauser Effect (*NOE*):

The steady-state heteronuclear Nuclear Overhauser Effect is the response of one spin to the perturbation of the population of another spin involved in a nuclear dipole-dipole interaction with the interrogated spin. In the present context, NOE measurements are made in order to

obtain the heteronuclear cross-relaxation rate, $J(\omega_H + \omega_N) - J(\omega_H - \omega_N)$. The relationship between the NOE, R_1 and the spectral density of motions that are the source of the relaxation is shown in equation 10[60]:

$$NOE = 1 + \frac{\gamma_H}{\gamma_N}\frac{d}{R_1}\,[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (10)$$

The NOE is also contingent upon the distances between the two spins, the CSA, the relative gyromagnetic ratios of the spins, and N-H bond motions[57].

The *NOE*s are acquired by recording peak intensities in two different experiments with (A_{sat}) and without (A_{eq}) saturation of the ^{1}H spins. In the former experiment, the saturated ^{1}H spin populations are equalized. The *NOE* is calculated by taking the ratio of the peaks in the two experiments as follows:

$$NOE = \frac{A_{sat} - A_{eq}}{A_{eq}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

There are differences between heteronuclear *NOE* experiments used to study dynamics aspects of proteins and homonuclear ^{1}H NOE experiments used to obtain distance restraints for the calculation of structures[57]. In the former experiment the differences in the *NOE* are associated with variations in the motions of the N-H bond vector whereas the N-H bond length is assumed to be constant during the experiment. On the other hand, the motional functions are approximately taken as constant during structure determination and the source of variation in NOE values is assumed to mainly be due to distances between the ^{1}H nuclei[57].

1.13 Spectral Density Mapping

Equations 1, 6 and 10 show the relationship between the measured relaxation rates (R_1, R_2, and NOE) and the values of the spectral density at the relevant NMR transition frequencies ($J(\omega)$). A

simple approach to visualizing the connection between the spin relaxation measurements and the motions associated with the bond vector is to map out the spectral density function values[65]. This approach does not require any physical model of the protein in question and is very general[65]. Thus, spectral density mapping is the simplest technique to analyze raw relaxation data.

Reduced spectral density mapping is an even simpler approach that assumes that the spectral density curve is flat at higher frequencies so that the high-frequency spectral density values are approximately equal[65]. However, the reduced spectral density mapping can differentiate among protein motions that contribute to relaxation on the basis of being fast, slower or intermediate. As explained by Morin[57], the simplest approach is first to calculate reduced spectral density values according to the following equations[65,66]:

$$J(0) = \frac{-1.5}{3d+c}\left(\frac{R_1}{2} - R_2 + 0.6\sigma_{NOE}\right) \quad \text{(13)}$$

$$J(\omega_N) = \frac{1}{3d+c}(R_1 - 1.4\sigma_{NOE}) \quad \text{(14)}$$

$$< J(\omega_H) >= \frac{\sigma_{NOE}}{5d} \quad \text{(15)}$$

where $<J(\omega_H)>$ approximates $J(\omega_H)$, $J(\omega_H+\omega_N)$ and $J(\omega_H-\omega_N)$ and $\sigma_{NOE} = (NOE - 1)R_1\frac{\gamma_N}{\gamma_H}$. $<J(\omega_H)>$ is also symbolized as $J(0.87)$. Re-expressing the high-frequency spectral densities in terms of $J(0)$ gives:

$$J(\omega_N) = \frac{J(0)}{1+(6.25\omega_N J(0))^2}$$

and

$$< J(\omega_H) >= \frac{J(0)}{1+(6.25\omega_H J(0))^2}$$

Then, plotting $J(\omega_N)$ or $<J(\omega_H)>$ against $J(0)$ can lead to insights into the dynamic behaviour of different parts of a protein. Figure 23, the graphical analysis can be used for qualitative analysis to differentiate between residues on the basis of their dynamics[57]. In the diagram, relaxation of

residues near point P is dominated by slow overall tumbling of the molecule on the ns timescale whereas residues at point Q have relaxation dominated by fast motions on the ps timescale. Most residues will fall on a line between P and Q indicating contributions from slow and fast dynamics. On the other hand, residues where relaxation has significant contributions from dynamics on the ms – μs timescale are displaced to the right and exhibit larger *J(0)* values[67].

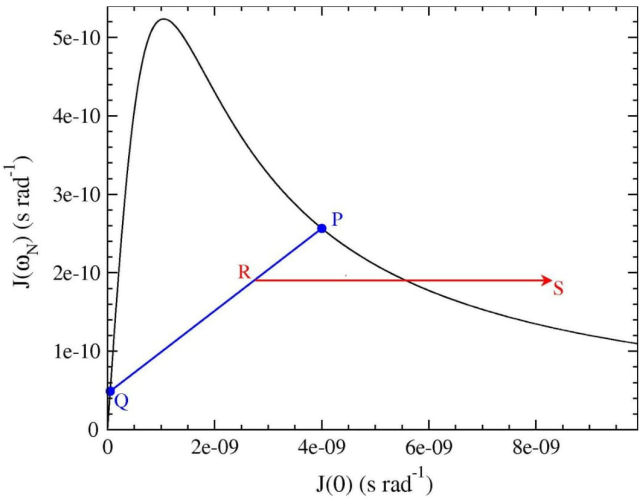


Figure 23. Reduced spectral density mapping analysed graphically[57]. The P point in the graph shows residues undergoing slow overall tumbling (nanosecond timescale), Q point implies residues undergoing relaxation by fast internal motions, and the residues falling between P and Q indicate contributions from both the slow and fast scale motions.

Figure 24 is particularly instructive. It shows the changes in behaviour of residues in MUP-1 protein at different temperatures and in the presence or absence of substrate[68]. Residues that relax predominantly by motions associated with overall rotational correlation of the protein on the ns timescale have a low $J(\omega_N)/J(0)$ and $J(0.87\omega_H)/J(0)$ and appear at the middle of the plots A and B (dark blue). As the temperature is increased those residues move leftward and

upward in the diagram as the rotational correlation time of the protein decreases. Residues in the protein that are disordered cluster on the left side of the plot and move in the opposite direction as the temperature is increased. Their relaxation is dominated by motions faster than the overall molecular rotational correlation as shown in Figure 24C. Residues in the protein that are in slow conformational exchange show a different behaviour with temperature as illustrated in Figure 24D.

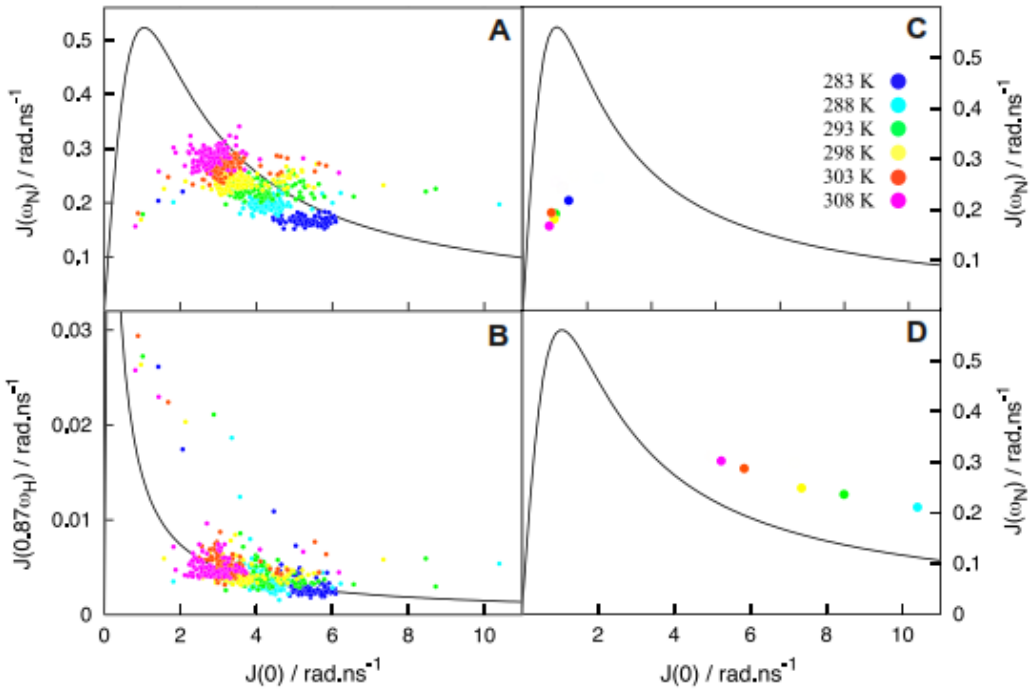


Figure 24. The temperature dependence spectral density mapping analysis for MUP-1 protein[57]. Different colors correlate with different temperatures. The protein residues in the middle of Panel A and B are associated with overall rotational correlation of the protein on the ns timescale. The protein residues in Panel C at higher temperature shows disordered residues clustering on the left hand side of the plot and their relaxation is faster than the overall molecular rotational correlation. Panel D show residues in the slow conformational exchange regime.

1.14 Model-free Formalism

The Lipari-Szabo model-free analysis applies a mathematical model to the relaxation data that is completely independent of any physical model of the protein. Although it is abstract, this model yields valuable information about residue-specific dynamics in proteins that can yield insights into their biological function[69,70,71,72]. There are three main parameters within this model-free formalism: S^2, τ_c, and τ_e. The S^2 generalized order parameter provides residue-specific information about the extent of N-H bond vector motion on a scale of 0 – 1. When a residue has motion restricted to that of the overall rotational correlation of the protein S^2 is 1. If motion of the backbone is completely unrestricted S^2 is 0. Values between 0 and 1 indicate the relative freedom of motion of the backbone. The average τ_c for all restricted residues indicates the overall rotational correlation of the protein. The Model also yields the timescale of the "internal" motion for each residue τ_e. However, it should be noted that τ_e contains information both about the spatial extent of the internal motion and its timescale. One of the assumptions in this approach is that the internal and overall motions occur on significantly different timescales so that they occur independently of each other. Sometimes a fourth parameter, R_{ex}, is added when the fits to some residues are poor as this sometimes indicates the presence of conformational exchange at those residues. R_{ex} thus can indicate the presence of microsecond-millisecond timescale motions accounting for slow motions and contributions to T_2 $(1/R_2)$ [69,70,71,72]

Lipari and *Szabo*[69, 70] first introduced model-free formalism and Clore *et al.*[71] and d'Auvergne[72] further extended this formalism which revolves about the hypothesis that the internal motions and the global tumbling of the molecule can be separated from each other. However, this theorem is only considered when the global tumbling range is between a few ns to a few tens of ns and the local motions span from tens to hundreds of ps timescales. The

decoupled global and internal motions are defined by a correlation function which forms a Fourier pair with the reduced spectral density. The correlation functions for global tumbling (Brownian rotational diffusion) and the internal bond motions of the N-H groups are defined as follows[73]:

$$C_O(\tau) = \frac{1}{5}\sum_{i=-k}^{k} W_i . e^{-\tau/\tau i} \quad \text{(16)}$$

$$C_I(\tau) = S^2 + \left(1 - {S_f}^2\right)e^{-\tau/\tau f} + ({S_f}^2 - S^2)e^{-\tau/\tau s} \quad \text{(17)}$$

Where, C(τ) is a correlation function, O = overall tumbling, I = internal motions, *W* is a weighting factor, and τ_s and τ_f refer to slow and fast internal motions. The index *k* ranges over the number of exponential terms used to describe the correlation function. For a molecule with spherical symmetry k = 0 and τ_i is replaced by τ_c. For an ellipsoidal molecule k = 2 and both prolate and oblate ellipsoid dynamics can be described.

If a Lorentzian Spectral Density and isotropic diffusion are assumed the following equation can be derived that is the other half of the Fourier pair:

$$J(\omega)_{LS} = \frac{2}{5}\left[\frac{S^2\tau_c}{1+(\omega\tau_C)^2}\right] + \frac{(1-S^2)\tau}{1+(\omega\tau)^2} \quad \textit{(18)}$$

Where, $(1/\tau) = (1/\tau_c) + (1/\tau_e)$.

A number of different models including or excluding fast and slow internal motion and R_{ex} can be used to fit the NMR relaxation data and generate parameters such as S^2 and τ_e[69, 70 71 72 73 74 75 76 77 78 72, 79]. The process by which models are selected use statistical methods to pick the best model and will not be described here[80,81].

There are various programs for calculation and optimisation of model-free parameters such as Modelfree[81,82,83], FAST model-free[83], DASHA[84], and a Mathematica™ (Wolfram Research, Champaign, IL) notebook developed by Leo Spyracopoulos[85].

1.15 *Carr-Purcell-Meiboom-Gill* Relaxation Dispersion (CPMG-RD) Experiments

Protein functions are thought to depend on the transition from the ground-state to the excited state[86]. The changes in the ground-state energy level are the result of conformational flexibility and may even affect chemical reactivity in enzymes. Understanding the kinetics and conformational dynamics of proteins can reveal more information about their biological functions[87 86]. These dynamic processes can be easily studied by various NMR spectroscopy techniques which are designed to study these processes on different time-scales. Protein dynamics on the ps-ms time-scale can be obtained from NMR spin relaxation experiments[86]. For example, heteronuclear laboratory frame spin relaxation[88] NMR spectroscopy is used to characterize the intramolecular motions of proteins on ps-ns timescales by measuring the relaxation rate constants and heteronuclear NOE (Nuclear Overhauser Effect). The relaxation rate constants are contingent upon the spectral density of motion functions that contain information about the frequency dependence of stochastic protein motions[87,89,90]. The motions include overall rotational diffusion and intramolecular dynamics at atomic sites and can be extracted during interpretation of the relaxation data[87,90,91].

A number of methods are available for characterizing the conformational dynamics of protein molecules from slow motions on the μs-ms timescales. This is important since many biological processes occur within this timescale. In solution NMR, experiments such as ^{13}C and ^{15}N heteronuclear ZZ-exchange, *Carr-Purcell* Meiboom Gill relaxation dispersion experiments[92], and $R_1\rho^{15}N$ relaxation methods can be used to obtain dynamics information on these timescales[93]. In addition, there are other techniques such as water relaxation dispersion and amide proton solvent exchange[93] that are informative. A major shortcoming of these latter

complementary techniques is that they reveal less direct information about protein motions needed for hydrogen exchange with the bulk solvent[93].

The *Carr-Purcell*-Meiboom-Gill relaxation dispersion (CPMG-RD) experiment measures the effective transverse relaxation rate, R_{2eff}, of a spin probe to monitor conformational exchange on the ms timescale. The principle is fairly simple. The measurement of intrinsic transverse relaxation rates (R_2) is done using a CPMG train of 180° pulses to refocus magnetization that is dephasing in the transverse plane as a result of, for example, field inhomogeneities across the sample. If dephasing occurs owing to conformational exchange, high-frequency CPMG pulse trains will be more effective at refocusing transverse magnetization than low frequency pulse trains. A series of experiments are done with different frequencies of CPMG trains. The effect of conformational exchange is revealed when the ratios of the signal intensities compared to a reference signal intensity in the absence of a CPMG pulse train is plotted as a function of CPMG frequency. If the profiles are flat, this indicates that no conformational exchange is occurring on the ms timescale. Decays in the profiles as a function of CPMG frequency indicate conformational exchange and fits of the exchange profiles to exchange models can yield the conformational exchange rate constant (see section 2.11). Different models have been developed to characterize exchange between two or more conformations[94]. Exchanging populations can also be acquired as well as structural insights from the chemical shift differences between the conformations. There has been enormous progress in applying the CPMG-RD experiments to elucidate site-specific information of protein dynamics. The most important impediment of CPMG-RD experiment in studying dynamics of complex macromolecules such as proteins is the need for high spectral resolution and higher dimensionality NMR. Hence, two-dimensional

correlation maps are now acquired through employment of one-bond coupled ^{1}H-^{15}N or ^{1}H-^{13}C spin pairs. This approach resolves the issues of low resolution and improves sensitivity[94].

1.16 OBJECTIVES OF MY THESIS

Nuclear Magnetic Resonance (NMR) spectroscopy is not only of great use for elucidating the atomic structures of proteins, but arguably it is the most information-rich technique for studying the dynamic properties of proteins. My hypothesis is that the conformational dynamics of CCHFV OTU protease change upon interaction with substrate and that this contributes to catalysis. I also hypothesize that all of the conformational states that the protein encounters during catalytic turnover are present in the conformational ensemble of the apoprotein. My hypothesis will be addressed by over-expressing and purifying OTU protease from bacteria grown in isotopically-enriched medium. NMR spectroscopy will be used to assign the resonances and characterize the backbone dynamics of the apoprotein on timescales ranging from milliseconds to nanoseconds. Experiments like *Carr-Purcell*-Meiboom-Gill relaxation dispersion and spectral density mapping will help reveal apoOTU protease polypeptide catalytic binding site dynamics to provide insights into the conformational changes that occur during catalysis. The findings of my study might provide valuable information for advancing the development of pharmaceutical inhibitors for all viral-specific OTU proteases in the *Nairovirus* family.

2. MATERIALS AND METHODS

2.1 Chemicals and Reagents

Unless indicated below, all the reagents and biochemicals utilized during the study were procured from Sigma-Aldrich (Missouri, USA), Acros Organics (New Jersey, USA), Mallinckrodt (Kentucky, USA), Fisher Scientific (New Jersey, USA) and, Cambridge Isotope Laboratories (Massachusetts, USA).

2.2 Cloning, Protein Over-expression and Purification

The gene encoding CCHFV-OUT was cloned into the pGEX6P vector (Stratagene, La Jolla, CA) with a Glutathione-S-Transferase (GST) purification segment and transformed into *E. coli* BL21-Gold (DE3) cells for protein over-expression by T. James of the Department of Microbiology, University of Manitoba. Isotope-labelled (^{15}N) proteins were prepared by inoculating with these transformed cells 5 ml of LB consisting of 10 g/L of Tryptone, 5 g/L of Yeast extract (Fluka Analytical, MO, USA), 10 g/L NaCl, pH (6.8-7.2) and also containing the antibiotic ampicillin at 150 μg/mL. The cells were grown overnight (18-20 hrs) at 37°C. The overnight cell culture was added to 1000 mL of LB medium containing ampicillin and grown to an OD^{600} of 0.8-1.2 (usually about 2 hr 30 min) in an orbital shaker at 260 rpm at 37°C. The cells were collected by centrifugation at 4000 rpm at 4°C for 15 min. The cell pellets were washed, re-suspended and grown in 500 mL of M9 Minimal Medium containing 6.4 g Na_2HPO_4, 1.5 g KH_2PO_4, pH 6.5, 0.25 g NaCl, 0.25 g $^{15}NH_4Cl$, 1mL 1 M $MgSO_4$, 50 μL 1 M $CaCl_2$, and 20 mL of 1 M D-glucose. For the purpose of amino acid unlabelling the cell culture also contained 1 g/L of a natural abundance isotopes amino acids. Acid-soluble amino acids (Asn, Asp, Cys, Glu, Gln, Leu, Met,

Trp, Tyr) were first dissolved in 50 mL of Millipore water and 5 mL of concentrated 98% hydrochloric acid; the base-soluble amino acids (Ile, Phe) were dissolved in 5 mL of concentrated sodium hydroxide before addition to the growth medium. The cell culture was incubated for 45 min at 37°C and shaken at 260 rpm. After 45 min, 1mM IPTG and 20 mL of 1 M D-glucose were added to the M9 Minimal Medium cell culture and cells were grown overnight (18-20 hrs) at 220 rpm and 16°C. The cells were then centrifuged and the pellets were suspended in 20 mL of Lysis buffer (50 mM Tris, pH 7.2, 200 mM NaCl, 5 mM EDTA, and 5 mM DTT) after overnight expression. The cell contents from the lysate were extracted by French Press using a 20,000 pounds per square inch (psi) limit cell and the clarified lysate was then centrifuged at 12,000 rpm for 1 hr at 4°C. The soluble lysate supernatant was passed over a 3 mL pre-equilibrated Glutathione-S-Transferase-Bind ^{TM} (Novagen) resin column and washed with 30 mL of Lysis buffer at 1 mL/minute. The resin was next washed with 20 mL of Lysis buffer, followed by 40 mL of Wash buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 5 mM EDTA, 5 mM DTT) and finally the bound OTU protein was eluted by washing the column with Elution buffer (50 mM Tris, pH 7.2, 500 mM NaCl, 5 mM EDTA, 15 mM glutathione, 5 mM DTT) at 1 mL/ min. HRV3c protease (200 µL of 20 mg/ml) was directly added to the eluted sample before dialyzing it over-night with Cleavage buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) to cleave the GST tag from the OTU protein. The protease-treated sample was again passed over the Glutathione-S-Transferase-Bind^{TM} resin to separate GST tags from the protein sample the following morning. The eluted protein was dialyzed overnight in Gel-filtration buffer (50 mM phosphate buffer, pH 6.5, 150 mM NaCl, 1 mM DTT). Next, the dialyzed protein was gel-filtered using an AKTA-FLPC (Fast Pressure Liquid Chromatography) (GE Healthcare) column and eluted with Gel Filtration Buffer at 1 mL/min with continuous absorption at 280 nm

detection of column effluent. Gradient buffer pumps and automated fraction collection were used for large-scale protein purification. The fractions of monomeric OTU protein corresponding to peaks between 70-80 min having an absorption at 280 nm in the chromatograms were pooled and collected for analysis (see Section 3.1). The sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out by Dr. B. Khare from the department of Microbiology to detect and validate OTU protein collected from the FPLC-pooled fractions before proceeding with the NMR experiments.

2.3 NMR Sample Preparation and Processing

The ^{15}N-labelled OTU protein samples (0.5 mM-1.2 mM) containing 50 mM phosphate buffer, pH 6.5, 150 mM NaCl, and 1 mM DTT were transferred into 5 mm NMR tubes (Wilmad-LabGlass, Vineland, NJ). Between 35 and 42 µL of 75 µM, 2, 2-dimethyl-2-silapentane-5-sulfonic acid (DSS) and 5 µL of deuterium oxide (D_2O) (99.9%) were directly added to the protein solution in the NMR tube. The NMR tube was purged with argon gas for 5 min to prevent cysteine oxidation and then sealed with Teflon tape. The NMR acquisition of 2 dimensional ^{1}H/^{15}N-Heteronuclear Single Quantum Coherence (2D-HSQC) experiments for the 169-residue ^{15}N-OTU were acquired on a 600 MHz Varian INOVA spectrometer using a triple-resonance probehead temperature-calibrated with methanol at 20°C. The spectra were processed with NMRPipe[95] and visualized with Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco)[96, 97]

2.4 Backbone Assignments

Samples for the complete 3-dimensional backbone assignment of ^{15}N/^{13}C-labeled OTU were prepared by T. James (Department of Microbiology). Three-dimensional NMR spectra (HNHA,

HNCO, HNCACB, HNCACO, CBCACONH) were acquired on an 800 MHz Varian spectrometer at 25 °C equipped with a triple resonance Cold Probe calibrated with methanol[40], using standard Varian BioPack pulse sequences[61, 98-101]. The spectra were processed with NMRpipe[95]; they were linear predicted, zero-filled, multiplied by a cosine-squared bell function, and base-line corrected before Fourier transformation. ^{1}H and ^{15}N chemical shift referencing were done using the ^{1}H water signal at 293 K that resonates at 4.821 ppm relative to 2, 2-dimethyl-2-silapentane sulfonate[40]. Assignments were done manually by J. O'Neil using Sparky (T. D. Goddard and D. G. Kneller, SPARKY, University of California, San Francisco) and PINE-SPARKY with assistance from the Pine server[102, 103].

2.5 Selection and Assignment of Amino Acids for "*Unlabelling*"

The confirmation of missing peaks in the ^{15}N-OTU and 18 single residue unlabelled 2D-HSQC spectra were validated by comparing residue-unlabeled spectra with a uniformly-labelled ^{15}N OTU 2D-HSQC spectrum displaying all the amino acid peaks as a reference sample. The experiment was validated by dividing the peak height intensities from each amino acid from the uniformly labelled ^{15}N-OTU 2D-HSQC spectrum into the corresponding volume of each amino acid missing or weak peak in the unlabelled ^{15}N-OTU 2D-HSQC spectrum. In this experiment, glycine was neglected because it exhibits distinct $^{13}C^{\alpha}$ and ^{15}N shifts in spectral regions that can be easily identified by examining 2D and 3D spectra.

2.6 Secondary Structure Estimates using Chemical Shift Analysis

The average values for random coil backbone chemical shifts for individual amino acids were subtracted from observed ^{1}H, ^{13}C, and ^{15}N chemical shifts for OTU to determine the chemical shift differences from a random coil according to the method described in Schwarzinger *et al.*

(2001)[104] . Since H^N, H^α, C^α and C^β chemical shifts are very sensitive to the local amino acid sequence, the random coil values were corrected for the effects of the neighbouring residues, as per the following equation:

$$\delta_{corrected(i)} = \delta_{rc}(i)+\Delta\delta(i-1)+\Delta\delta(i+1)+\Delta\delta(i-2)+\Delta\delta(i+2)\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Where;

- $\delta_{corrected(i)}$ = corrected chemical shift difference at a position *i* for a residue in the sequence,
- $\delta_{rc}(i)$ = random coil chemical shift for residue (*i*) in question; and
- $\Delta\delta$ terms = correction factors for the two residues following and preceding residue (*i*)[105].

We used also a Secondary Structure Propensity (SSP)[105] method that integrates the chemical shifts from various nuclei to estimate the fraction α or β-structure in the OTU protein. This method weights the sensitivity of each nucleus to secondary structure. It is observed that the most important nuclei for this purpose are $^{13}C\alpha$, $^{13}C\beta$ and $^{1}H\alpha$. We used the software available at: http://pound.med.utoronto.ca.uml.idm.oclc.org/software.html to calculate the SSP scores and create the graphs. The algorithm also applies a correction for chemical shift referencing errors.

2.7 Secondary Structure Estimates using $^3J^{HNH\alpha}$ Coupling Constants

The ratio of the intensities of the cross-peaks and diagonal-peaks in a 3D HNHA experiment were used to calculate the $^3J^{HNH\alpha}$ coupling constants to a first approximation, under the assumption that the cross-peaks and diagonal-peaks line shapes are identical[60]. The following relation was used to obtain the $^3J^{HNH\alpha}$ coupling constants: (neglecting the relaxation effects)

$$I_{cross}/I_{diag} = -\tan^2 (^3J^{HNH\alpha}2\pi\delta_2)$$

Where I_{cross} and I_{diag} are the intensities of the cross- and diagonal peaks in the experiment; $2\delta_2$ is the ^1H_N magnetization dephasing period that was set to 12.5 ms. The results obtained will most likely be 5-10% underestimated because during the approximation does not take into account the longitudinal relaxation of the H^{α} proton during the $2\delta_2$ period. The sequence-corrected COIL values were subtracted from the above approximation for the $^3J^{HNH\alpha}$ coupling constants to obtain $\Delta^3J^{HNH\alpha}$ values[106].

2.8 NMR Relaxation Measurements

The ^{15}N-labelled OTU NMR relaxation data were collected on a Varian 600 MHz NMR spectrometer at the University of Manitoba and an 800 MHz NMR spectrometer at the University of Alberta (NANUC) using triple-resonance probes and BioPack pulse sequences at 25 °C[61, 98-101]. Pascal Mercier made the measurements at 800 MHz and Vu To made the measurements at 600 MHz. Measurement parameters for all of the NMR data are summarized in Table 1. For both 600 and 800 MHz measurements, the R_1 relaxation rates were obtained from a total of nine data sets using relaxation delays of 10, 80, 170, 270, 400, 550, 760, 1080 and 1800 ms. To measure R_2 relaxation rates with the *Carr-Purcell-Meiboom–Gill* sequence[107] a total of 10 data sets were acquired using relaxation delays of 10, 30, 50, 70, 90, 110, 130, 150, 170 and 190 ms. Post-acquisition delays of 4 s were used in both experiments whereas a 5 s relaxation delay was used for the NOE experiments. The steady-state $^1H/^{15}N$ NOE values were obtained from experiments with (I_{NOE}) and without (I_{noNOE}) saturation using equation 12 from Chapter 1 above.

Table 1. Acquisition parameters for the NMR experiments.

Experiment	Scans	Complex points	SW[^{1}H] (ppm)	SW[^{13}C] (ppm)	SW[^{15}N] (ppm)	Field (tesla)
^{1}H-^{15}N HSQC	32	1024	16 ppm		32 ppm	14.1
HNCO	8	1024×64×32	15 ppm	16 ppm	33 ppm	18.8
HN(CA)CO	8	1024×64×32	15 ppm	16 ppm	33 ppm	18.8
HNCACB	8	1024×80×32	15 ppm	80 ppm	33 ppm	18.8
CBCA(CO)NH	8	1024×80×32	15 ppm	80 ppm	33 ppm	18.8
HNHA	4	512×128×32	15/12.5		34 ppm	18.8
R_1	16	1024×64	16 ppm		32 ppm	14.1/ 18.8
R_2	16	1024×64	16 ppm		32 ppm	14.1/ 18.8
NOE	128	1024×64	16 ppm		32 ppm	14.1/ 18.8
CPMG	16	1024 ×64	16 ppm		32 ppm	14.1 T

The relationship between the rates of relaxation (R_x) and the NOE at the nuclear spin transition frequencies to the spectral density values were described in equations 1, 6 and 10 in Chapter 1.

Relaxation rates were obtained from fits done in the program RELAX[108]. The signal/noise ratios of the individual peaks were utilized to calculate the relaxation rate errors.

2.9 Reduced Spectral Density Mapping

Spectral density analysis was done using software generously provided by Dr. Leo Spyracopoulos (Department of Biochemistry, University of Alberta)[109]. Uncertainties in the calculated spectral densities were determined by Monte Carlo methods as described previously[110, 111]. The standard deviations from the NMR measurements were calculated 500 times to determine the error in the spectral density values and the normal distribution was generated by the Monte Carlo method. Mathematica™ was used for all calculations (Wolfram Research, Inc., 2004).

The simplified spectral density mapping described in 1.13 was refined by Farrow *et al*., (1995)[65]. They made the assumption of a Lorentzian Spectral Density in which $J(\omega)$ is proportional to $1/\omega^2$. With this assumption equations 1, 6, and 10 can be rewritten as follows and the reduced spectral densities extracted from the relaxation rate measurements:

$$R_1 = d[3J(\omega_N) + 7J(\beta_1\omega_H)] + cJ(\omega_N) \ldots\ldots\ldots\ldots (2)$$

$$R_2 = \frac{d}{2}\left[4J_{eff}(0) + 3J(\omega_N) + 13J(\beta_2\omega_H)\right] + \frac{c}{6}[4J_{eff}(0) + 3J(\omega_N)] \ldots(3)$$

$$NOE = 1 + \frac{\gamma_H}{\gamma_N}\frac{d}{R_1}[5J(\beta_3\omega_H)] \ldots\ldots\ldots(4)$$

where, β_1=0.921, β_2=*0.955,* and β_3=0.87.

The $J_{eff}(0)$ term is put in place to account for motion on the millisecond-microsecond timescale that can lead to erroneously high values of $J(0)$ and R_2 where:

$$J_{\text{eff}}(0) = J(0) + \Lambda R_{ex} \quad \text{(5)}$$

and the constant $$\Lambda = \frac{6}{3\,d^2 + 4c^2} \quad \text{(6)}$$

2.10 *Lipari-Szabo* Model-Free analysis

Lipari-Szabo Model-Free[69, 70] analysis was done using *Mathematica*™ scripts generously provided by Dr. Shaheen Shojania (Department of Biochemistry, University of British Columbia) as described previously[111, 112]. Uncertainties in the calculated Lipari-Szabo parameters were determined by Monte Carlo methods also as described previously.

The analysis of the relaxation data was carried out by optimizing all three *Lipari-Szabo* parameters for each individual residue and this approach provided a good fit to the NMR data[113]. The initial analyses were carried out by using equation 18 in Chapter 1, and then was further tested using additional models with variations of the *extended model-free* approach[71]. The correlation times for internal motions are separated into fast (τ_f) and slow (τ_s) components in the extended model-free method.

A series of models were tested that use both single field relaxation data and two-field data using the program Mathematica™[85]. Due to the number of parameters, two-field data were required for the extended model-free approach.

Monte Carlo simulations were used to estimate spectral density value errors in the Lipari-Szabo parameters and 500 points were calculated. Model 3 was tested and chosen on the basis of both R-factors (R_f)[114] and the Akaike information criterion (*AIC*)[115]. The AIC method applies a penalty (see below) to the models that increases linearly with complexity. A non-linear minimization of the χ^2 error function is carried out where:

$$\chi^2 = \sum_{i=1}^{N} \frac{(R_i - P_i)^2}{\sigma_i^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots (7)$$

and, i = an index over N, the total number of relaxation measurements (R_1, R_2, NOE), R_i is the measured value for each residue, P_i is the theoretical value for each residue and σ is the uncertainty in the relaxation constant. For *AIC,* the model chosen is the one where $\chi^2 + 2p$ is minimum with p being the number of optimized parameters.

The Models displaying low R_f values such as 0.15 or less were selected for estimating the Monte-Carlo error estimates. Model 3 was chosen on the basis of its R_f value (less than 0.15) and the mean *AIC* ± 1 standard deviation.

Table 2. *Lipari-Szabo* methods used for testing Model 3.

Model	Model type	Optimized parameters	Field
Model-3	*LS*	S_2, τ_c, τ_e, R_{ex}	14.1 T, 18.8 T

2.11 Relaxation Dispersion Measurements

The backbone ^{15}N CPMG relaxation dispersion experiments were acquired at the University of Manitoba 14.1 T Varian/Agilent instrument on 650 μL of 750 μM OTU protein. The two-dimensional relaxation dispersion experiments were done at 14.1 T at 25^{o}C with a CPMG relaxation time delay (τ_{cp}) of 40 ms and effective CPMG field strengths of 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1000 Hz[116, 117]. The peak heights were extracted from the data and analyzed using NESSY software[118]:

Relaxation dispersion peak heights were converted into effective relaxation rates in the usual way using the following equation[119]: $R_{2,eff} = -1/T_{relax} \cdot \ln (I/I_0)$, where I is the peak height measured at each CPMG field strength and I_0 is the peak height in a spectrum recorded without application of the CPMG pulse train. Relaxation dispersion data were then fit to models with no exchange (Model 1), 2-site exchange where the exchange rate $k_{ex} >> \Delta\omega$ (Model 2 – fast exchange) or 2-site exchange where the exchange rate $k_{ex} << \Delta\omega$ (Model 3 – slow exchange) using the software NESSY[118] and Akaike information theory to select the best model[115].

3 RESULTS

3.1 Protein Expression and Purification

The amino acid sequence of CCHFV OTU protease is shown in Figure 1[120]. The two N-terminal residues, Gly-Pro, belong to the linker between GST and OTU protease. The catalytic triad ***Cys[40]-His[151]-Asp[153]*** are highlighted in the sequence. At the C-terminus, residues from Thr[163] to Leu[169] are not visible in the X-ray diffraction structure[28] and are presumed disordered.

Gly-Pro-Met[1]-Asp[2]-Phe[3]-Leu[4]-Arg[5]-Ser[6]-Leu[7]-Asp[8]-Trp[9]-Thr[10]-Gln[11]-Val[12]-Ile[13]-Ala[14]-Gly[15]-Gln[16]-Tyr[17]-Val[18]-Ser[19]-Asn[20]-Pro[21]-Arg[22]-Phe[23]-Asn[24]-Ile[25]-Ser[26]-Asp[27]-Tyr[28]-Phe[29]-Glu[30]-Ile[31]-Val[32]-Arg[33]-Gln[34]-Pro[35]-Gly[36]-Asp[37]-Gly[38]-Asn[39]-***Cys***[40]-Phe[41]-Tyr[42]-His[43]-Ser[44]-Ile[45]-Ala[46]-Glu[47]-Leu[48]-Thr[49]-Met[50]-Pro[51]-Asn[52]-Lys[53]-Thr[54]-Asp[55]-His[56]-Ser[57]-Tyr[58]-His[59]-Tyr[60]-Ile[61]-Lys[62]-Arg[63]-Leu[64]-Thr[65]-Glu[66]-Ser[67]-Ala[68]-Ala[69]-Arg[70]-Lys[71]-Tyr[72]-Tyr[73]-Gln[74]-Glu[75]-Glu[76]-Pro[77]-Glu[78]-Ala[79]-Arg[80]-Leu[81]-Val[82]-Gly[83]-Leu[84]-Ser[85]-Leu[86]-Glu[87]-Asp[88]-Tyr[89]-Leu[90]-Lys[91]-Arg[92]-Met[93]-Leu[94]-Ser[95]-Asp[96]-Asn[97]-Glu[98]-Trp[99]-Gly[100]-Ser[101]-Thr[102]-Leu[103]-Glu[104]-Ala[105]-Ser[106]-Met[107]-Leu[108]-Ala[109]-Lys[110]-Glu[111]-Met[112]-Gly[113]-Ile[114]-Thr[115]-Ile[116]-Ile[117]-Ile[118]-Trp[119]-Thr[120]-Val[121]-Ala[122]-Ala[123]-Ser[124]-Asp[125]-Glu[126]-Val[127]-Glu[128]-Ala[129]-Gly[130]-Ile[131]-Lys[132]-Phe[133]-Gly[134]-Asp[135]-Gly[136]-Asp[137]-Val[138]-Phe[139]-Thr[140]-Ala[141]-Val[142]-Asn[143]-Leu[144]-Leu[145]-His[146]-Ser[147]-Gly[148]-Gln[149]-Thr[150]-***His[151]***-Phe[152]-***Asp[153]***-Ala[154]-Leu[155]-Arg[156]-Ile[157]-Leu[158]-Pro[159]-Gln[160]-Phe[161]-Glu[162]-Thr[163]-Asp[164]-Thr[165]-Arg[166]-Glu[167]-Ala[168]-Leu[169]

Figure 1. Amino acid sequence of the CCHFV OTU protease[120]. The GST-affinity tag residues Gly-Pro are shown in bold at the N-terminal of the CCHFV OTU protease sequence giving rise to a 171-residue OTU protein.

The growth of the *E. coli* BL21-Gold (DE3) cells containing the pGEX6P vector in 500 millilitres of minimal medium overnight typically yielded about 6 grams of cells (wet weight). The purification of the fusion protein (GST-OTU) was carried out on a GST Bind^{TM} resin (Novagen) – a glutathione matrix column. This column has a strong affinity towards Glutathione-S-Transferase-containing proteins. The HRV 3c protease cleaved at the recognition sequence site (Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro) between GST and the CCHFV OTU protease leaving behind Gly-Pro at the N-terminal region of OTU. OTU protease was separated from uncleaved fusion protein and HRV-3c protease by size-exclusion chromatography as shown in Figure 2. Size-exclusion Fast Protein Liquid Chromatography (FPLC) is a high-performance automated chromatography technique that isolates and purifies globular proteins on the basis of their hydrodynamic volume[121]. This chromatography technique can also be used for estimating the molecular mass of proteins because the elution volume of a protein is linearly correlated to its molecular weight[122]. The typical final yield of soluble CCHFV OTU protease after spin concentrating to 600 μL was usually between 600-650 μM. This amounts to about 13 mg of OTU per litre of culture medium.

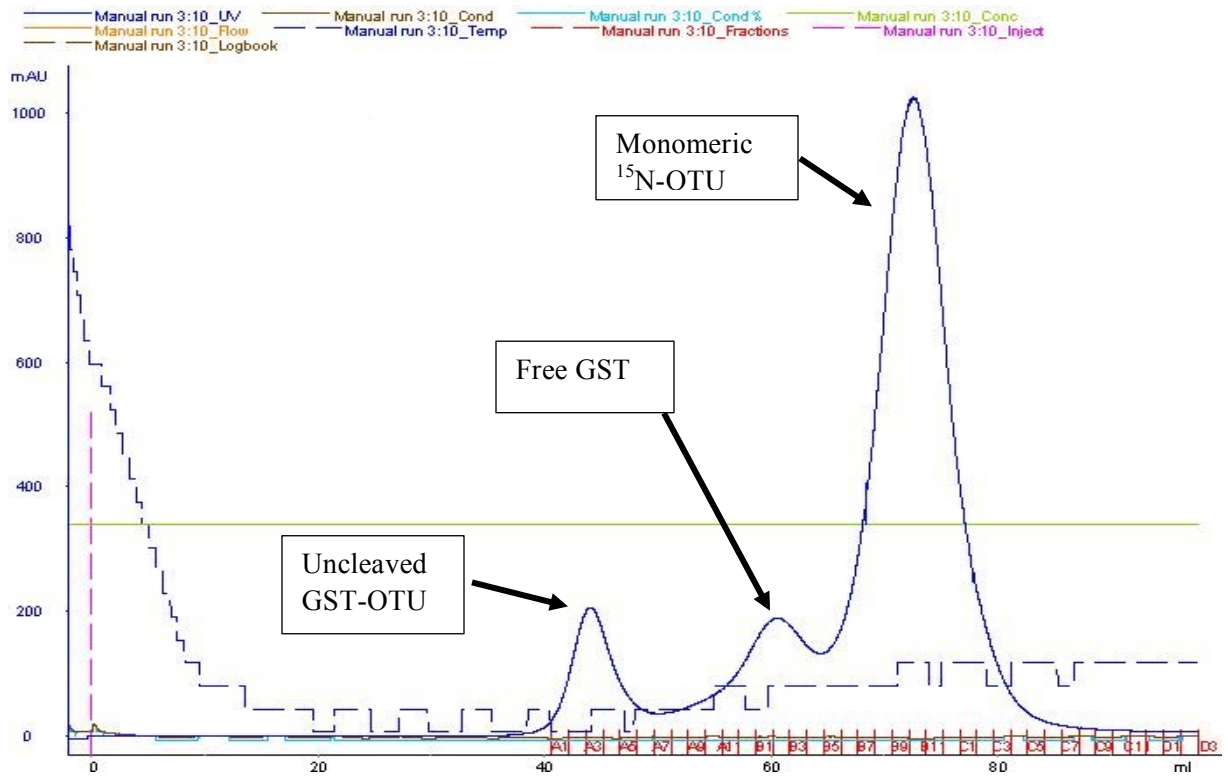


Figure 2. AKTA-FPLC elution profile for the separation of CCHFV OTU protease from the fusion protein and HRV3c protease. The large peak between 70-80 min giving an absorbance of around 1100 mAU corresponds to cleaved GST-free ^{15}N-CCHFV OTU protease (18 kDa). The other two small peaks around 42 min and 60 min correspond to free GST (25 kDa, dimer = 50 kDa) and uncleaved fusion protein (43 kDa, dimer = 86 kDa), respectively.

SDS-PAGE was performed on the purified protein and elution fractions from the S75 chromatography column to confirm the purity of the OTU protein. Figure 3 shows the SDS-PAGE electrophoregram of the S75 elution fractions of CCHFV OTU protease after cleaving its Glutathione-S-transferase tag by HRV 3c protease. The fractions from lanes 7-10 show single bands corresponding to a molecular weight of about 20 kDa, suggesting that highly purified CCHFV OTU protease has been prepared free of GST and the fusion protein. The bands

observed from lanes 4-6 indicate free GST corresponding to mass of about 25 kDa and the band in lane 3 at about 40 kDa likely corresponds to the uncleaved fusion protein.

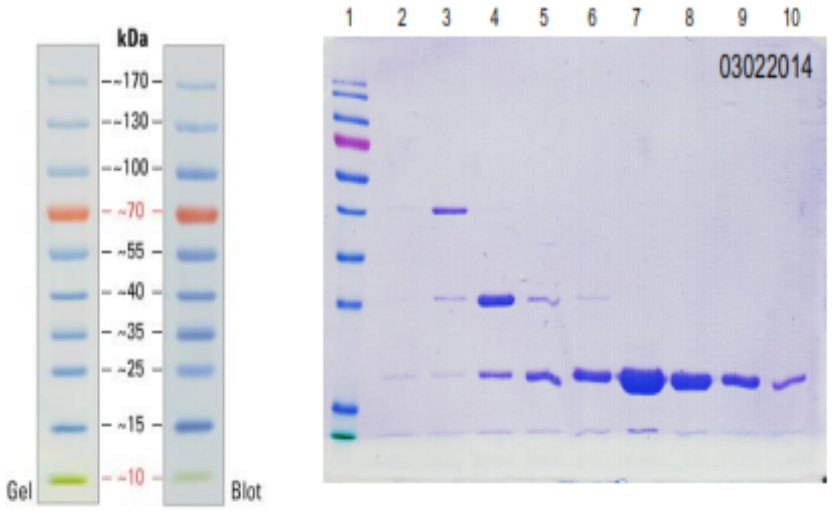


Figure 3. SDS-PAGE denaturing gel of GST-tagged CCHFV OTU protease following cleavage and purification by the FPLC.

3.2. NMR Spectroscopy and Resonance Assignments

The ^{1}H-^{15}N 2D-HSQC CCHFV OTU protease spectrum in Figure 4, shows amide resonances of the backbones as well as the side-chains of Asn and Gln. The spectrum shows CCHFV OTU protease cross-peak resonances spread out across the spectrum displaying the typical pattern of a folded protein. Figure 4 also indicates that nearly every backbone amide resonance was assigned. The complete backbone assigned resonances are shown in Table 1 in the appendix section. The lone Cys-40 was assigned based upon a ^{1}H-^{15}N 2D-HSQC spectrum of a C40A mutant (not shown) and the triple-resonance spectra (see below). The NMR resonances in the 2D HSQC spectrum of CCHFV OTU in Figure 4 show a wide array of peak widths spanning from extremely narrow intense lines in the C-terminus region of the protease sequence (e.g. Thr^{165})

indicating backbone disorder and flexibility on the ps-ns timescales to very broad lines near the active site triad (Cys^{40}-His^{151}-Asp^{153}) and elsewhere suggesting conformational fluctuations on the ms-μs timescales[123] (see also Figure 5). The weaker resonances in the spectrum may arise from slow conformational exchange on the ms-s timescale. An intriguing aspect of the spectrum is shown by the doubling of resonances for residues Gly^{15}, Gln^{16}, Phe^{29}, Ile^{31}, Val^{32}, His^{59}, Asp^{96}, Gly^{100}, Trp^{119}, Thr^{120}, Val^{121}, Ser^{124}, and Asn^{143} suggesting slow conformational exchange between two conformers. Figure 4, also shows multiple cross-peak intensities for several residues such as Gly^{136}, Leu^{48}, and Ile^{117}. Along with the weak peak intensity of some amino acid resonances the peak multiplicity suggests conformational exchange on the ms-s time-scales. The most challenging areas during the backbone assignment of the protease were around residues Glu^{98}, Trp^{99}, and Phe^{152} that all remain unassigned. Dilution of the protein did not eliminate these peaks as might be expected if they arise from dimer formation or aggregation.

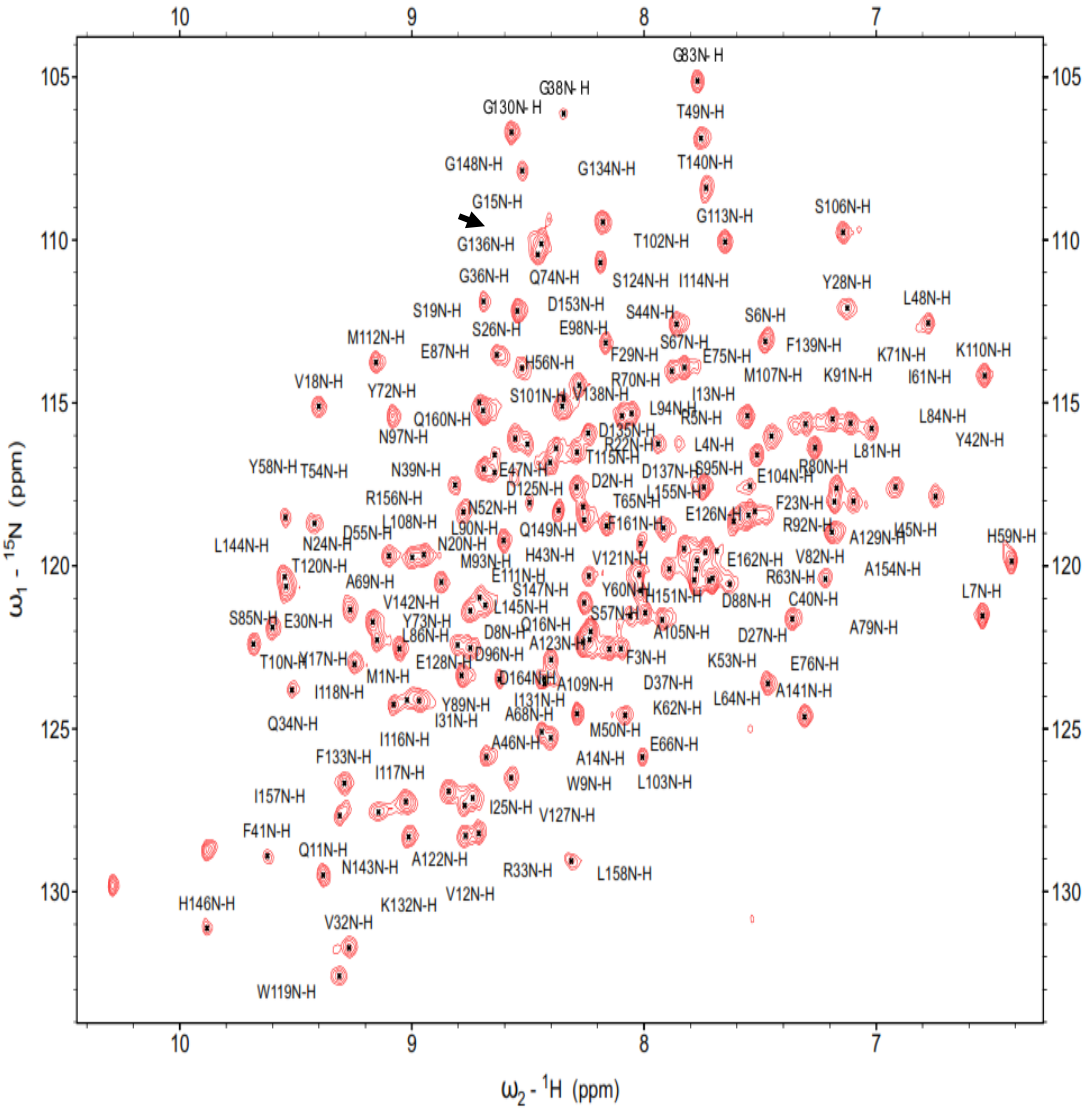


Figure 4. The assigned amide-backbone resonances of ^{15}N-CCHFV OTU in a ^{1}H-^{15}N-HSQC spectrum at pH 6.5 and 25 °C on a Varian INVOVA 600 MHz spectrometer at the University of Manitoba.

Figure 5 shows cross-peak intensities for CCHFV OTU protease from a ^{1}H-^{15}N-HSQC spectrum. A flexible disordered C-terminus is suggested by the narrowest and most intense lines from Phe^{161}-Leu^{169}. This agrees with the crystal structure that lacks electron density for residues Thr^{163} – Leu^{169} presumably owing to static disorder in the crystal[28]. Residues Asn^{52}, Gly^{100}, and Ala^{123} display moderately strong peaks suggesting some degree of disorder there. The residues spanning from Gly^{38}-Phe^{41}, and near Asn^{97}, Trp^{119} and Thr^{150} show the weakest peaks in the spectrum. For a detailed description of the possible peak intensity origins in protein NMR spectra see Park *et al*., (2011)[124].

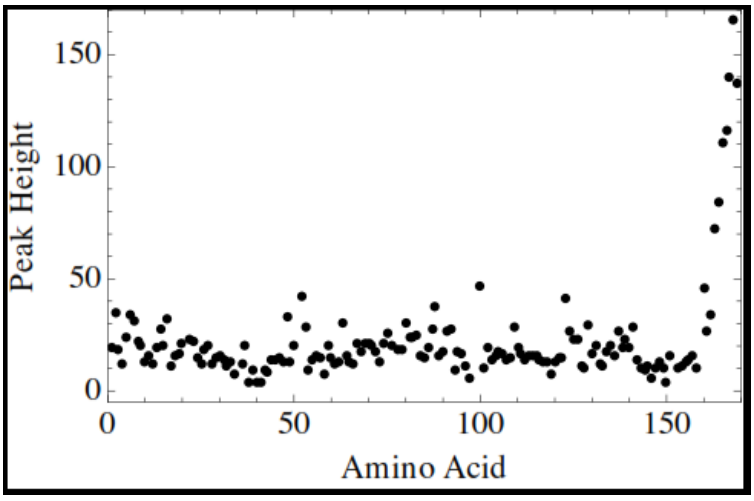


Figure 5. The backbone amide-resonances relative intensities from a ^{15}N-CCHFV OTU protease ^{1}H-^{15}N-HSQC spectrum at pH 6.5 and 25°C on a Varian INVOVA 600 MHz spectrometer at the University of Manitoba.

3.3 Unlabelling experiments and 3D assignment

Sequential assignments of the ^{1}H, ^{15}N, C', C^{α}, H^{α}, and C^{β} resonances of the $^{13}C/^{15}N$-labelled CCHFV OTU protease were done as described in the Materials and Methods section of the thesis. In the three-dimensional experiments, neighbouring amino acids were sequentially interlinked along the polypeptide chain by mapping the resonances onto the primary sequence. These assignments were done with 3D heteronuclear triple resonance experiments that connect the atoms by utilizing one- and two-bond scalar couplings[125]. Figures 6(A) illustrates an example of an HN(CA)CO experiment for the backbone assignments for amino acid residues Phe^{3}-Leu^{4}-Arg^{5}, Ser^{6}, Leu^{7}, Asp^{8}, and Trp^{9}.

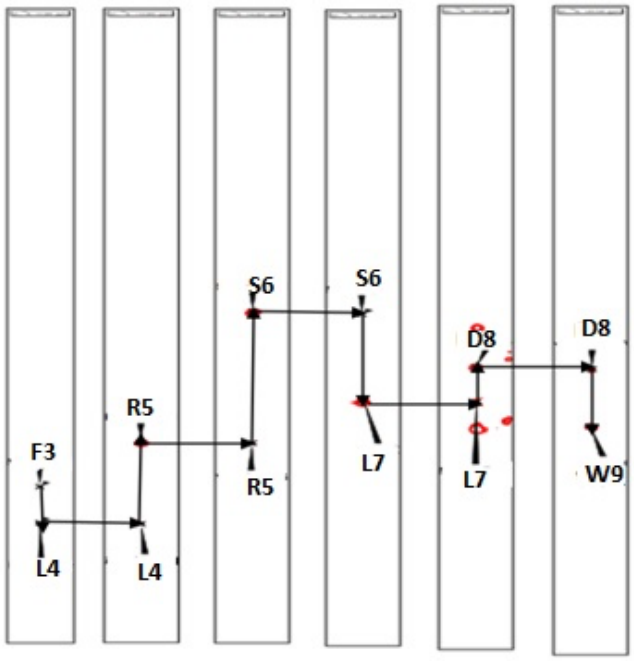


Figure 6 Amide-detected heteronuclear NMR backbone assignment of CCHFV OTU protease showing strip-plots extracted from the three dimensional HN(CA)CO experiment showing inter- and intra-residual correlations between HN(i) and N(i) with C'(i) and C'(i-1) resonances betweenresidues Phe^{3}-Trp^{9}. Measurements were done at pH 6.5 and 25°C.

After assigning the backbone resonances, about 10-15% of the amides remained ambiguous. We wanted to confirm and extend the assignments by selective unlabelling of specific amino acids[41]. Although more than one amino acid can be chosen to be selectively unlabelled in a single protein sample, identification and assignment are contingent upon connections to other resonances in triple resonance experiments ($^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts) and their percentage in the polypeptide chain. Therefore, to avoid ambiguous assignments and aid direct identification, single ^{15}N-labelled amino acids were selectively unlabelled in each CCHFV OTU protease sample. The amino acid proline was not selected for unlabelling experiments since it does not contain an amide proton and therefore does not appear in the three-dimensional and two-dimensional HN-detected experiments[41]. Glycine was not unlabelled because its distinct $^{13}C^{\alpha}$ and ^{15}N shifts are in spectral regions that can be easily identified by examining 2D and 3D spectra[123].

In cases where the resonances have a high signal-to-noise ratio and good separation from other peaks, identification of unlabelled peaks can be done by visual inspection as shown in Figure 7 where alanine has been unlabelled. It can also be done by directly subtracting an unlabelled spectrum from the scaled uniformly labelled sample reference spectrum, where spectral overlap is evident. However, this approach has two main drawbacks: imperfect cancelling of peaks because of residual ^{15}N-labelling and difficulties associated with low signal-to-noise (S/N) ratios. Therefore, where peaks were overlapped we decided to divide the peak height intensities from each unlabelled amino acid resonance by the corresponding peak height in the uniformly labelled ^{15}N-CCHFV OTU protease HSQC spectrum. This approach helped identify many unlabelled resonances that were overlapped by labelled resonances.

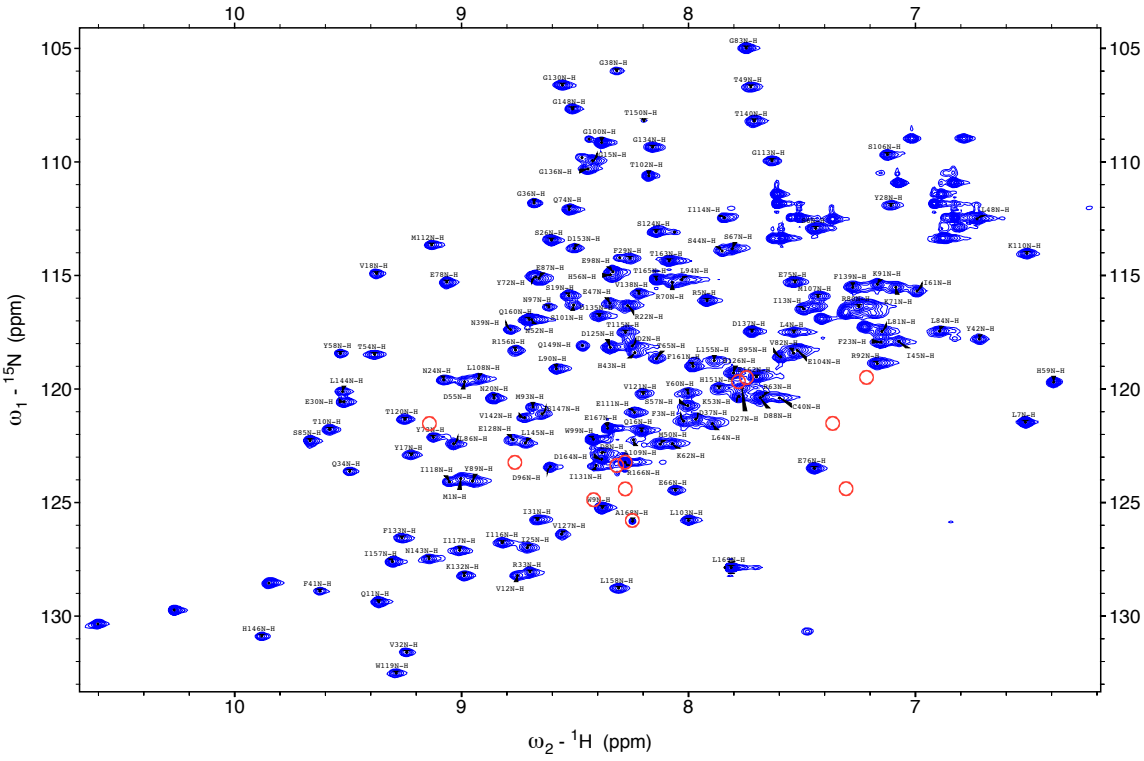


Figure 7. Two-dimensional ^{1}H-^{15}N HSQC spectrum of alanine-unlabelled CCHFV OTU protease at pH 6.5 and 25°C. The 12 missing Ala amino acid resonances in the CCHFV OTU protease spectrum are circled in red and were identified by comparison with a standard uniformly labelled ^{15}N-CCHFV OTU HSQC spectrum. (See Figure 4.)

3.4. Chemical Shift Analysis and Coupling Constants

The NMR chemical shift can be extremely sensitive to conformational changes and interpretation of the secondary structure is possible by comparing protein chemical shifts to random coil values

corrected for local sequence effects using the neighbour-corrected Intrinsically Disordered Protein Library[126, 127]. Shown in Figure 8 are chemical shift difference plots from random coil values for each nucleus for the CCHFV OTU protease. The plots also indicate the secondary structure information contained in the shifts from the random coil values. Examination of the plots shows that the chemical shifts are in fair to good agreement with the X-ray diffraction structures that have been published[28]. In general, the $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ resonances (Figure 8C-F) are known to be the most accurate indicators of secondary structure[128].

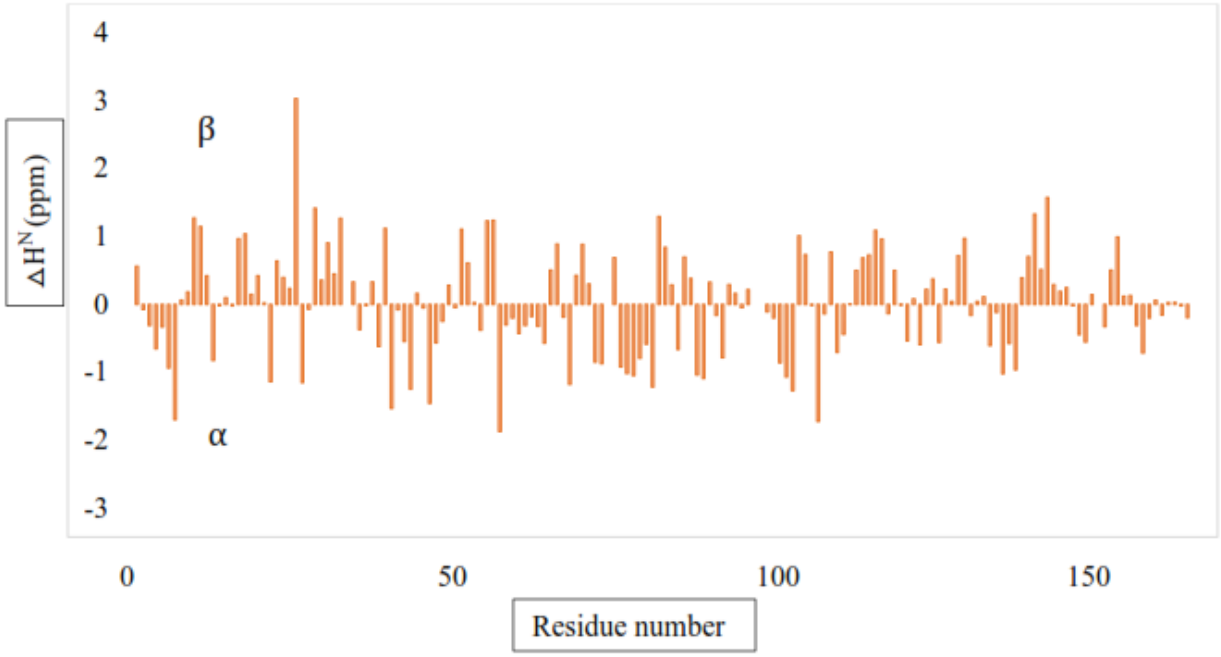


Figure 8A. Chemical shift difference plot from random coil values of ΔH^{N} (ppm) in CCHFV OTU protease adjusted for sequence dependence. Positive shifts suggest β-sheet conformation whereas negative shifts suggest α-helix.

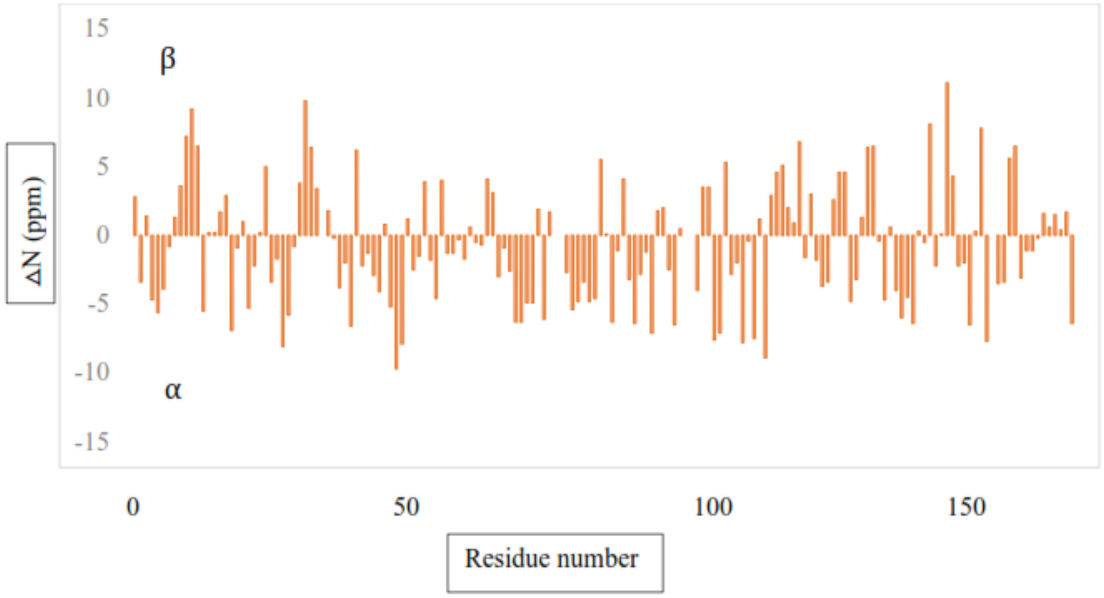


Figure 8B. Chemical shift difference plot from random coil values of $\Delta^{15}N$ (ppm) in CCHFV OTU protease adjusted for sequence dependence. Positive shifts suggest β-sheet conformation whereas negative shifts suggest α-helix conformation.

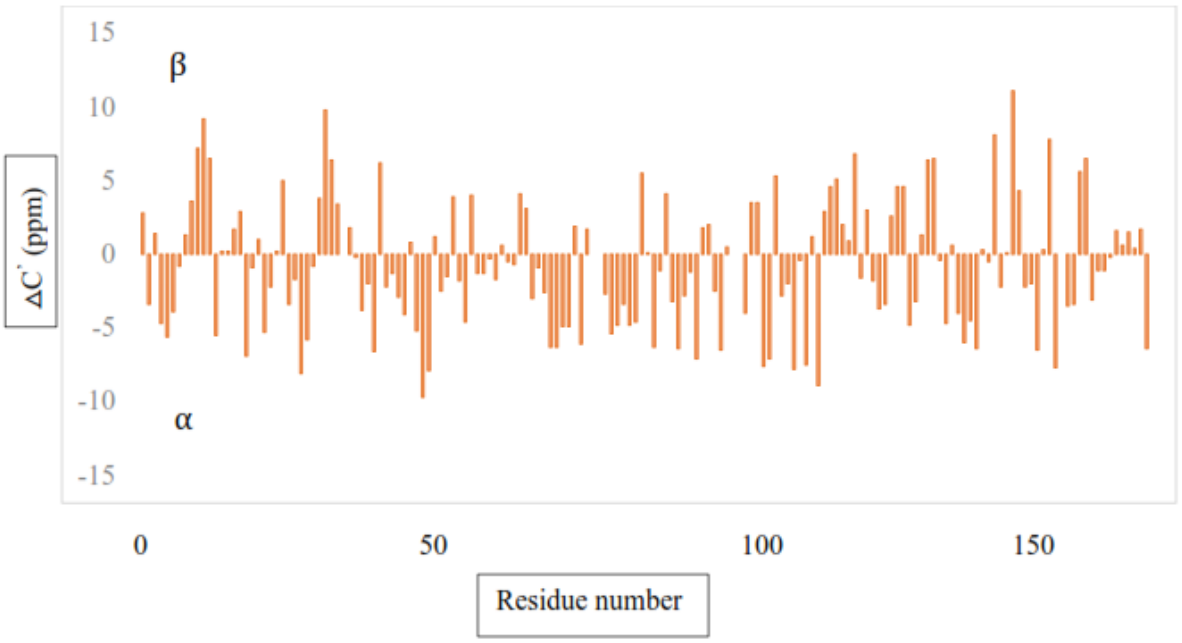


Figure 8C. Chemical shift difference plot from random coil values of $\Delta^{13}C'$ (ppm) in CCHFV OTU protease adjusted for sequence dependence. Positive shifts suggest α-helix conformation whereas negative shifts suggest β-sheet conformation.

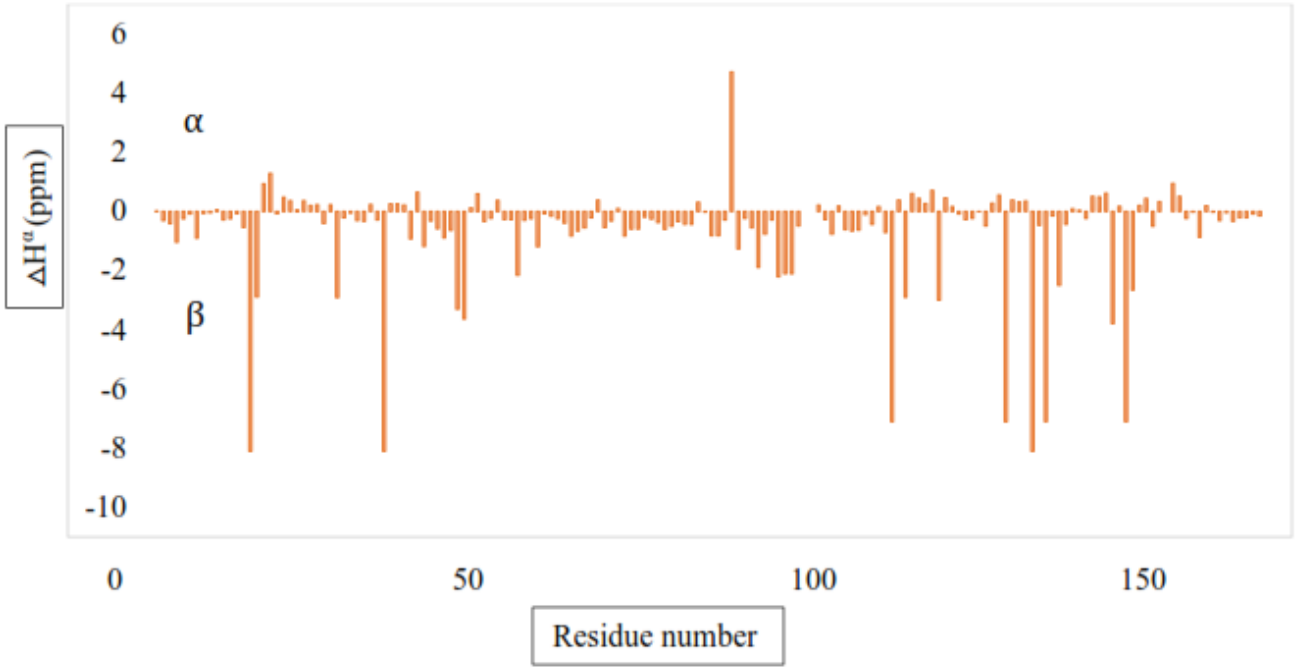


Figure 8D. Chemical shift difference plot from random coil values of ΔH^{α} (ppm) in CCHFV OTU protease adjusted for sequence dependence. Positive shifts suggest α-helix conformation whereas negative shifts suggest β-sheet conformation.

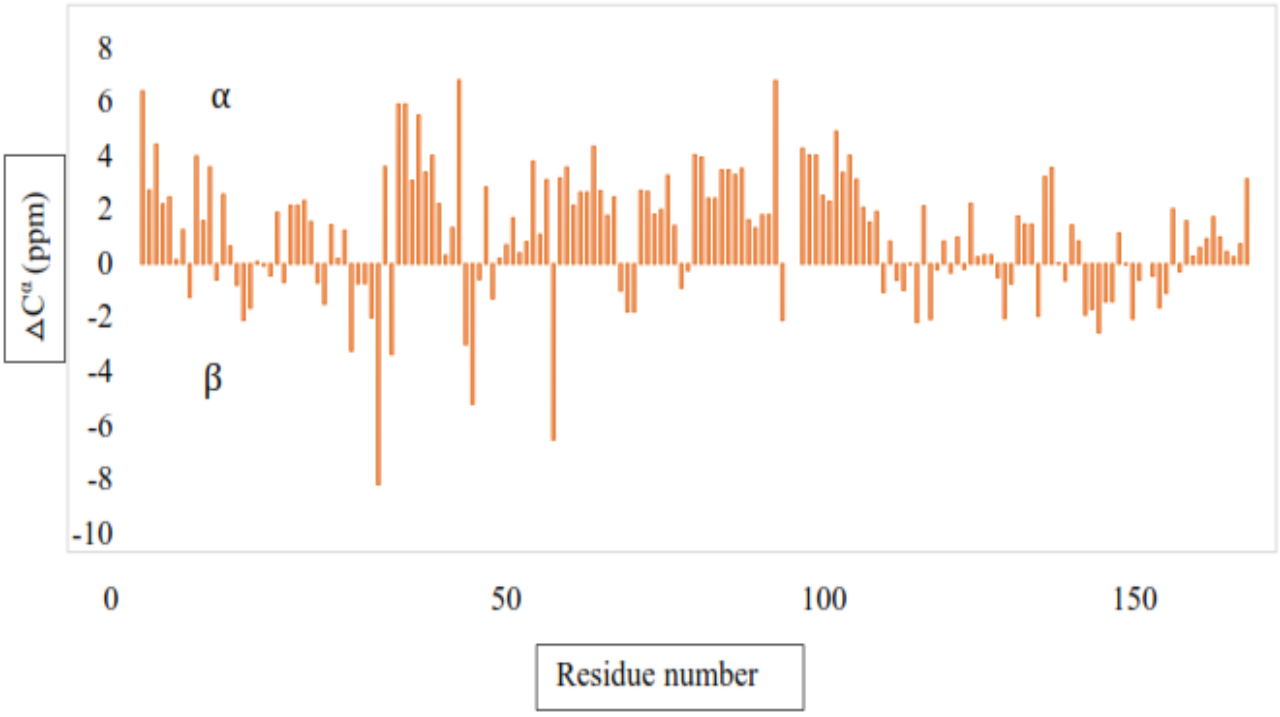


Figure 8E. Chemical shift difference plot from random coil values of $\Delta^{13}C^{\alpha}$ (ppm) in CCHFV OTU protease adjusted for sequence dependence. Positive shifts suggest α-helix conformation whereas negative shifts suggest β-sheet.

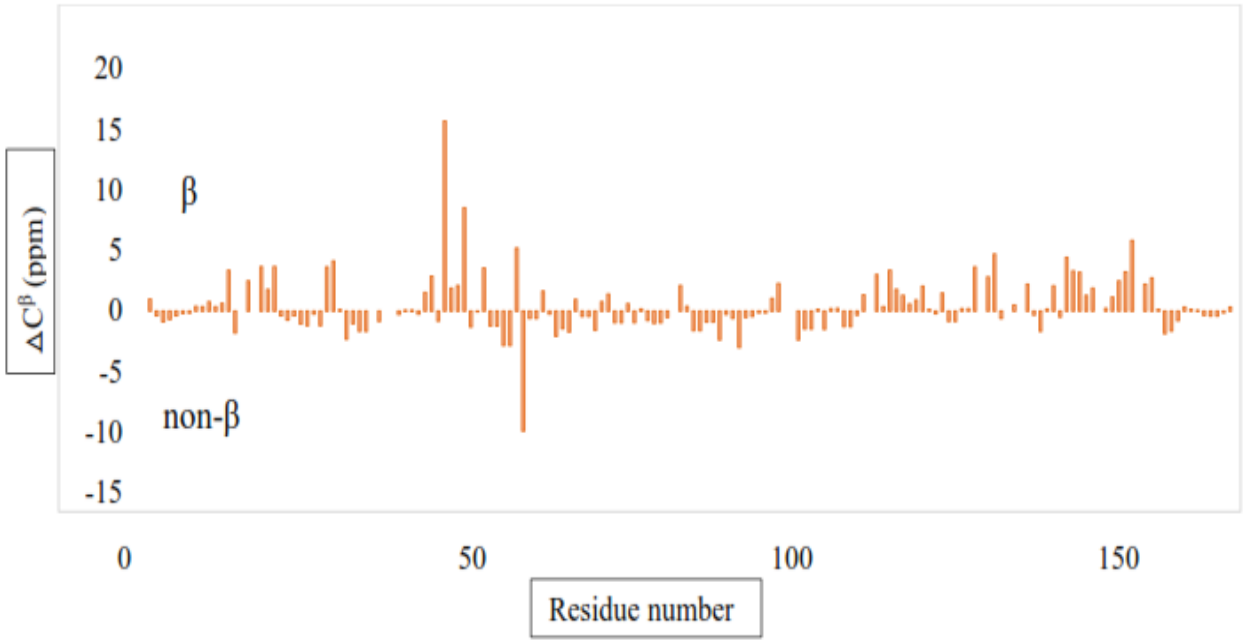


Figure 8F. Chemical shift difference plot from random coil values of $\Delta^{13}C^{\beta}$ (ppm) in CCHFV OTU protease. No correction for sequence dependence has been applied because these are not available from the literature. Positive shifts suggest β-sheet conformation whereas negative shifts suggest non β-sheet conformation.

Several algorithms are available that combine the secondary-structure information content of several nuclear chemical shift differences from random coil values to yield more accurate measures of secondary structure. The Chemical Shift Index (CSI)[129] consensus method combines C', C^{α}, C^{β}, and H^{α} chemical shift differences from random coil values, however these consensus CSI calculations are not corrected for local sequence effects. The results applying this algorithm to CCHFV OTU protease data collected at pH 6.5 at 25°C are illustrated in Figure 9. Note that at least four consecutive residues are needed for identifying α-helix, since 1 turn of α-helix consists of 3.6 amino acids[130]. The algorithm can identify 9 different types of secondary

structure including several types of turn as shown in the figure. The positive (red) and negative (shades of blue) bars indicate α-helix and β-sheet structure. In general, the results shown in Figure 9 agree with the x-ray diffraction structure of OTU protease although the secondary structure boundaries are not well very defined (see below).

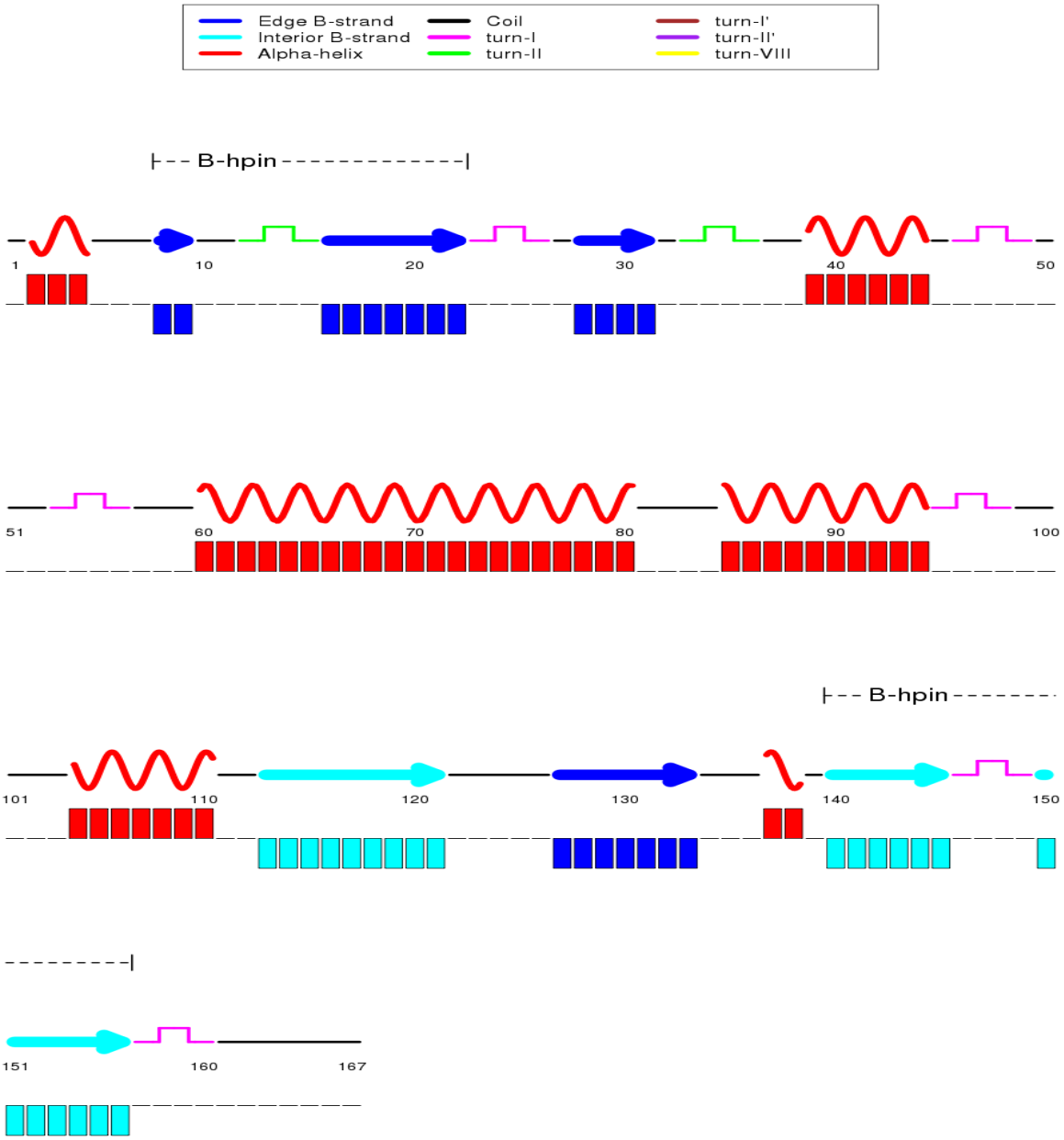


Figure 9. The backbone secondary structure calculated for CCHFV OTU protease by the Chemical Shift Indexing (CSI) 3.0 webserver to identify α-helix, β-strands, β-turns, coil regions, β-hairpins, interior β-strands and edge β-strands[131].

Figure 10 shows the results from a different algorithm called Secondary Structure Propensity (SSP) applied to the CCHFV OTU protease. The secondary structure propensities were calculated as described in Materials and Methods section 2.6. The algorithm uses a weighted average of the C^{α}, C^{β} and H^{a} chemical shift differences to calculate a secondary structure propensity. The SSP scores are interpreted as follows: a score of 1 indicates α-structure, a score of -1 indicates β-structure, a score near 0 indicates a disordered region and a numerical value of 0.5 suggests a marginally stable α-helical ensemble. This method provides a quantitative determination of the secondary structure propensity at a given position. Figure 10 shows the SSP scores of CCHFV OTU protease and they[28] are in remarkably good agreement with the published X-ray diffraction structure. (See Table 1 below). A region of minor disagreement is Thr^{10}-Asn^{20} that is composed of two short β-strands but shows only weak β-propensity in the NMR data. At the C-terminus the SSP scores suggest weak α-helix propensity whereas in the crystal structure[28] the lack of electron density suggest this region is disordered and this agrees with other NMR measurements below. Also, near residue 35 is a region of weak secondary structure propensity suggesting dynamic flexibility there. The excellent agreement between the chemical shift measurements and X-ray diffraction structure shows that the overall structure and secondary structure boundaries of OTU protease measured by NMR are identical to those measured in the crystal[28] and allowed us to utilize the crystal structure as a model for the solution dynamics studies. These results also suggest that the chemical shift assignments are mostly correct.

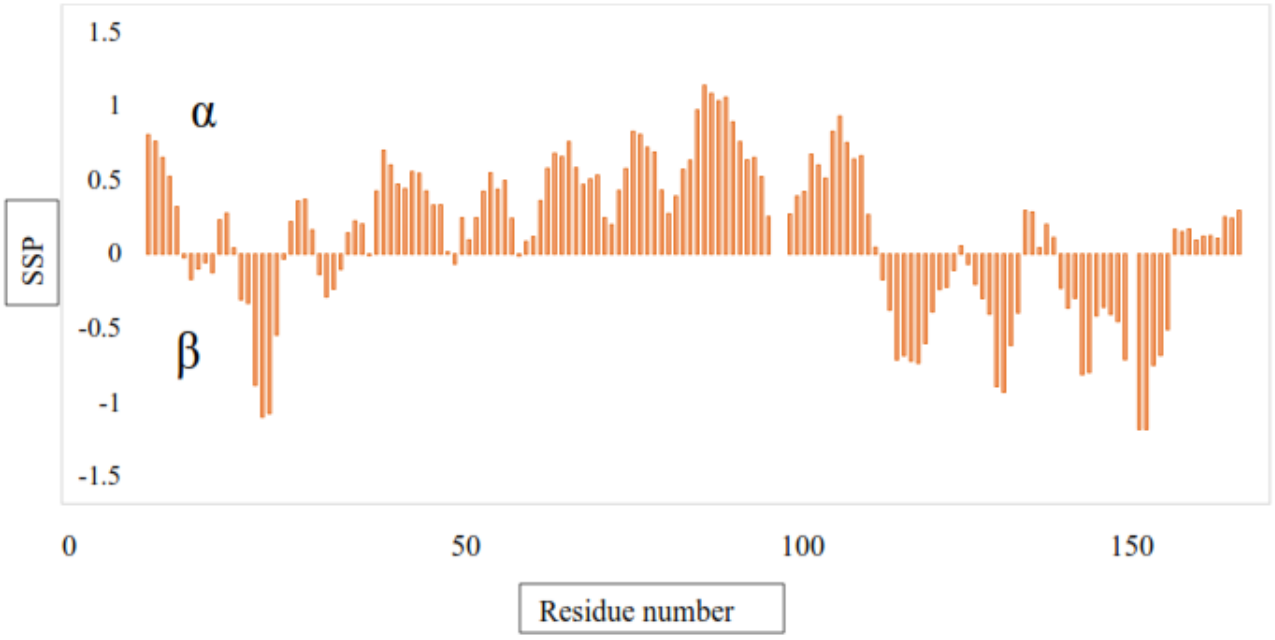


Figure 10. The Secondary Structure Propensity chemical shifts plot (SSP) ($\Delta\delta C^{\alpha}$-$\Delta\delta C^{\beta}$) of CCHFV OTU protease indicated by positive values (α-helix propensity) and negative values (β-sheet propensity).

Secondary Structure	Start	End
HELIX 1	ASP 2	ARG 5
HELIX 2	ILE 25	ASP 27
HELIX 3	CYS 40	THR 49
HELIX 4	SER 57	TYR 72
HELIX 5	PRO 77	VAL 82
HELIX 6	LEU 86	LEU 94
HELIX 7	THR 102	MSE 112
HELIX 8	PRO 159	PHE 161
SHEET 1	THR 10	ILE 13
SHEET 2	GLN 16	SER 19
SHEET 3	VAL 127	PHE 133
SHEET 4	THR 115	VAL 121
SHEET 5	ALA 141	SER 147
SHEET 6	HIS 151	ILE 157
SHEET 7	PHE 29	VAL 32

Table 1. Secondary structure boundaries as defined by the X-ray diffraction structure of OTU protease bound to ubiquitin. PDB ID: 3PT2[28].

The $^{3}J_{HNHA}$ coupling constants can also be used to decipher the secondary structure of the protease in solutions[132]. Figure 11 shows $^{3}J_{HNHA}$ coupling constants for 128 CCHFV OTU protease residues and the corresponding secondary structure in the crystal[28]. Although the measured values are expected to be 5-10 % lower than the actual values[133] in general there is an excellent agreement with the crystal structure[28]. (Because of faster relaxation of the antiphase magnetization the cross-peaks have lower intensities and artificially low coupling constants[133]).

In the figure, the residues in α-helix and β-sheet from the crystal structure[28] are indicated by horizontal lines at 1 Hz and 11 Hz. Residues in an α-helix are usually indicated by coupling constants smaller than 6 Hz and β-sheets are suggested by coupling constants larger than 8 Hz. The residues corresponding to coupling constants between 6 and 8 Hz suggest flexible residues. The dashed lines drawn at 5 Hz and 7 Hz indicate secondary structure thresholds. The results from the coupling constants agree fairly well with the crystal structure suggesting N-and C-terminal β-sheets and a central α-helical region. Fifty-one residues in the protease amino acid sequence were found to exhibit coupling constants in the range of 5-7 Hz indicating a flexible backbone. Some of these residues are found to be situated near the loop regions but many are located in secondary structure regions of the CCHFV OTU protease suggesting increased dynamics there. Cross-peaks were not observed for some residues and hence do not appear in the plot. Some of these residues are situated near residues Cys^{40}, Trp^{99}, and His^{151} implying slow conformational exchange in those regions.

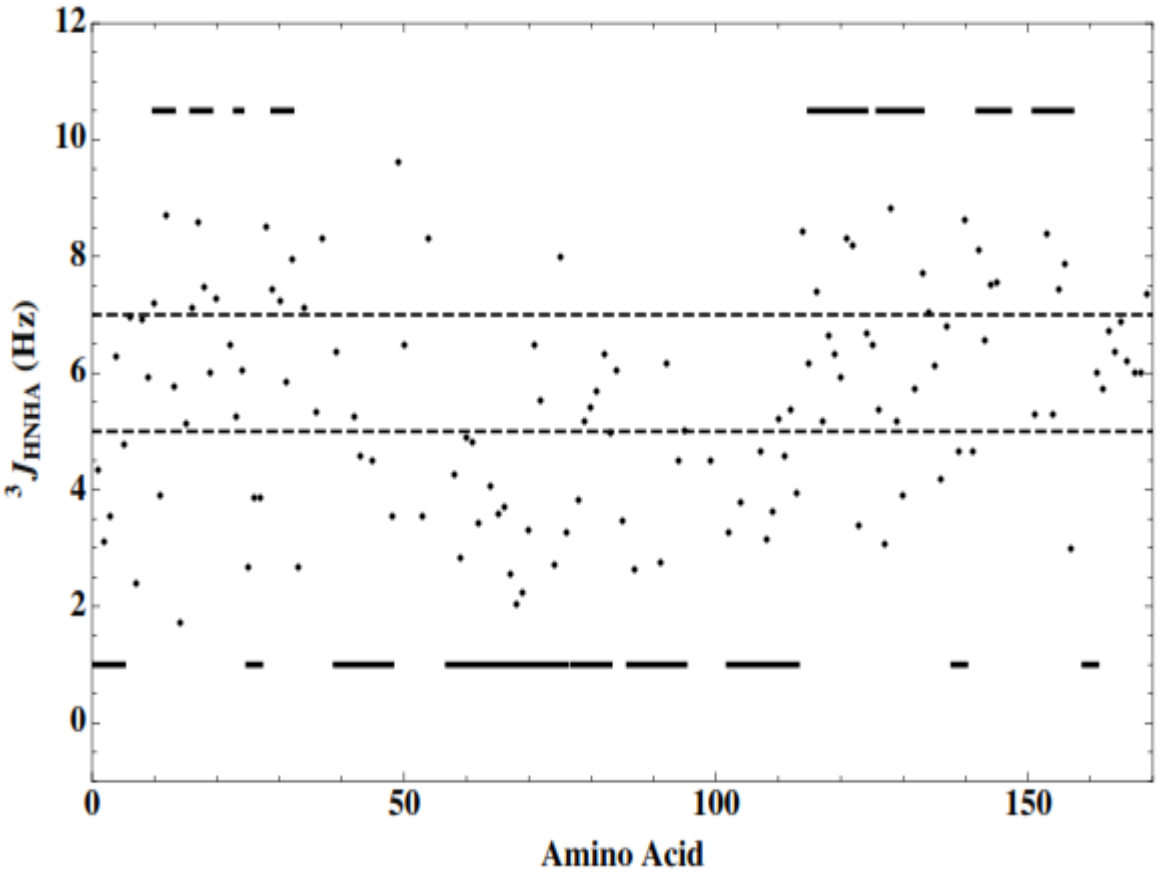


Figure 11. The $^3J_{HNHA}$ coupling constants for the CCHFV OTU protease were determined from the ratios of cross-peaks to diagonal peaks in a 3-D HNHA experiment measured at 800 MHz (18.8 T). The dashed lines indicate the thresholds for α-helix (below) and β-sheet (above). Between the lines are residues most likely in disordered or flexible segments. The solid bars indicated the regions of secondary structure in the crystal structure[28].

3.5. NMR Relaxation Dynamics

NMR relaxation measurements were used to probe the backbone dynamics of CCHFV OTU protease in solution on different time-scales. The studies are aimed at a deeper understanding of the molecular dynamics of the CCHFV OTU protease by measurements of the ^{15}N-longitudinal relaxation (T_1), ^{15}N-transverse relaxation (T_2), and heteronuclear ^{1}H-^{15}N *NOE* measured at 14.1 and 18.8 T for observable resonances**.** Figures 12A and 12B show sample heteronuclear NOE spectra for saturation (*NOE*) and no saturation (*noNOE)* for CCHFV OTU protease at pH 6.5 and 25 °C at 14.1 Tesla**.** The ratio $(I_{NOE}-I_{noNOE})/I_{noNOE}$ was used to obtain the heteronuclear steady-state ^{1}H-^{15}N *NOE* values that are plotted in Figures 13A and 14A. The observed maximum, minimum and average *NOE* values at 14.1 Tesla were 1.3, 0.1 and 0.816. The positive *NOE* values indicate more restricted dynamics on the ps-ns timescales for the CCHFV OTU protein. The end of the CCHFV OTU protease at the C-terminus region spanning from Thr^{163}-Leu^{169} showed lower *NOE* values indicating unrestricted faster dynamics at 14.1 T. This strongly meshes with the X-ray crystal structure which shows a lack of electron density from Thr^{163}-Leu^{169} likely because of static disorder[28]. There are additional low *NOEs* around Thr^{49} and Gly^{100} suggesting possible backbone flexibility in these regions of the CCHFV OTU protease as well.

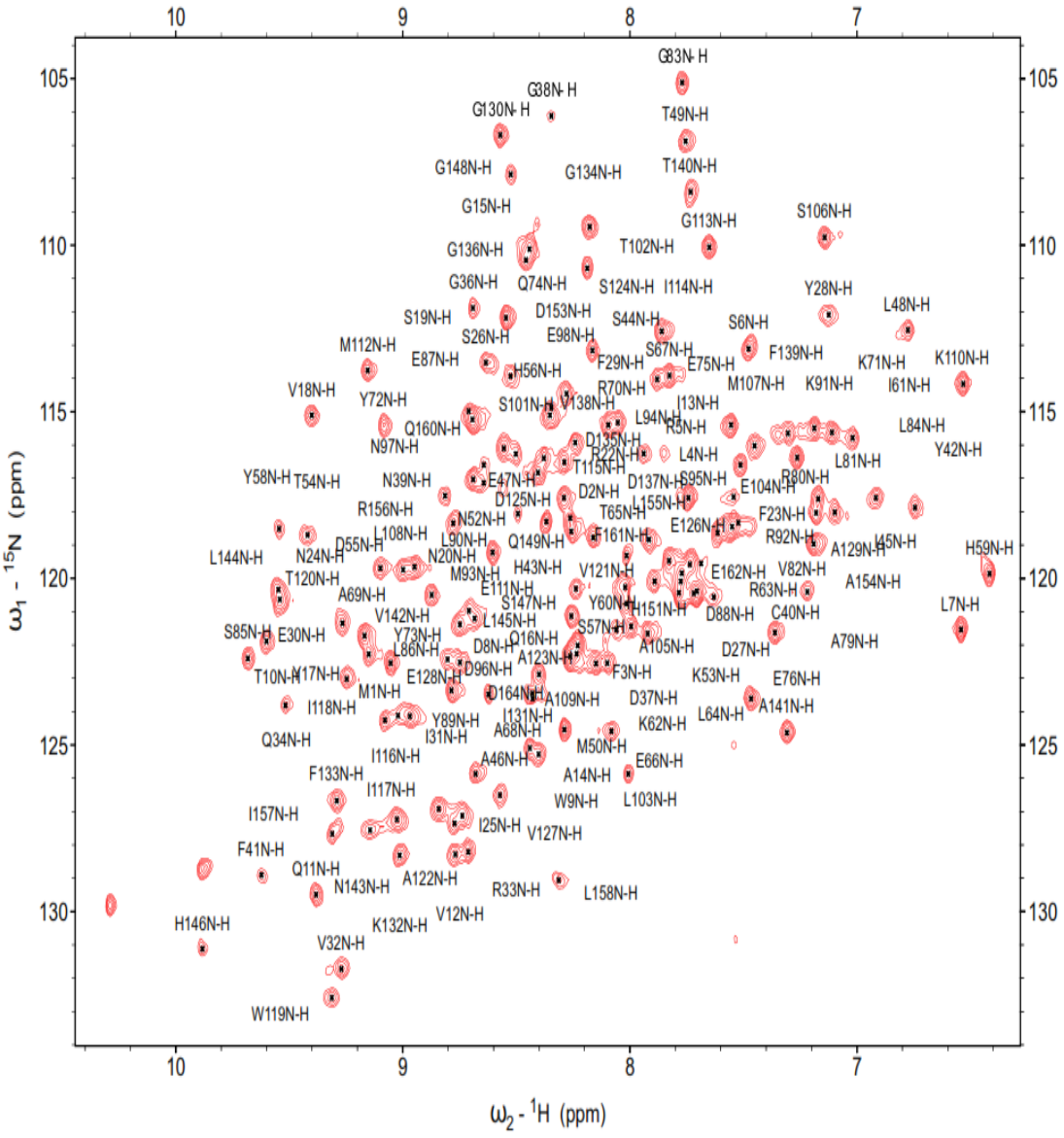


Figure 12A. The steady-state heteronuclear NOE spectrum of ^{15}N-labeled CCHFV OTU protease at pH 6.5 and 25 °C acquired on a Varian INOVA 600 MHz (14.1 T) spectrometer with a pre-saturation (NOE) period of 5s.

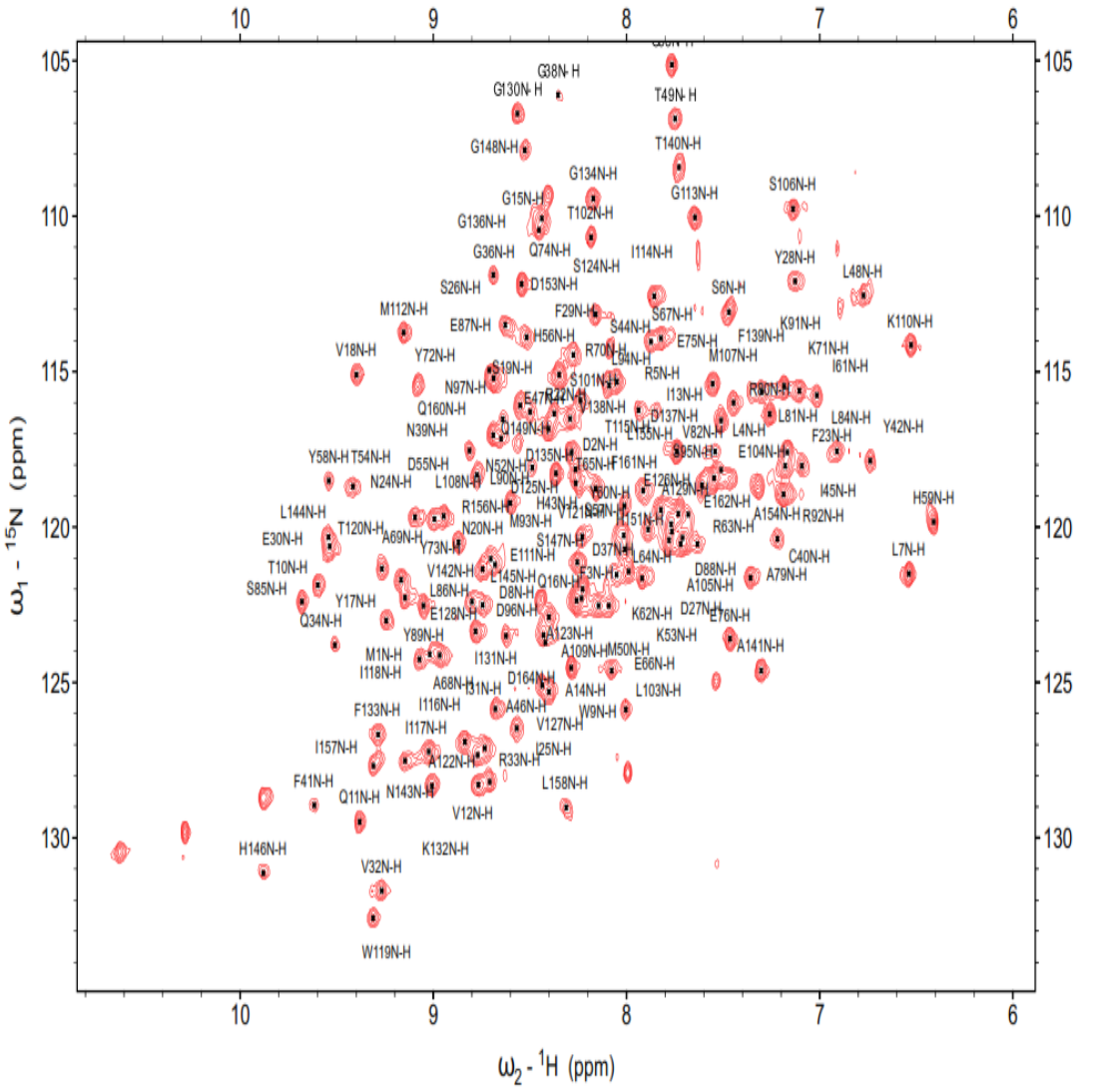


Figure 12B. The steady-state heteronuclear NOE spectrum of ^{15}N-labeled CCHFV OTU protease at pH 6.5 and 25°C acquired on a Varian INOVA 600 MHz spectrometer with no saturation (no*NOE*).

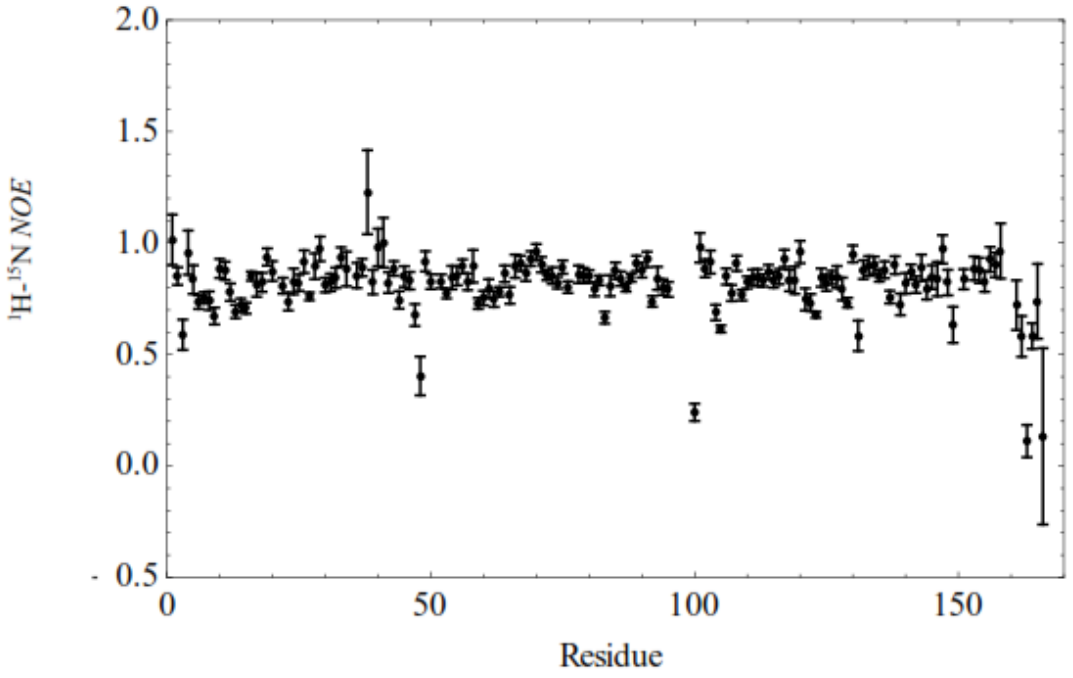


Figure 13A. Heteronuclear *NOE* measurements of ^{15}N-labeled CCHFV OTU protease at pH 6.5 and 25 °C determined at 14.1 T (600 MHz) field strength.

The T_1 measurements at 600 MHz in Figure 13B, showed a variation across the sequence that is similar to that exhibited by the NOE values. This is reasonable because both measurements are sensitive to dynamics on similar timescales. The maximum, minimum and average T_1 values at 600 MHz are 1.1 s, 0.15 s and 0.85 s. At 800 MHz a similar pattern was observed (Figure 14B) and the maximum, minimum and average values are and 1.25 s, 0.6 s and 1.15 s. The T_1 values decline steeply towards the C-terminus of the protein indicating faster relaxation and shorter rotational correlation times there. This suggests disorder at the C-terminus, and this pattern is similar to the *NOE* values. The average T_1 of 0.85 s at 14.1 T is compatible with what would be expected for a protein with a well-folded backbone. However, higher rates of relaxation were observed for amino acid residues between Pro^{35}-Cys^{40} which are near the catalytic active site region of the protein and are consistent with literature[29] data indicating faster dynamics in that segment. The amino acids Thr^{49}, Asp^{55}-Tyr^{60}, and Gly^{100} also show low T_1 values suggesting faster relaxation there and possibly indicating flexibility. It should also be mentioned that erroneous assignments might explain some of these values.

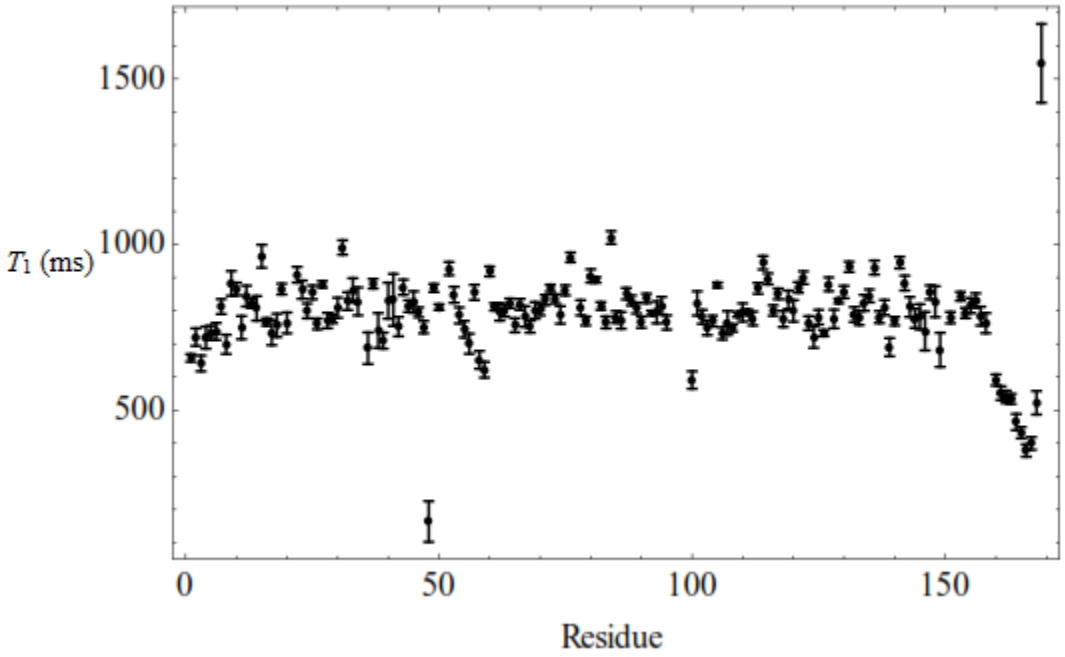


Figure 13B. Longitudinal relaxation times T_1 (ms) for ^{15}N-labeled CCHFV OTU protease at pH 6.5 and 25 °C, determined at 14.1 T (600 MHz) field strength.

The T_2 relaxation measurements yield significant insights into both the fast and the slow dynamics timescales. High T_2 values usually suggest unrestricted fast dynamics on the picoseconds to nanoseconds time-scales. Low T_2 values suggest restricted dynamics on the picoseconds to nanoseconds timescale with possible contributions from slow conformational exchange on the millisecond to second timescale. In Figures 13C and 14C the maximum, minimum and average values of the T_2 measurement rates for 600 (and 800 MHz) field strength are 150 ms (270 ms), 10 ms (40 ms) and 80.5 ms (90.3 ms). The T_2 values increase dramatically towards the C-terminus suggesting high flexibility, unrestricted fast motion and disorder at this region. The region from Ala^{105}-Lys^{110} in both the 600 and 800 MHz data also show relatively high T_2 values indicating possible unrestricted fast dynamics and a flexible backbone in this segment of the protein. A notable feature from the T_2 measurements is that the region spanning from Cys^{40}-Met^{50} shows the fewest number of dynamics measurements because this region contains peaks with the lowest intensities, suggesting that the residues in this region undergo slow conformational exchange. This is in contrast to a report in the literature[29] showing elevated T_2 values in this region. In addition, there are low T_2 values near residues Asn^{39} and Gln^{149} suggesting restricted fast dynamics and possible contributions from slow conformational exchange. The value measured for residue Thr^{49} is likely to be an artefact.

One further observation is worth noting. The 18.8 T relaxation data are missing peaks from the C-terminus especially in the NOE measurements (Figure 14A). The 18.8 T measurements were made in Edmonton which meant that the samples were not as fresh as the ones measured at 14.1 T. And the NOE measurements, that are missing the most peaks, were the last to be measured. One explanation for the missing peaks is that the C-terminal residues were proteolyzed to individual amino acids that would not be visible in the spectra. Another

possibility is that the C-terminus, possibly through protein:protein interactions, slowly adopted a conformation ensemble in the slow exchange regime (ms – s) broadening their resonances beyond detection.

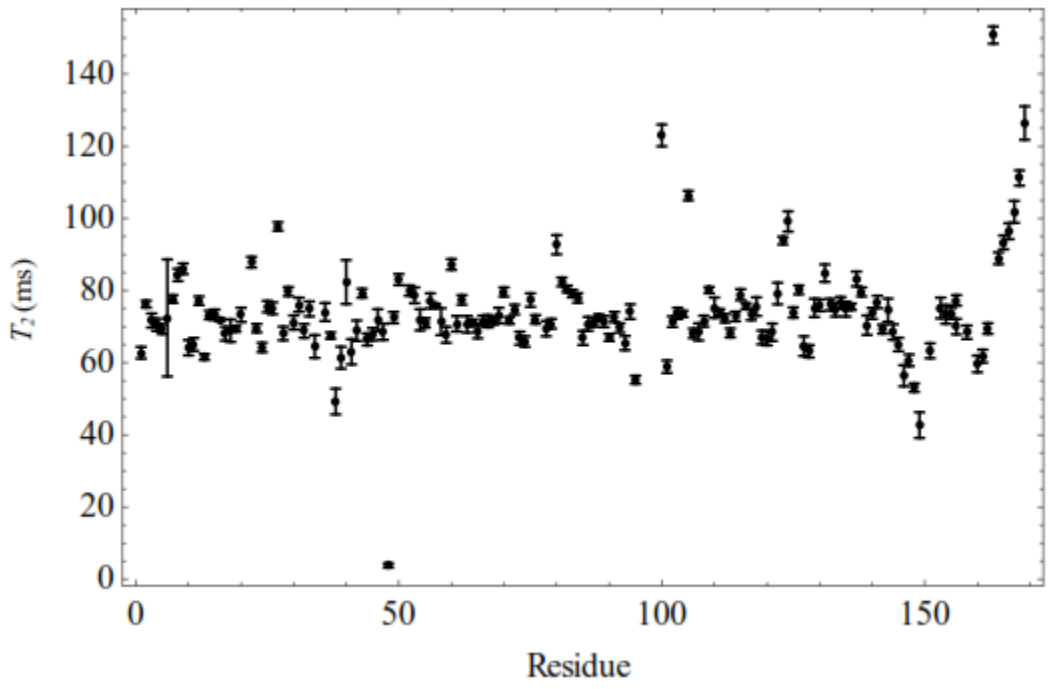


Figure 13C. The transverse relaxation times T_2 (s) for ^{15}N-labeled OTU at pH 6.5 and 25 °C, determined at 14.1 T (600 MHz) field strength.

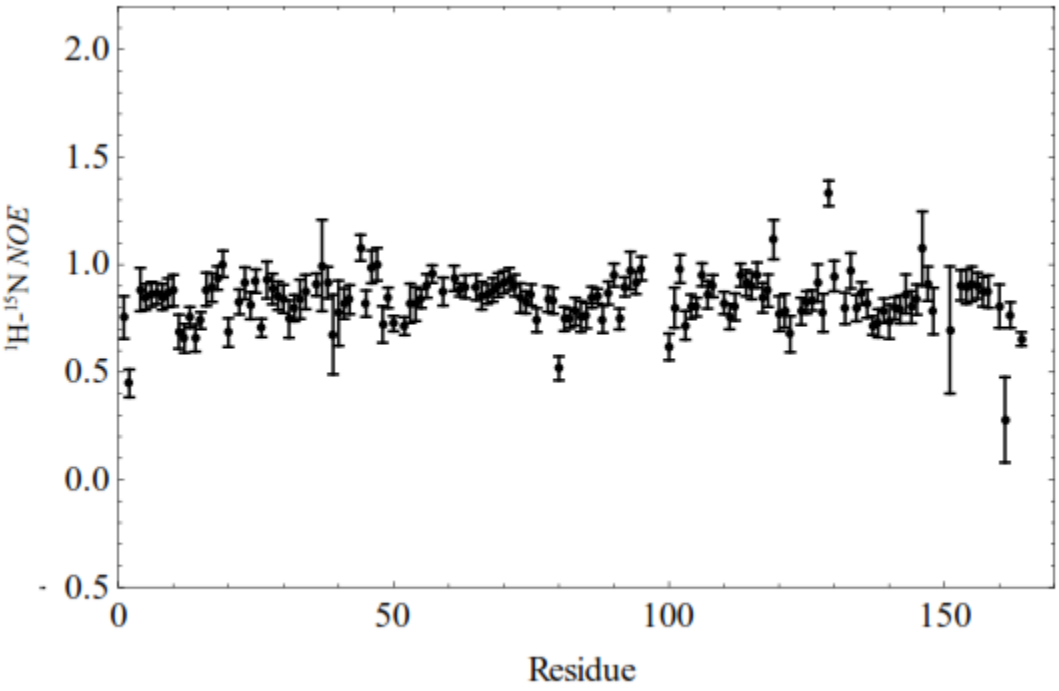


Figure 14A. The heteronuclear NOEs for ^{15}N-labeled CCHFV OTU protease at pH 6.5 and 25 °C, determined at 18.8 T (800 MHz) field strength.

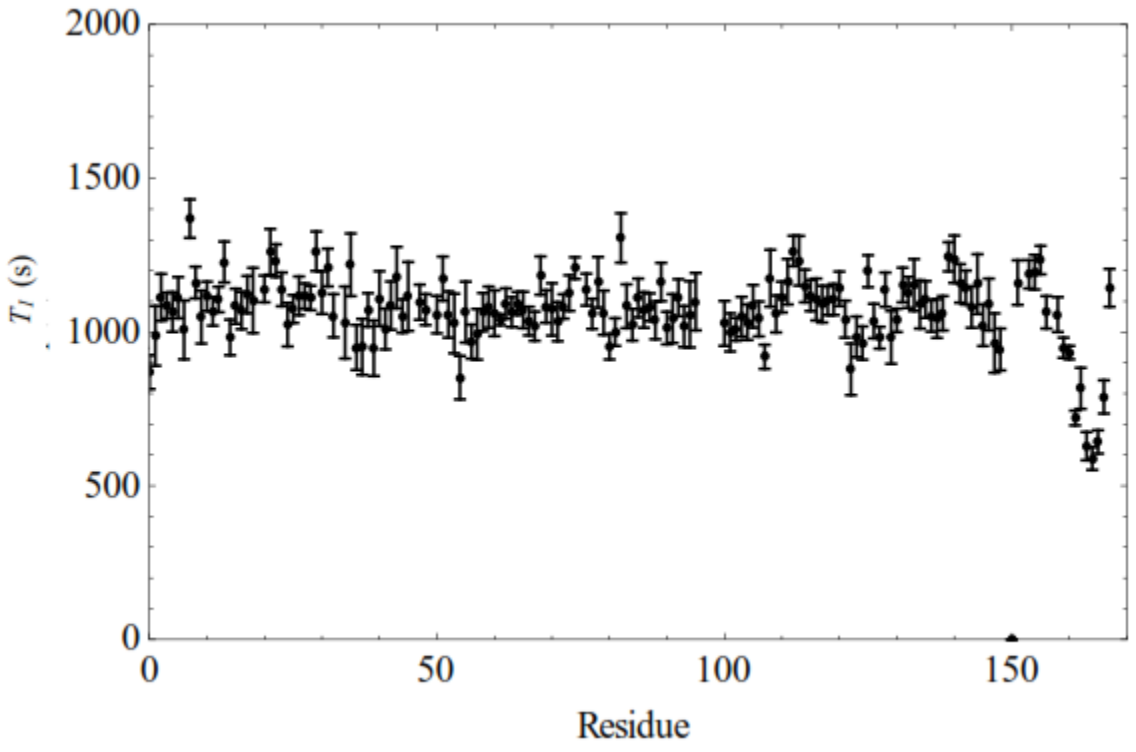


Figure 14B. The longitudinal relaxation times T_1 (s) for ^{15}N-labeled CCHFV OTU protease at pH 6.5 and 25 °C, determined at 18.8 T (800 MHz) field strength.

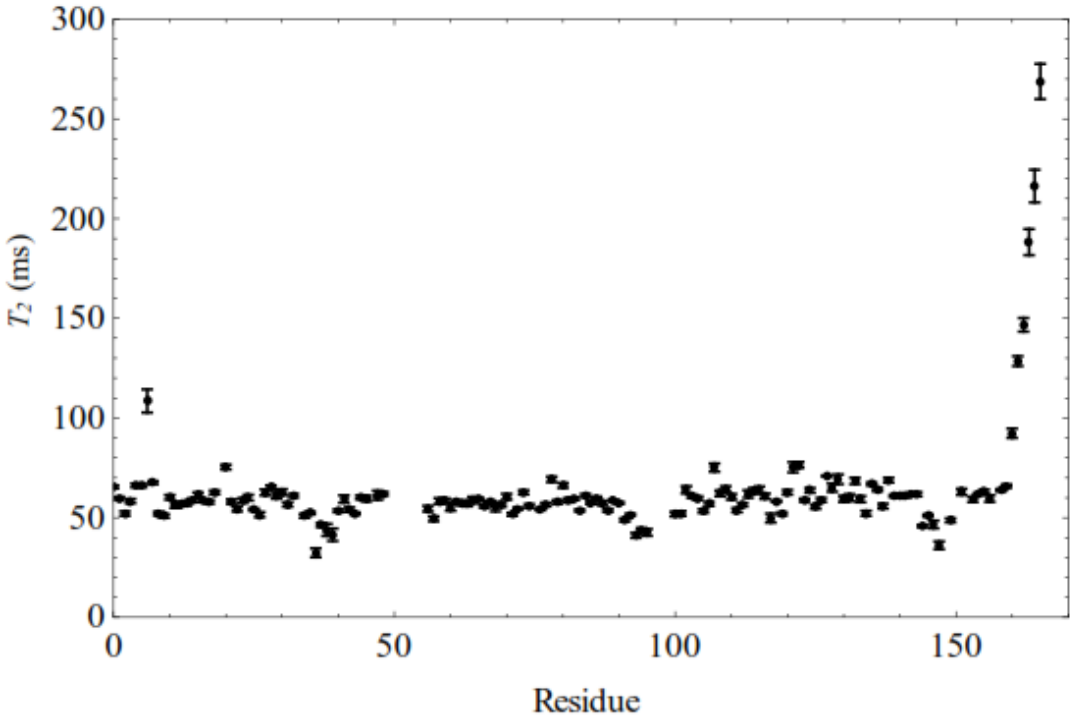


Figure 14C. The measured transverse relaxation times T_2 (s) of ^{15}N-labeled CCHFV OTU protease at pH 6.5 and 25 °C, determined at 18.8 T (800 MHz) field strength.

3.6 Reduced Spectral Density Mapping

Reduced spectral density mapping is a graphical data presentation method that illustrates the relative contributions to NMR relaxation of motions in different time domains, does not involve any assumptions about molecular dynamics, is easy to interpret and can lead to some insight about the relative flexibility of different parts of a protein[57,65,66,117]. The reduced spectral density function at zero frequency $J(0)$ provides insight into protein global rotational tumbling, $J(\omega_N)$ provides insights about faster regional motions in the proteins, and $J(\omega_H)$ depicts information like bond vector fluctuations (even faster internal motions) for each amino acid residue in the protein sequence. The reduced spectral density functions were calculated at three different frequencies $J(0)$ @ 0MHz, $J(\omega_N)$ @ 61MHz and 81 MHz, and $J(0.87\omega_H)$ @ 522MHz and 696 MHz from the relaxation data measured at two different magnetic field strengths (600 and 800 MHz). Below, I present the data measured at 600 MHz because of the problems with the 800 MHz (18.8 T) NOE data described earlier.

The frequencies $J_{eff}(0.87)$ in Figure 15A corresponding to 522 MHz are 0.87 times the ^{1}H Larmor frequencies. High values are found near the C-terminus region of the protease. This tells us that at the C-terminus dynamics on the ps time-scales contribute more to relaxation of the highly disordered C-terminus than in the rest of the protein that is well folded.

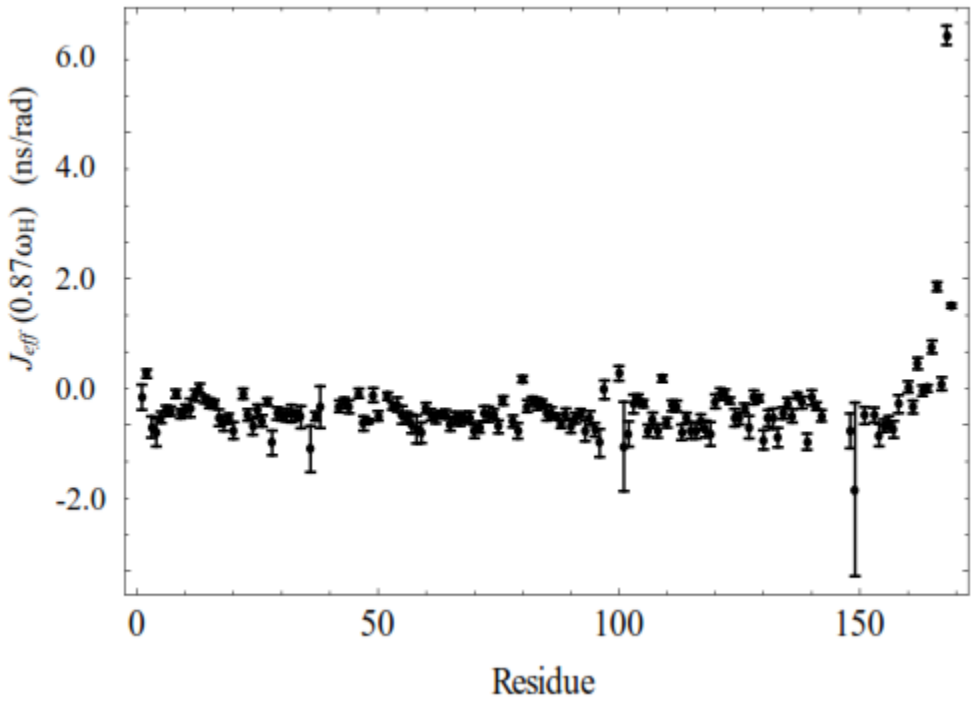


Figure 15A. The reduced spectral density values for mapping motions at 522 MHz (J 0.87ωH) (ns/rad) of ^{15}N labelled CCHFV OTU protease at pH 6.5 and 25 °C estimated according to the Farrow *et al*., method[65].

The $J(\omega_N)$ values plotted in Figure 15B measure contributions to relaxation from motions that are 10-fold slower than $J_{eff}(0.87)$. The residues Asp[55], Gly[100], Ala[123], and Gly148 show some of the highest contributions at mid-frequencies. There are also significant motions at the C-terminus region of the protease. The residues Tyr[17] and Ser[85] showed the smallest contributions at mid-frequencies.

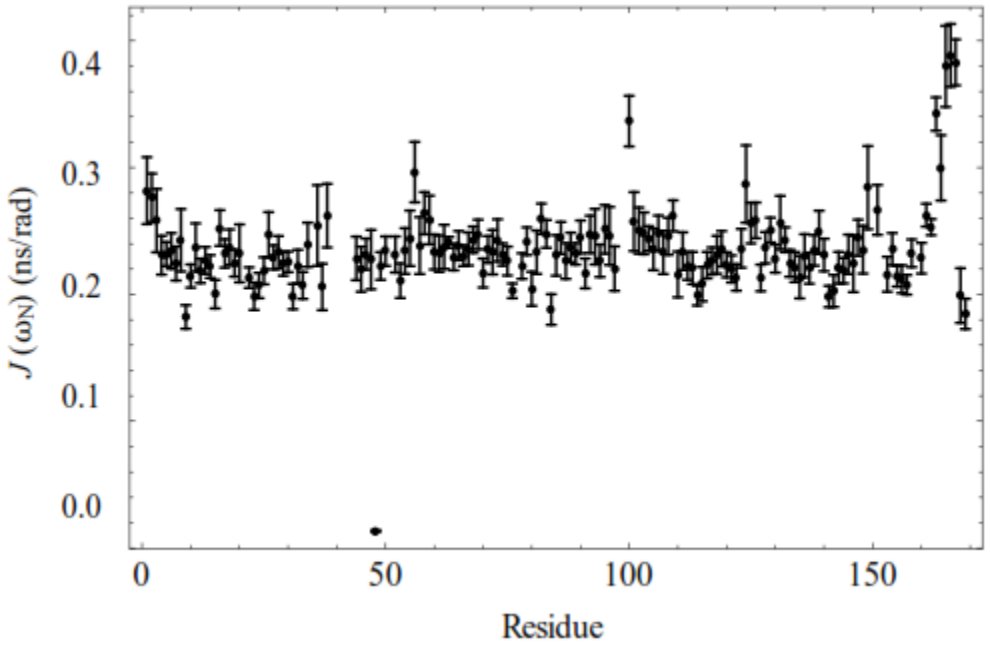


Figure 15B. The reduced spectral density values for mapping motions at 61 MHz $J(\omega_N)$ of ^{15}N-labelled CCHFV OTU protease at pH 6.5 and 25 °C.

The low frequency $J_{eff}(0)$ spectral density values are shown in Figure 15C. The most notable feature of the plot is the low values in the C-terminus. This indicates a low contribution of slow motions to relaxation in the highly flexible disordered C-terminus compared to the rest of the protein. Interestingly, contributions to relaxation at low frequencies are highest around residues 95 and 150 where other data suggest slow conformational events. However, it must be noted that the errors are highest in these regions.

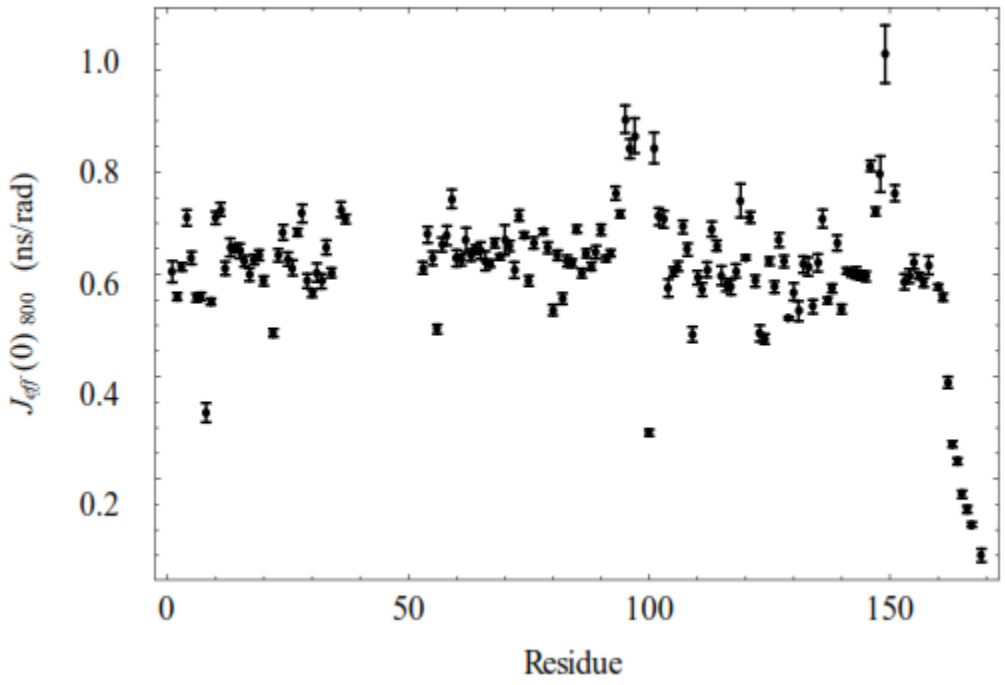


Figure 15C. The reduced spectral density values $J_{eff}(0)$ (ns/rad) from 600 MHz measurements for mapping motions of ^{15}N labelled CCHFV OTU protease at pH 6.5 and 25 °C.

3.7 *Lipari-Szabo* Model-Free Formalism

The relaxation rates of amino acids in the CCHFV OTU protease at 14.1 and 18.8 Tesla magnetic fields were fit by employing *Lipari-Szabo* model-free analysis to gain insights into its range and timescale of the backbone dynamics[113, 115, 134, 135]. In contrast to the original Lipari-Szabo method we used an approach here that derives overall and internal rotational correlation times for each residue as this method is reported to provide a statistically significant improvement to relaxation data[113]. Among the dozen or so possible versions of the Lipari-Szabo spectral density function[71] we chose so-called Model 3 to fit the relaxation data to. This model-free system uses four principle parameters – S^2 which gives information about the range of

motions responsible for spin relaxation, τ_c that provides information about the overall rotational correlation of the entire protein, τ_e that provides information about fast internal N-H bond motions on the ps-ns time-scales and the R_{ex} parameter that provides insights into slow motion on the millisecond-second time-scale[136]. As described in the Methods section, Model 3 was chosen based on a consideration of the total *R*-factors (R_f) and Akaike information criterion (AIC). For the Model 3 parameters are shown in Table 2 below:

Model	R_f [R_1] (s^{-1})	R_f [R_2] (s^{-1})	R_f [*NOE*]	R_f *Total*	**Field-1 (11.7 T]**	**Field-2 [14.1 T]**	**Mean [AIC]**	**SD [AIC]**
3	0.092	0.083	0.438	0.0835	600 MHz	800 MHz	43. 20	31.6

Table 2. Individual and total R-factors (R_f) and Akaike information criterion values (AIC) for two-field extended *Lipari-Szabo* analysis for Model 3.

Figure 16A shows that most of the residues exhibit a τ_c of between 9 - 10 ns in good agreement with what would be expected for a protein with a mass of 18 kDa. Near residues 28, 38-42, 60, 100, 105, 151 and in the C-terminus, the rotational correlation times are shorter. This suggests increased backbone flexibility in those regions. However, some of the errors are high especially for residues 38-42 and for some of the C-terminal residues. For residues 38-42 this is most likely because the peaks are so broad and have low intensity. In the C-terminus the high error bars may be caused by the lack of NOE data collected at 18.8 Tesla that was described earlier.

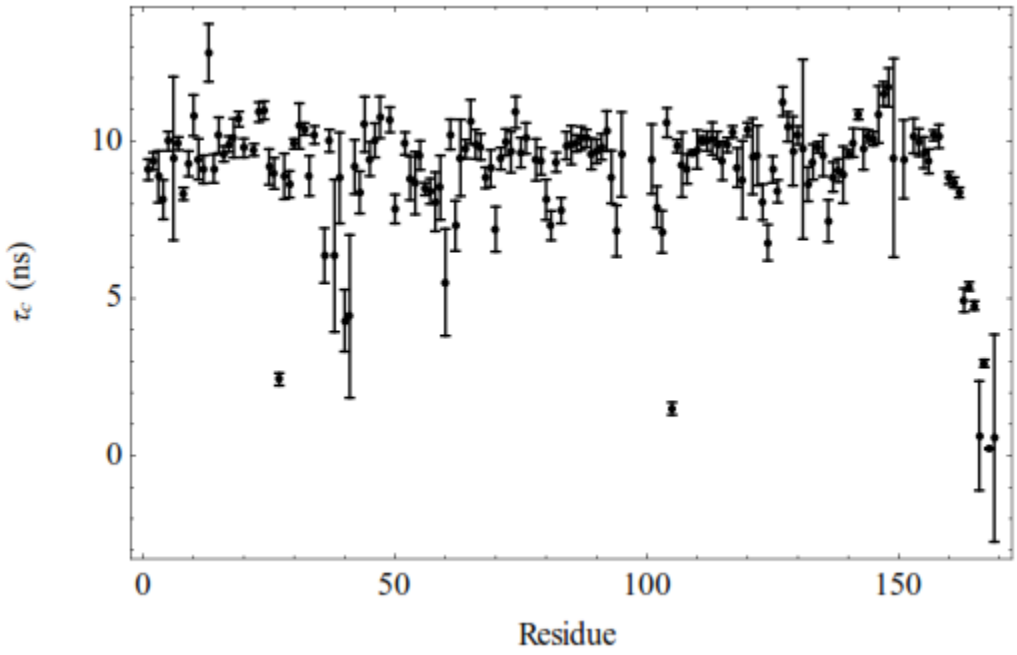


Figure 16A. Model-free τ_c parameters with errors estimated from 500 Monte Carlo sets using Model 3 from the relaxation data (T_1, T_2, and *NOE*) acquired at two fields (14.1 T and 18.8 T) in CCHFV OTU protease.

The order parameter values (S^2) are shown in Figure 17B. For most residues they are in the range of 0.85-0.95 suggesting a restricted dynamic spatial range on the ns-ps timescale. This is consistent with a stably folded protein backbone. In addition, the same residues that showed short rotational correlation times (τ_c) also show lower order parameters (S^2) possibly indicating increased dynamics in those regions of the protein. Note however, that the same regions that showed high error bars in the τ_C data also show high errors in the S^2 data.

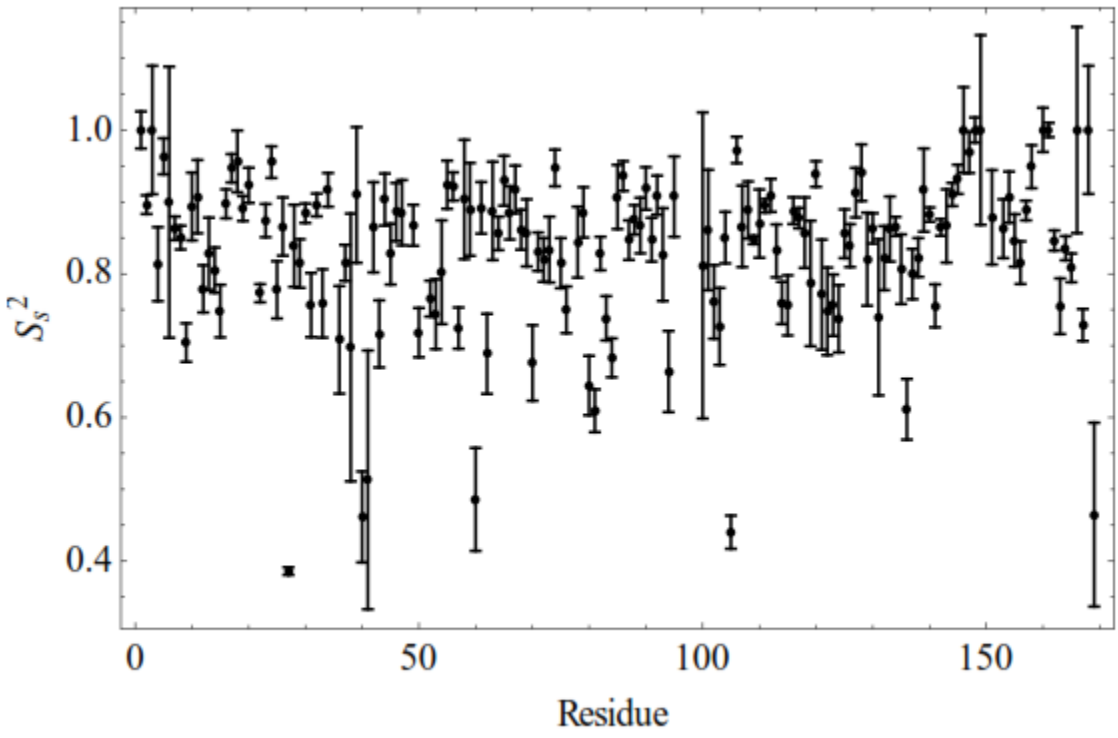


Figure 16B. The Model-free generalized order parameters (S^2) with errors estimated from 500 Monte Carlo sets using Model 3 with the relaxation data (T_1, T_2, and *NOE*) acquired at two fields (14.1 T and 18.8 T) in CCHFV OTU protease.

Figure 16C shows the distribution of internal correlation times (τ_e) across the protein sequence. Most of the residues show values between 20 – 40 ps and there is little noticeable variation across the sequence. Some values have very large error bars and the significance of these results is minimal.

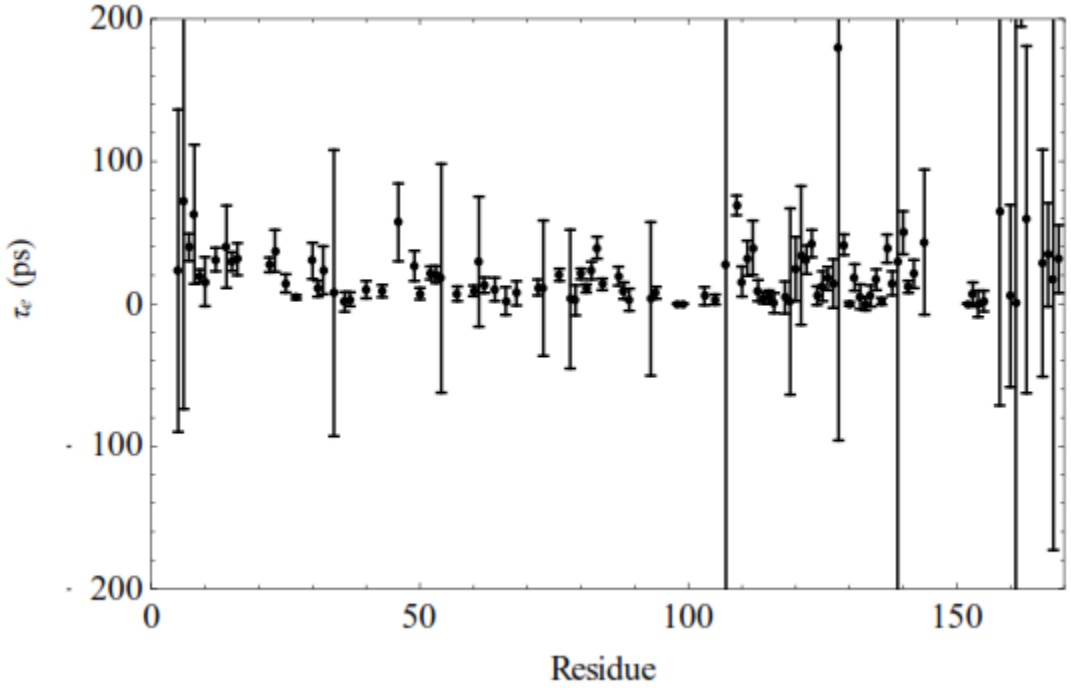


Figure 16C. Model-free internal rotational correlation times τ_e (ps) with errors estimated from 500 Monte Carlo sets using Model 3 with the relaxation data (T_1, T_2, and *NOE*) acquired at two fields (14.1 T and 18.8 T) in CCHFV OTU protease.

On the other hand, significant differences in R_{ex} across the amino acid sequence are observed in Figure 16D. R_{ex} measures slow conformational events on the millisecond timescale and it's worth noting that many residues show elevated R_{ex}. In particular, high R_{ex} values near Asn-39 suggest slow conformational events near the active site and high R_{ex} near Gly-100 suggest slow conformational events at the Ubiquitin binding cleft. Recall that the T_2 measurements obtained from our data measured at both the magnetic fields as shown in Figure 13C and 14C also showed elevated values in the same regions of the CCHFV OTU protease segments suggesting slow conformational events in these regions.

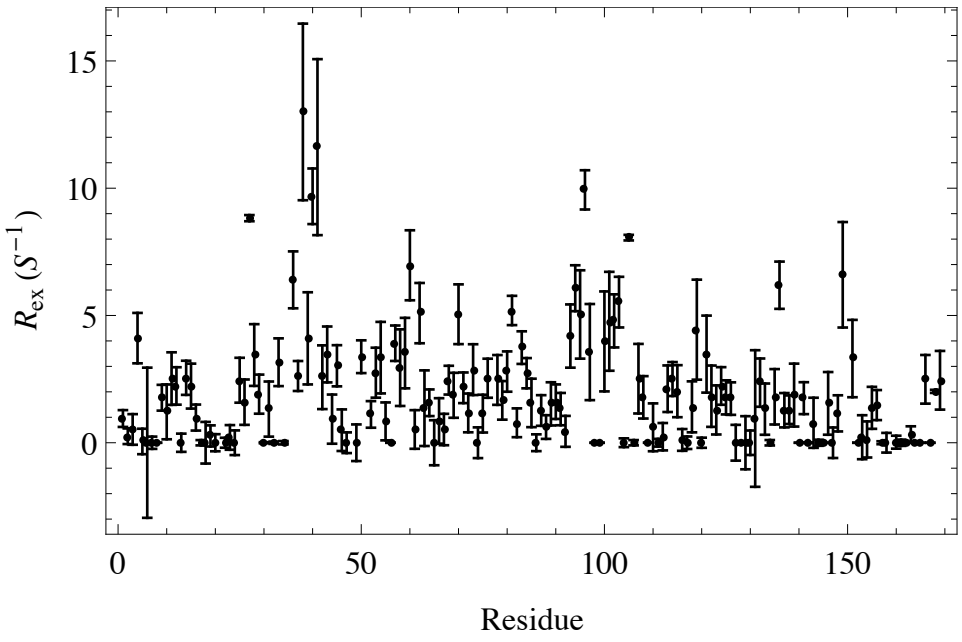


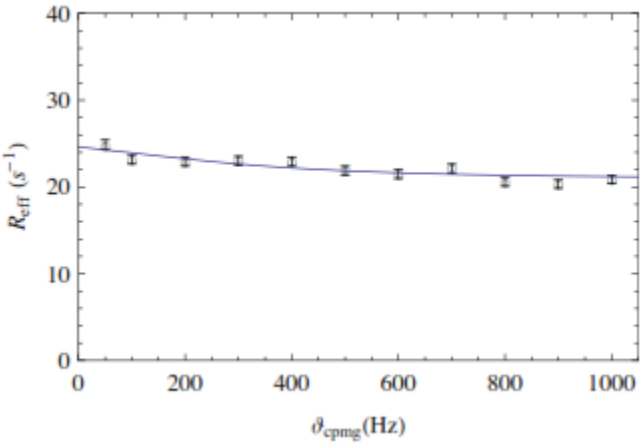
Figure 16D. Model-free R_{ex} values with errors estimated from 500 Monte Carlo sets using Model 3 with the relaxation data (T_1, T_2, and *NOE*) acquired at two fields (14.1 T and 18.8 T) in CCHFV OTU protease.

3.8 *Carr-Purcell-Meiboom-Gill* Relaxation Dispersion Experiment

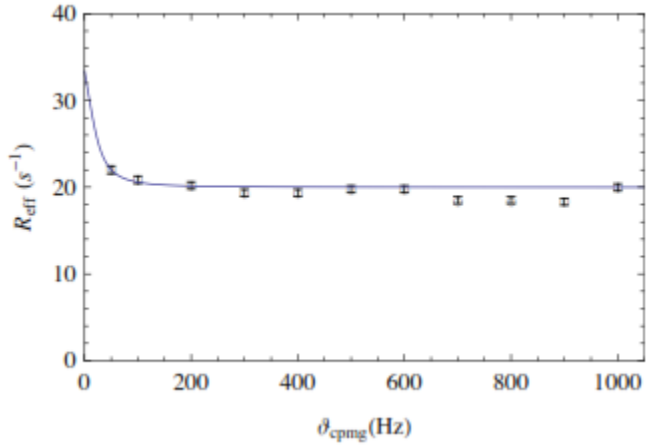
The CPMG relaxation dispersion experiments provide important quantitative insights into conformational exchange events on the milliseconds-to-millisecond time-scales[136]. These are of interest as they may represent rate-limiting protein motions in an enzyme's catalytic pathway. The ^{15}N CPMG relaxation dispersion experiment effectively records the transverse relaxation rate constants, T_2, in during a CPMG pulse train as a function of interpulse delay and allows identification of ^{1}H-^{15}N bond vector equilibrium exchange processes[107]. Quantitatively, the relaxation dispersion experiment can yield significant information pertaining to chemical shift

differences between the ground and excited state conformations as well as the rate of exchange (k_{ex}) between the equilibrium populations (p_a and p_b)[136].

The NMR dispersion experiments were carried out at 25 ^{0}C and 14.1 T and the results fitted using the program NESSY[118]. Thirty-one residues were fit best to Model-2 (fast exchange) and 24 residues were best fit to Model 3 (slow exchange) and the remaining residues fit best to Model-1 (no exchange). The residues exhibiting conformational exchange on the millisecond time-scale are evident when R_{eff} data values are plotted as a function of $1/\tau_{cp}$ and curvature is seen in the graphs[116, 117, 119]. Several examples are shown in Figure 17. Flat profiles are evidence that the ^{15}N nuclear spins do not encounter any motion on the ms time-scale[107]. The relaxation dispersion curves of few residues displaying significant changes in intensity on the ms timescale are shown in Figure 17A-F and they are: Thr^{10}, Ile^{13}, Glu^{30}, Thr^{54}, Leu^{84} and Ser^{95} in the CCHFV OTU protease sequence. Note that many of these residues are situated near the binding cleft for ubiquitin (or ISG-15) although several a distant from the active site[28, 29]. Our findings strongly mesh with recently published[29] CCHFV OTU protease T_2 relaxation dispersion data acquired at 25^{0}C although we see more residues exhibiting slow exchange. Table 3 shows the best Models and the best fit values for R_{eff} and k_{ex} for the above mentioned amino acids.

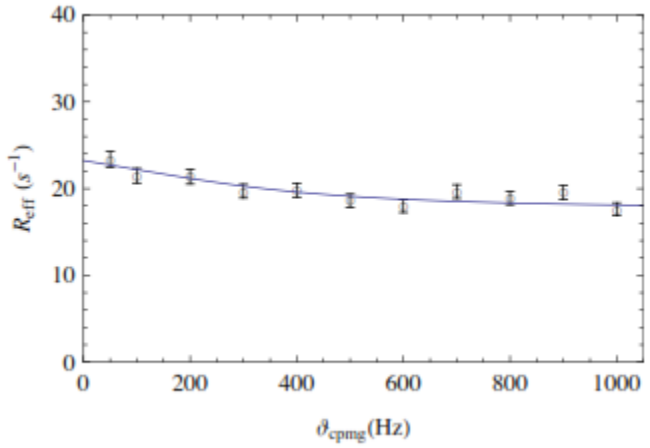


(A)

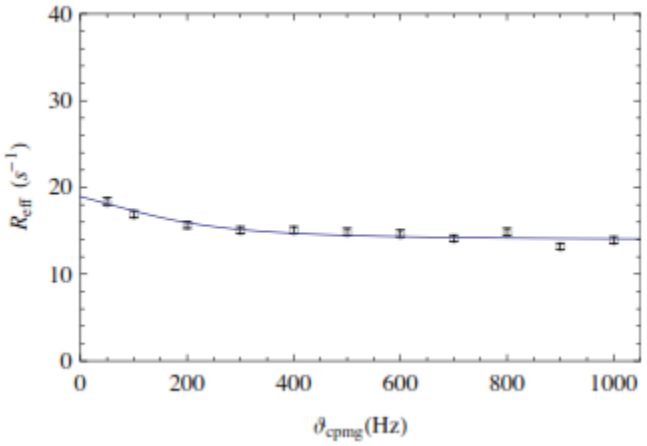


(B)

Figures 17A and B show relaxation dispersion curves measured at 14.1 T and 25°C for CCHFV OTU protease amino acid residues (A) Thr[10] and (B) Ile[13]. The data were fit using the NESSY software[118] to Model 3 (A) or Model 2 (B).

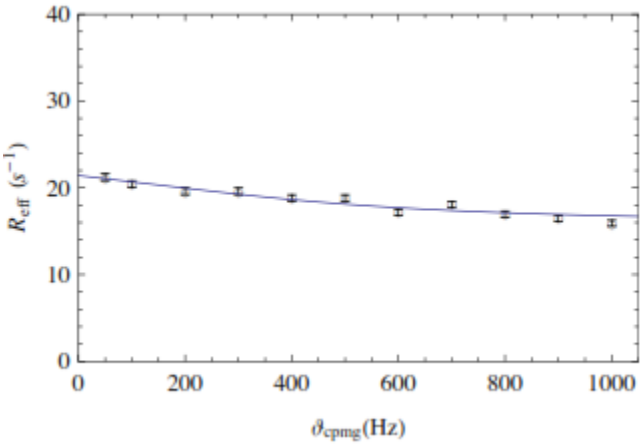


(C)

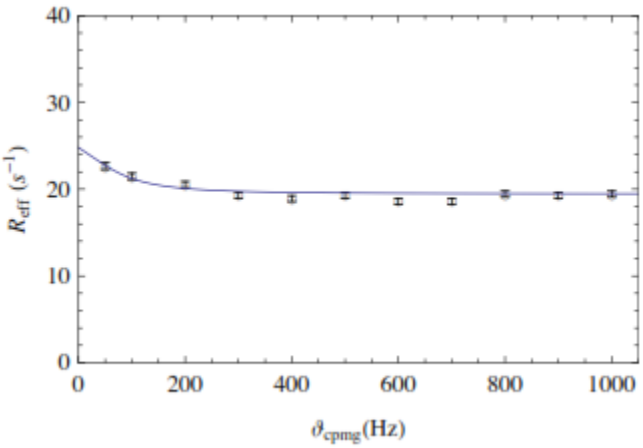


(D)

Figures 17C and D show relaxation dispersion curves measured at 14.1 T and 25°C for CCHFV OTU protease amino acid residues (C) Glu^{30} and (D) Thr^{54}. The data were fit using the NESSY software[118] to Model 2 (C) or Model 3 (D).



(E)



(F)

Figures 17E and F show relaxation dispersion curves measured at 14.1 T and 25°C for CCHFV OTU protease amino acid residues (E) Leu[84] and (F) Ser[95]. The data were fit using the NESSY software[118] to Model 2.

Table 3. The best-fit relaxation dispersion Models and best fits of R_{eff} and k_{ex} by residue.

Residue	Model	R_{eff} (s^{-1})	error	k_{ex} (s^{-1})	error
Thr^{10}	3	20.2	0.13	3000	236
Ile^{13}	2	18.9	0.07	810	97
Glu^{30}	2	19.0	0.08	902	153
Thr^{54}	3	18.3	0.11	1417	203
Leu^{84}	2	13.8	0.11	1394	142
Ser^{95}	2	13.5	1.2	5204	1130

4.0 Discussion

Cysteine thiolesterases are now emerging as possible drug targets for myriad diseases including cancer, multiple sclerosis and osteoarthritis[137-139]. The pathogenesis of these diseases is indicated by excess proteolysis and there is currently a paucity of inhibitors to be used as therapy to curb these diseases. Thus, synergistic development of potent drug design along with understanding the physiological functions of papain-like cysteine protease are both required to fight against various diseases. Enzyme:drug interaction studies can be further deepened by understanding the interrelationship between structure, function and conformational dynamics of enzyme-substrate interactions. Protein dynamics or flexibility on different time-scales have been shown to crucial in understanding such interactions[38]. In my research, I have used NMR spectroscopy as a tool to probe dynamic conformational changes in a viral enzyme CCHFV OTU protease, which causes fatal hemorrhagic fever in humans for which currently no therapeutic drugs are available. A recently published NMR dynamics study[29] on the CCHFV OTU protease showed conformational dynamics on multiple time-scales along with weaker binding affinity towards ubiquitin and ISG-15-like molecules. Slow dynamics on the ms timescale was found largely localized to the catalytic region[29]. In view of advancing our understanding, we envisaged to further extend the research to reveal broader information about CCHFV OTU protease flexibility by using various NMR experiments. The backbone amide dynamics on uniformly ^{15}N-labelled CCHFV OTU protease, spin relaxation measurements, reduced-spectral density mapping, and model-free analysis revealed that most of the amino acid residues participate in a well-folded secondary structure conformation but the C-terminus is disordered and some residues show evidence of slow conformational exchange. I will now discuss some of the major findings and in each case

all of the figures and tables referred to in the remaining discussion can be found in Chapter 3 – Results.

4.1 Protein Expression and Purification

The pGEX6P-1-CCHFV-OTU vector with its N terminal fused to a cleavable Glutathione-S-transferase tag was used for over-expression of CCHFV OTU protease in overnight culture. This GST-fusion protein system has been successfully employed in other NMR studies[140]. Glutathione-S-transferase (GST) is a homodimeric 25 kDa eukaryotic protein[141] which has a strong affinity for glutathione linked to a GST Bind™ resin (reduced glutathione matrix) (Novagen). This binding is reversible and aids in eluting the GST-CCHFV OTU protease fusion protein under mild and non-denaturing conditions (pH = 6.5). The PreScission protease HRV 3c cleaves at the Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro recognition sequence between the GST tag and the CCHFV OTU protease and was used for separating GST from the folded CCHFV OTU protease. DTT was used a reducing agent in the elution buffers to keep glutathione reduced as well as the cysteine residue in CCHFV OTU protease[205]. The cleaved CCHFV OTU protease was further purified and separated from the GST and pre-scission protease by gel-filtration (FPLC) method. The Superdex 75 size-exclusion gel filtration column was used to separate the globular protease and OTU because it has a suitable fractionation range (3,000 – 70,000 kDa) it minimizes absorption of CCHFV OTU protease to the matrix, and allowed easy separation and elution of CCHFV OTU protease in the elution buffer[142].

4.2 NMR Spectroscopy and Backbone Assignment

Figure 4 shows a very typical ^{15}N-HSQC spectrum of the CCHFV OTU protease. The crosspeaks in are widely dispersed suggesting a folded conformation of the CCHFV OTU protease[143]. The average ^{1}H and ^{15}N resonance line-widths are 9 and 3 Hz respectively also as expected for a folded proteins[143]. The very narrow lines in the C-terminus suggest rapid rotational correlation times and disorder in the last 7 residues. However, the spectrum also contains a number of very weak peaks. The broader lines and weak peaks exhibited by residues near Cys^{40}, His^{151} and Asp^{153} (Figure 5) suggest conformational exchange on the microsecond-to-millisecond time-scales in the catalytic active site. Although the crystal structures[28] do not indicate conformational flexibility it's possible that this is owing to crystal constraints or that the structure may stabilize after binding with ubiquitin or ISG-15. Even at 800 MHz (18.8 T) residues such as Asn^{98}, Trp^{99}, and Phe^{153} remain unassigned and this suggests that the area near Trp^{99} also undergoes slow conformational exchange.

4.3 Sequence Specific Resonance Assignment

4.3.1 Three-dimensional Experiments

The preliminary sequence-specific assignment of all amino acids in the CCHFV OTU protease (18 kDa) was accomplished by Dr. Joe O'Neil using the pine server assignment algorithm and samples prepared by Terry James. They used the three-dimensional experiments HNCO, HN(CA)CO (Figure 6), CBCA(CO)NH, HNCACB, and HNHA. The principal goal of these experiments was to examine and assign all backbone ^{1}H, ^{13}C and ^{15}N resonances in the CCHFV

OTU protease sequence. These experiments connected the neighbouring residues of amino acids in the CCHFV OTU protease along the polypeptide chain sequence[144].

4.3.2. "*Unlabelling*" of Amino Acids in CCHFV OTU protease

Although 3-D experiments were able to successfully assign many amino acids in the CCHFV OTU protease sequence, however there were some residues such as Gly^{136}, Leu^{48}, and Ile^{117} which showed weak peak intensity in the 2D spectrum (see Figure 4), thereby questioning the validity of the assignments. Other ambiguous assignments included amino acids G^{15}, Q^{16}, F^{29}, I^{31}, V^{32}, C^{40}, H^{59}, D^{96}, Gly^{100}, W^{119}, T^{120}, V^{121}, S^{124}, and N^{143} that showed peak-doubling, peak multiplicity or extremely broad resonances. To confirm these ambiguous assignments and to determine different amino acid residues in the CCHFV OTU protease on the basis of its spin-system, I decided to use inexpensive selective isotope "*unlabelling*" experiments of all amino acids except proline and glycine, instead of the more expensive selective isotope labelling strategy[144]. The principal advantage of the former experiments over the latter is that the cross-peaks of the desired amino acids are not observed in the 2-D spectrum of a uniformly-labelled protein. This is because the bacterial amino acid biosynthesis is regulated at the enzymatic and gene level[145] and supplying the growing *E. coli* bacterial cells expressing CCHFV OTU protease with unlabelled amino acids causes the cells to restrict endogenous amino acid biosynthesis and incorporate the unlabelled amino acid. One problem that can arise with this strategy is isotope scrambling. However, most of the amino acids in the CCHFV OTU protease sequence showed either no or very minimal cross-metabolism except Glu and Gln which showed high isotope scrambling. The result of this approach to assignment was a nearly complete backbone

assignment that was validated by secondary structure analysis and comparison to the crystal structure[28].

4.4 Chemical Shifts and Coupling Constants

The chemical shift analysis was utilized for estimating the secondary structure (α-helix, turns and *β*-sheet conformation) of CCHFV OTU protease[146-148]. Two methods were used including the chemical shift index (CSI) method developed by Wishart and Skyes[149] (Figure 9) and the Secondary Structure Propensity method developed by Julie Forman-Kay and coworkers[105] (Figure 10). The remarkable agreement between the secondary structure boundaries observed in the crystal structure[28] (see Table 1) and those determined by NMR is an important result. The agreement between the structures measured by the two methods suggests that the vast majority of our assignments are correct. It also allowed us to us the X-ray crystal structure as a model for interpreting the solution dynamics studies. Although the secondary structure analysis of the $^{3}J_{HNH\alpha}$ coupling constants (Figure 11) generally agreed with the chemical shift measurements there are many places where the agreement is weak. It is noteworthy that the coupling constants suggest a number of residues in the 6-8 Hz region indicating dynamic flexibility and suggesting that the protein may be more flexible than suggested by the X-ray diffraction[28] structure and the SSP analysis. This interpretation is supported by the relaxation data including the large number of residues exhibiting slow dynamics in the relaxation dispersion experiments.

4.5. NMR Relaxation Experiments

Assignment of the majority of backbone ^{1}H and ^{15}N resonances allowed a comprehensive set of NMR relaxation data to be acquired spanning the picosecond to millisecond timescales for 156 observable resonances at two magnetic field strengths (14.1 and 18.8 T). The relaxation data (Figures 13 and 14) corroborated the linewidth, chemical shift, and coupling constant analysis described above. The C-terminal residues exhibited short T_1's and NOE's, and long T_2's clearly indicating that the protein is disordered there in agreement with the missing electron density in the X-ray diffraction data[28]. The processed relaxation data suffered from weak intensities for some peaks as well as missing data for the 18.8 T NOE experiment that led to high noise in the Lipari-Szabo parameters and missing spectral density data at the C-terminus. Despite these deficiencies, the Lipari-Szabo overall rotational correlation times (τ_c) in the C-terminus are significantly lower than in the rest of the protein (Figure 17C) confirming the unfolded state of the C-terminus in solution. It's also telling that the spectral density values at the C-terminus suggest a large contribution to relaxation from dynamics on the ps – ns timescale as expected for a disordered segment (Figure 15A). Disorder at the ends of proteins is not unusual and is often referred to as "end effects"[150, 151].

In addition to the C-terminus, residues near the active site (Cys^{40}, His^{151}, Asp^{153}) and substrate binding cleft (Gly^{100} forms part of the β-lobe that lines the binding cleft) showed some deviations from average relaxation values. For example, those two regions show low T_2 values suggesting possible relaxation by slow conformational exchange on the ms –μs timescale (Figure 14C). In addition, Lipari-Szabo analysis and Reduced Spectral Density analysis both showed significant relaxation contributions to T_2 by slow exchange as suggested by significant R_{ex} values

(Figure 15D and 17D). The noise in these plots suggestion caution in this interpretation. However, direct measurements of slow exchange by relaxation dispersion suggest that as many as 55 residues in the protein are involved in slow conformational dynamics (Figures 18). Many, but not all of these residues are located in the regions earlier suspected of exhibiting slow conformational exchange: the active site and the binding cleft. But other residues distant from the substrate binding surface and active site also appear to be involved.

While my work was in progress, another group published an NMR study of CCHFV deubiquitinase[29]. The study reported that NMR assignments had been accomplished but they were not published or deposited in the BMRB data bank. Neither did this group analyze their NMR data by Lipari-Szabo or Reduced Spectral Density analyses so these are novel aspects of my research presented here. They did report relaxation dispersion measurements but they observed significant slow conformational events for only a few residues. The slowly exchanging residues were found near the N-terminus of the protein and in the catalytic cleft in agreement with what I report here. However, I observed far more residues undergoing slow conformational exchange. In Figure 18, the amplitudes of the dispersion curves for Thr^{10}, Ile^{13}, Glu^{30}, Thr^{54}, Leu^{84}, and Ser^{95} are small suggesting that the differences in chemical shift between the ground and excited states being detected are small. This suggests that the slow conformational change that is detected by NMR is a subtle one. Not all enzymes undergo a large conformational change during catalysis yet they exhibit significant dynamics on the microsecond-millisecond timescales suggesting that dynamics is important in catalysis[63]. The crystal structures of CCHFV OTU deubiquitinase suggest no large change in conformation upon binding ubiquitin or ISG15[152,28, 153] My NMR results show that subtle conformational changes occur on the ms-μs timescale throughout much of the protein and these events may reflect important rate-limiting steps in the

catalytic cycle. Another possibility is that these conformational events may be important in enabling the broad substrate specificity of the CCHFV deubiquitinase allowing it to bind equally well to ubiquitin and ISG15. To distinguish between these possibilities and to determine the role of the detected conformational exchange will require measuring the changes in dynamics caused by substrate binding.

5. Conclusion and Future Research Goals

The research in my thesis was focused on understanding the conformational exchanges and dynamic behavior of a viral enzyme Ovarian tumour domain protease from the Crimean-Congo hemorrhagic fever virus (CCHFV) on different time-scales in solution. This viral enzyme causes serious fatal hemorrhagic fever among humans in more than 30 countries in the world for which currently no therapeutic drug interventions are available. I utilized NMR spectroscopy as a tool to probe CCHFV OTU protease backbone degree of flexibility and conformational changes on different time-scales. We successfully assigned nearly all of the H, C and N backbone resonances in the OTU protease by standard 3 dimensional experiments. NMR chemical shifts and coupling constants agreed very well with the X-ray diffraction structure. This is important because it allowed us to use the X-ray structure to interpret the NMR data. The narrow peaks and relaxation data strongly suggest that the last 6 or 7 amino acids at the C-terminus of the OTU protease are disordered. As I mentioned before, this agrees very well with the X-ray diffraction structure which showed no electron density for the last 7 amino acids suggesting that they are disordered in the crystal. Broadened peaks, intermediate coupling constants, and slow exchange in the CPMG experiments suggested the possibility of dynamic flexibility near the active site residues (Cys 40, His 151, and Asp 153) of the enzyme. However, the CPMG relaxation dispersion curves suggest that the slow conformational event extends throughout the protein. The CPMG relaxation dispersion curves also suggested that the slow conformational event extends throughout the protein.

An important goal for the immediate future is to repeat the 18.8 T NOE experiments because of the loss of data at the C-terminus. In the longer term, using the assigned spectra we

can now measure dynamics in the presence of substrates in order to connect the conformational events to the catalytic cycle. The first step in an enzyme dynamics analysis is the identification of the protein segments that change their mobility upon binding substrates. This study reports the dynamic properties of the protein backbone in the absence of substrate. In the future, circular dichroism (CD) spectroscopy and further NMR relaxation studies and chemical shift difference mapping should be carried out to see if there are any significant conformational changes in the CCHFV OTU protease upon binding with substrates in solution.

References

[1] Weber, D. J., and Rutala, W. A. (2001) Risks and prevention of nosocomial transmission of rare zoonotic diseases, *Clin Infect Dis 32*, 446-456.

[2] Williams, R. J., Al-Busaidy, S., Mehta, F. R., Maupin, G. O., Wagoner, K. D., Al-Awaidy, S., Suleiman, A. J., Khan, A. S., Peters, C. J., and Ksiazek, T. G. (2000) Crimean-congo haemorrhagic fever: a seroepidemiological and tick survey in the Sultanate of Oman, *Trop Med Int Health 5*, 99-106.

[3] Duh, D., Nichol, S. T., Khristova, M. L., Saksida, A., Hafner-Bratkovic, I., Petrovec, M., Dedushaj, I., Ahmeti, S., and Avsic-Zupanc, T. (2008) The complete genome sequence of a Crimean-Congo hemorrhagic fever virus isolated from an endemic region in Kosovo, *Virol J 5*, 7.

[4] Morikawa, S., Saijo, M., and Kurane, I. (2007) Recent progress in molecular biology of Crimean-Congo hemorrhagic fever, *Comp Immunol Microbiol Infect Dis 30*, 375-389.

[5] Hornak, K. E., Lanchy, J. M., and Lodmell, J. S. (2016) RNA Encapsidation and Packaging in the Phleboviruses, *Viruses 8*.

[6] Deyde, V. M., Khristova, M. L., Rollin, P. E., Ksiazek, T. G., and Nichol, S. T. (2006) Crimean-Congo hemorrhagic fever virus genomics and global diversity, *J Virol 80*, 8834-8842.

[7] Ergonul, O., Tuncbilek, S., Baykam, N., Celikbas, A., and Dokuzoguz, B. (2006) Evaluation of serum levels of interleukin (IL)-6, IL-10, and tumor necrosis factor-alpha in patients with Crimean-Congo hemorrhagic fever, *J Infect Dis 193*, 941-944.

[8] Soares-Weiser, K., Thomas, S., Thomson, G., and Garner, P. (2010) Ribavirin for Crimean-Congo hemorrhagic fever: systematic review and meta-analysis, *BMC Infect Dis 10*, 207.

[9] Reyes-Turcu, F. E., Ventii, K. H., and Wilkinson, K. D. (2009) Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes, *Annu Rev Biochem 78*, 363-397.

[10] Suryadinata, R., Roesley, S. N., Yang, G., and Sarcevic, B. (2014) Mechanisms of generating polyubiquitin chains of different topology, *Cells 3*, 674-689.

[11] Heideker, J., and Wertz, I. E. (2015) DUBs, the regulation of cell identity and disease, *Biochem J 467*, 191.

[12] Ventii, K. H., and Wilkinson, K. D. (2008) Protein partners of deubiquitinating enzymes, *Biochem J 414*, 161-175.

[13] Liz, M. A., and Sousa, M. M. (2005) Deciphering cryptic proteases, *Cell Mol Life Sci 62*, 989-1002.

[14] Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D., and Hill, C. P. (1997) Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 A resolution, *EMBO J 16*, 3787-3796.

[15] Andersson, F. I., Werrell, E. F., McMorran, L., Crone, W. J., Das, C., Hsu, S. T., and Jackson, S. E. (2011) The effect of Parkinson's-disease-associated mutations on the deubiquitinating enzyme UCH-L1, *J Mol Biol 407*, 261-272.

[16] Hu, M., Li, P., Song, L., Jeffrey, P. D., Chenova, T. A., Wilkinson, K. D., Cohen, R. E., and Shi, Y. (2005) Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14, *EMBO J 24*, 3747-3756.

[17] Zhong, X., and Pittman, R. N. (2006) Ataxin-3 binds VCP/p97 and regulates retrotranslocation of ERAD substrates, *Hum Mol Genet 15*, 2409-2420.

[18] Winborn, B. J., Travis, S. M., Todi, S. V., Xu, P., Peng, J. M., Cohen, R. E., and Paulson, H. L. (2008) The deubiquitinating enzyme ataxin-3 edits K63-linkages in mixed linkage ubiquitin chains, *Faseb J 22*.

[19] Weeks, S. D., Grasty, K. C., Hernandez-Cuebas, L., and Loll, P. J. (2011) Crystal structure of a Josephin-ubiquitin complex: evolutionary restraints on ataxin-3 deubiquitinating activity, *J Biol Chem 286*, 4555-4565.

[20] Cope, G. A., Suh, G. S., Aravind, L., Schwarz, S. E., Zipursky, S. L., Koonin, E. V., and Deshaies, R. J. (2002) Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of Nedd8 from Cul1, *Science 298*, 608-611.

[21] McCullough, J., Clague, M. J., and Urbe, S. (2004) AMSH is an endosome-associated ubiquitin isopeptidase, *J Cell Biol 166*, 487-492.

[22] Shrestha, R. K., Ronau, J. A., Davies, C. W., Guenette, R. G., Strieter, E. R., Paul, L. N., and Das, C. (2014) Insights into the mechanism of deubiquitination by JAMM deubiquitinases from cocrystal structures of the enzyme with the substrate and product, *Biochemistry 53*, 3199-3217.

[23] Goodrich, J. S., Clouse, K. N., and Schupbach, T. (2004) Hrb27C, Sqd and Otu cooperatively regulate gurken RNA localization and mediate nurse cell chromosome dispersion in Drosophila oogenesis, *Development 131*, 1949-1958.

[24] Komander, D., and Barford, D. (2008) Structure of the A20 OTU domain and mechanistic insights into deubiquitination, *Biochem J 409*, 77-85.

[25] Steinhauer, W. R., Walsh, R. C., and Kalfayan, L. J. (1989) Sequence and structure of the Drosophila melanogaster ovarian tumor gene and generation of an antibody specific for the ovarian tumor protein, *Mol Cell Biol 9*, 5726-5732.

[26] Capodagli, G. C., McKercher, M. A., Baker, E. A., Masters, E. M., Brunzelle, J. S., and Pegan, S. D. (2011) Structural analysis of a viral ovarian tumor domain protease from the Crimean-Congo hemorrhagic fever virus in complex with covalently bonded ubiquitin, *J Virol 85*, 3621-3630.

[27] Lin, S. C., Chung, J. Y., Lamothe, B., Rajashankar, K., Lu, M., Lo, Y. C., Lam, A. Y., Darnay, B. G., and Wu, H. (2008) Molecular basis for the unique deubiquitinating activity of the NF-kappaB inhibitor A20, *J Mol Biol 376*, 526-540.

[28] James, T. W., Frias-Staheli, N., Bacik, J. P., Levingston Macleod, J. M., Khajehpour, M., Garcia-Sastre, A., and Mark, B. L. (2011) Structural basis for the removal of ubiquitin and interferon-stimulated gene 15 by a viral ovarian tumor domain-containing protease, *Proc Natl Acad Sci U S A 108*, 2222-2227.

[29] Eisenmesser, E. Z., Capodagli, G. C., Armstrong, G. S., Holliday, M. J., Isern, N. G., Zhang, F., and Pegan, S. D. (2015) Inherent dynamics within the Crimean-Congo Hemorrhagic fever virus protease are localized to the same region as substrate interactions, *Protein Sci 24*, 651-660.

[30] Baldwin, A. J., and Kay, L. E. (2009) NMR spectroscopy brings invisible protein states into focus, *Nat Chem Biol 5*, 808-814.

[31] Wolynes, P. G. (2005) Recent successes of the energy landscape theory of protein folding and function, *Q. Rev. Biophys. 38*, 405–410.

[32] Henzler-Wildman, K., and Kern, D. (2007) Dynamic personalities of proteins, *Nature 450*, 964-972.

[33] Smock, R. G., and Gierasch, L. M. (2009) Sending signals dynamically, *Science 324*, 198-203.

[34] Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Freer, S. T., and Rose, P. W. (2002) Complexity and simplicity of ligand-macromolecule interactions: the energy landscape perspective, *Curr Opin Struct Biol 12*, 197-203.

[35] Stank, A., Kokh, D. B., Fuller, J. C., and Wade, R. C. (2016) Protein Binding Pocket Dynamics, *Acc Chem Res 49*, 809-815.

[36] Boehr, D. D., Dyson, H. J., and Wright, P. E. (2006) An NMR perspective on enzyme dynamics, *Chem Rev 106*, 3055-3079.

[37] Sullivan, S. M., and Holyoak, T. (2008) Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection, *Proc Natl Acad Sci U S A 105*, 13829-13834.

[38] Boehr, D. D., McElheny, D., Dyson, H. J., and Wright, P. E. (2006) The dynamic energy landscape of dihydrofolate reductase catalysis, *Science 313*, 1638-1642.

[39] Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley, New York ; Chichester.

[40] Cavanagh, J. F., Wayne J.; Palmer, Arthur G. III; Skelton, Nicholas J. (1996) *Protein NMR Spectroscopy: Principles and Practise*, Academic Press, San Diego.

[41] Krishnarjuna, B., Jaipuria, G., Thakur, A., D'Silva, P., and Atreya, H. S. (2011) Amino acid selective unlabeling for sequence specific resonance assignments in proteins, *J Biomol NMR 49*, 39-51.

[42] Goto, N. K., and Kay, L. E. (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy, *Curr Opin Struct Biol 10*, 585-592.

[43] Barnwal, R. P., Rout, A. K., Atreya, H. S., and Chary, K. V. R. (2008) Identification of C-terminal neighbours of amino acid residues without an aliphatic C-13(gamma) supercript

stop as an aid to NMR assignments in proteins, *Journal of Biomolecular Nmr 41*, 191-197.

[44] Atreya, H. S., and Chary, K. V. R. (2000) Amino acid selective 'unlabelling' for residue-specific NMR assignments in proteins, *Curr Sci India 79*, 504-507.

[45] Rasia, R. M., Brutscher, B., and Plevin, M. J. (2012) Selective Isotopic Unlabeling of Proteins Using Metabolic Precursors: Application to NMR Assignment of Intrinsically Disordered Proteins, *Chembiochem 13*, 732-739.

[46] Whitehead, B., Craven, C. J., and Waltho, J. P. (1997) Double and triple resonance NMR methods for protein assignment, *Methods Mol Biol 60*, 29-52.

[47] Rule, G. S., and Hitchens, T. K. (2006) *Fundamentals of protein NMR spectrosopy*, Springer, Dordrecht.

[48] Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) 3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins, *J Magn Reson 89*, 496-514.

[49] Grzesiek, S., and Bax, A. (1992) Improved 3d Triple-Resonance Nmr Techniques Applied to a 31-Kda Protein, *J Magn Reson 96*, 432-440.

[50] Grzesiek, S., and Bax, A. (1992) An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins, *J Magn Reson 99*, 201-207.

[51] Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego.

[52] Wittekind, M., and Mueller, L. (1993) Hncacb, a High-Sensitivity 3D Nmr Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins, *J Magn Reson Ser B 101*, 201-205.

[53] Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M., and Kay, L. E. (1990) Practical Aspects of Proton Carbon Carbon Proton 3-Dimensional Correlation Spectroscopy of C-13-Labeled Proteins, *J Magn Reson 87*, 620-627.

[54] Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) H-1-H-1 Correlation Via Isotropic Mixing of C-13 Magnetization, a New 3-Dimensional Approach for Assigning H-1 and C-13 Spectra of C-13-Enriched Proteins, *J Magn Reson 88*, 425-431.

[55] Weisemann, R., Lohr, F., and Ruterjans, H. (1994) Hncch-Tocsy, a Triple-Resonance Experiment for the Correlation of Backbone C-13(Alpha) and N-15 Resonances with Aliphatic Side-Chain Proton Resonances and for Measuring Vicinal (Co)-C-13,H-1(Beta) Coupling-Constants in Proteins, *Journal of Biomolecular Nmr 4*, 587-593.

[56] Grzesiek, S., Anglister, J., and Bax, A. (1993) Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in C-13/N-15-Enriched Proteins by Isotropic Mixing of C-13 Magnetization, *J Magn Reson Ser B 101*, 114-119.

[57] Morin, S. (2011) A practical guide to protein dynamics from 15N spin relaxation in solution, *Prog Nucl Magn Reson Spectrosc 59*, 245-262.

[58] Abragam, A. (1961) *The principles of nuclear magnetism*, Clarendon Press, Oxford.

[59] Levitt, M. H. (2001) *Spin dynamics : basics of nuclear magnetic resonance*, John Wiley & Sons, Chichester ; New York.

[60] Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M., and Skelton, N. J. (2007) Protein NMR Spectroscopy: Principles and Practice, 2nd Edition, *Protein Nmr Spectroscopy: Principles and Practice, 2nd Edition*, 1-888.

[61] Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity, *J Am Chem Soc 114*, 10663-10665.

[62] Millet, O., Loria, J. P., Kroenke, C. D., Pons, M., and Palmer, A. G. (2000) The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale, *J Am Chem Soc 122*, 2867-2877.

[63] Hansen, D. F., Yang, D. W., Feng, H. Q., Zhou, Z., Wiesner, S., Bai, Y. W., and Kay, L. E. (2007) An exchange-free measure of N-15 transverse relaxation: An NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange, *J Am Chem Soc 129*, 11468-11479.

[64] Hahn, E. L. (1950) Spin echoes, *Phys. Rev. 80* 580–594.

[65] Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) Spectral density function mapping using 15N relaxation data exclusively, *J Biomol NMR 6*, 153-162.

[66] Lefevre, J. F., Dayie, K. T., Peng, J. W., and Wagner, G. (1996) Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions, *Biochemistry 35*, 2674-2686.

[67] Fischer, M. W. F., Majumdar, A., and Zuiderweg, E. R. P. (1998) Protein NMR relaxation: theory, applications and outlook, *Prog Nucl Mag Res Sp 33*, 207-272.

[68] Krizova, H., Zidek, L., Stone, M. J., Novotny, M. V., and Sklenar, V. (2004) Temperature-dependent spectral density analysis applied to monitoring backbone dynamics of major urinary protein-I complexed with the pheromone 2- sec-butyl-4,5-dihydrothiazole, *J Biomol NMR 28*, 369-384.

[69] Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity, *J Am Chem Soc 104*, 4546-4559.

[70] Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results, *J Am Chem Soc 104*, 4559-4570.

[71] Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) Deviations from the Simple 2-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic-Relaxation of Proteins, *J Am Chem Soc 112*, 4989-4991.

[72] d'Auvergne, E. J., and Gooley, P. R. (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces, *Journal of Biomolecular Nmr 40*, 107-119.

[73] Bloembergen, N., Purcell, E. M., and Pound, R. V. (1948) Relaxation Effects in Nuclear Magnetic Resonance Absorption, *Phys Rev 73*, 679-712.

[74] Fushman, D., Cahill, S., and Cowburn, D. (1997) The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of 15N relaxation with monomer/dimer equilibration, *J Mol Biol 266*, 173-194.

[75] d'Auvergne, E. J. (2006) Protein dynamics : a study of the model-free analysis of NMR relaxation data, pp xxvi, 353 p., bound, University of Melbourne.

[76] Orekhov, V. Y., Korzhnev, D. M., Diercks, T., Kessler, H., and Arseniev, A. S. (1999) H-1-N-15 NMR dynamic study of an isolated alpha-helical peptide (1-36)- bacteriorhodopsin reveals the equilibrium helix-coil transitions, *Journal of Biomolecular Nmr 14*, 345-356.

[77] Korzhnev, D. M., Billeter, M., Arseniev, A. S., and Orekhov, V. Y. (2001) NMR studies of Brownian tumbling and internal motions in proteins, *Prog Nucl Mag Res Sp 38*, 197-266.

[78] Zhuravleva, A. V., Korzhnev, D. M., Kupce, E., Arseniev, A. S., Billeter, M., and Orekhov, V. Y. (2004) Gated electron transfers and electron pathways in azurin: A NMR dynamic study at multiple fields and temperatures, *Journal of Molecular Biology 342*, 1599-1611.

[79] d'Auvergne, E. J., and Gooley, P. R. (2006) Model-free model elimination: a new step in the model-free dynamic analysis of NMR relaxation data, *J Biomol NMR 35*, 117-135.

[80] Devore, J. L. (1982) *Probability and statistics for engineering and the sciences*, Brooks/Cole Pub. Co., Monterey, Calif.

[81] Mandel, A. M., Akke, M., and Palmer, A. G., 3rd. (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme, *J Mol Biol 246*, 144-163.

[82] Palmer, A. G., Rance, M., and Wright, P. E. (1991) Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-12 Heteronuclear Nmr-Spectroscopy, *J Am Chem Soc 113*, 4371-4380.

[83] Cole, R., and Loria, J. P. (2003) FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data, *J Biomol NMR 26*, 203-213.

[84] Orekhov, V. Y., Nolde, D. E., Golovanov, A. P., Korzhnev, D. M., and Arseniev, A. S. (1995) Processing of heteronuclear NMR relaxation data with the new software DASHA, *Appl Magn Reson 9*, 581-588.

[85] Spyracopoulos, L. (2006) A suite of Mathematica notebooks for the analysis of protein main chain 15N NMR relaxation data, *J Biomol NMR 36*, 215-224.

[86] Palmer, A. G., Williams, J., and McDermott, A. (1996) Nuclear magnetic resonance studies of biopolymer dynamics, *J Phys Chem-Us 100*, 13293-13310.

[87] Wrabl, J. O., Shortle, D., and Woolf, T. B. (2000) Correlation between changes in nuclear magnetic resonance order parameters and conformational entropy: molecular dynamics simulations of native and denatured staphylococcal nuclease, *Proteins 38*, 123-133.

[88] Carr, H. Y., and Purcell, E. M. (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments, *Phys Rev 94*, 630-638.

[89] Lee, A. L., Kinnear, S. A., and Wand, A. J. (2000) Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex, *Nat Struct Biol 7*, 72-77.

[90] Kay, L. E., Muhandiram, D. R., Wolf, G., Shoelson, S. E., and Forman-Kay, J. D. (1998) Correlation between binding and dynamics at SH2 domain interfaces, *Nat Struct Biol 5*, 156-163.

[91] Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G., 3rd. (1999) Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA, *J Mol Biol 285*, 2133-2146.

[92] Haber, E., and Anfinsen, C. B. (1962) Side-chain interactions governing the pairing of half-cystine residues in ribonuclease, *J Biol Chem 237*, 1839-1844.

[93] Palmer AG 3rd1, K. C., Loria JP. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules., *Methods Enzymol 339*, 204-238.

[94] Hansen, D. F., Vallurupalli, P., and Kay, L. E. (2008) An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins, *J Phys Chem B 112*, 5898-5904.

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

[96] Lee, W., Westler, W. M., Bahrami, A., Eghbalnia, H. R., and Markley, J. L. (2009) PINE-SPARKY: graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy, *Bioinformatics 25*, 2085-2087.

[97] Bahrami, A., Assadi, A. H., Markley, J. L., and Eghbalnia, H. R. (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy, *PLoS Comput Biol 5*, e1000307.

[98] Ikura, M., Kay, L. E., and Bax, a. (1990) A Novel-Approach for Sequential Assignment of H-1, C-13, and N-15 Spectra of Larger Proteins - Heteronuclear Triple-Resonance 3-Dimensional Nmr-Spectroscopy - Application to Calmodulin, *Biochemistry 29*, 4659-4667.

[99] Yamazaki, T., Lee, W., Revington, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H., and Kay, L. E. (1994) An Hnca Pulse Scheme for the Backbone Assignment of N-15,C-13,H-2-Labeled Proteins - Application to a 37-Kda Trp Repressor DNA Complex, *J Am Chem Soc 116*, 6464-6465.

[100] Grzesiek, S., and Bax, A. (1992) Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr, *J Am Chem Soc 114*, 6291-6293.

[101] Vuister, G. W., and Bax, A. (1993) Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins, *J Am Chem Soc 115*, 7772-7777.

[102] Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy, *Bioinformatics 31*, 1325-1327.

[103] Goddard, T. D. a. K., D. G. . (2008) Sparky 3, *University of California, San Francisco*.

[104] Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts, *J Am Chem Soc 123*, 2970-2978.

[105] Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation, *Protein Sci 15*, 2795-2804.

[106] Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A. G., Dobson, C. M., and Smith, L. J. (1997) NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein, *Journal of Molecular Biology 274*, 152-159.

[107] Gagne, D., Charest, L. A., Morin, S., Kovrigin, E. L., and Doucet, N. (2012) Conservation of flexible residue clusters among structural and functional enzyme homologues, *J Biol Chem 287*, 44289-44300.

[108] d'Auvergne, E. J., and Gooley, P. R. (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces, *J Biomol NMR 40*, 107-119.

[109] Spyracopoulos, L. (2006) A suite of Mathematica notebooks for the analysis of protein main chain 15N NMR relaxation data, *Journal of Biomolecular NMR 36*, 215-224.

[110] Yang, Z. R., Thomson, R., McNeil, P., and Esnouf, R. M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins, *Bioinformatics 21*, 3369-3376.

[111] Shojania, S., and O'Neil, J. D. (2006) HIV-1 Tat is a natively unfolded protein - The solution conformation and dynamics of reduced HIV-1 Tat-(1-72) by NMR spectroscopy, *Journal of Biological Chemistry 281*, 8347-8356.

[112] Spyracopoulos, L., Yee, A. A., and ONeil, J. D. J. (1996) Backbone dynamics of an alamethicin in methanol and aqueous detergent solution determined by heteronuclear H-1-N-15 NMR spectroscopy, *Journal of Biomolecular Nmr 7*, 283-294.

[113] Schurr, J. M., Babcock, H. P., and Fujimoto, B. S. (1994) A Test of the Model-Free Formulas - Effects of Anisotropic Rotational Diffusion and Dimerization, *J Magn Reson Ser B 105*, 211-224.

[114] Laskowski, R. A., MacArthur, M. W., and Thornton, J. M. (1998) Validation of protein models derived from experiment, *Curr Opin Struct Biol 8*, 631-639.

[115] d'Auvergne, E. J., and Gooley, P. R. (2003) The use of model selection in the model-free analysis of protein dynamics, *J Biomol NMR 25*, 25-39.

[116] Neudecker, P., Lundstrom, P., and Kay, L. E. (2009) Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding, *Biophys J 96*, 2045-2054.

[117] Mittermaier, A. K., and Kay, L. E. (2009) Observing biological dynamics at atomic resolution using NMR, *Trends Biochem Sci 34*, 601-611.

[118] Bieri, M., and Gooley, P. R. (2011) Automated NMR relaxation dispersion data analysis using NESSY, *BMC Bioinformatics 12*, 421.

[119] Meneses, E., and Mittermaier, A. (2014) Electrostatic interactions in the binding pathway of a transient protein complex studied by NMR and isothermal titration calorimetry, *J Biol Chem 289*, 27911-27923.

[120] James, T., and University of Manitoba. Department of Microbiology. (2010) A structural examination of the Crimean-Congo Hemorrhagic Fever Virus Otu protease domain in the presence of the Ubiquitin and ISG15 substrates, p 142 leaves, University of Manitoba, October 2010.

[121] Doonan, S. (1996) Protein purification protocols. General strategies, *Methods Mol Biol 59*, 1-16.

[122] Andrews, P. (1965) The gel-filtration behaviour of proteins related to their molecular weights over a wide range, *Biochem J 96*, 595-606.

[123] Park, S. J., Kostic, M., and Dyson, H. J. (2011) Dynamic Interaction of Hsp90 with Its Client Protein p53, *J Mol Biol 411*, 158-173.

[124] Park, S. J., Borin, B. N., Martinez-Yamout, M. A., and Dyson, H. J. (2011) The client protein p53 adopts a molten globule-like state in the presence of Hsp90, *Nat Struct Mol Biol 18*, 537-541.

[125] Ikura, M., Kay, L. E., and Bax, A. (1990) A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin, *Biochemistry 29*, 4659-4667.

[126] Schwarzinger, S. K., G.J; Foss, T.R.; Chung.J; Wright,P.E; Dyson, H.J. (2001) Sequence-dependent correction of random coil of NMR chemical shifts., *J. Am. Chem. Soc. 123*, 2970-2978.

[127] Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure, *J Mol Biol 222*, 311-333.

[128] Wang, Y., and Jardetzky, O. (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data, *Protein Sci 11*, 852-861.

[129] Wishart, D. S., and Sykes, B. D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data, *J Biomol NMR 4*, 171-180.

[130] Pauling, L., Corey, R. B., and Branson, H. R. (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain, *Proc Natl Acad Sci U S A 37*, 205-211.

[131] Hafsa, N. E., Arndt, D., and Wishart, D. S. (2015) CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts, *Nucleic Acids Res 43*, W370-377.

[132] Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M., and Dobson, C. M. (1996) Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations, *J Mol Biol 255*, 494-506.

[133] Cavanagh, J. F., Wayne J.; palmer, Arthur G. III; Skelton, Nicholas J. (1996) Protein NMR Spectroscopy: Principles and Practise, *Academic Press, San Diego.*

[134] Farina, B., Pirone, L., Russo, L., Viparelli, F., Doti, N., Pedone, C., Pedone, E. M., and Fattorusso, R. (2010) NMR backbone dynamics studies of human PED/PEA-15 outline protein functional sites, *FEBS J 277*, 4229-4240.

[135] Buevich, a. V., Shinde, U. P., Inouye, M., and Baum, J. (2001) Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies, *Journal of Biomolecular Nmr 20*, 233-249.

[136] Loria, J. P., Berlow, R. B., and Watt, E. D. (2008) Characterization of enzyme motions by solution NMR relaxation dispersion, *Acc Chem Res 41*, 214-221.

[137] Edelmann, M. J., Nicholson, B., and Kessler, B. M. (2011) Pharmacological targets in the ubiquitin system offer new ways of treating cancer, neurodegenerative disorders and infectious diseases, *Expert Rev Mol Med 13*, e35.

[138] Kaufman, J. M., Lapauw, B., and Goemaere, S. (2014) Current and future treatments of osteoporosis in men, *Best Pract Res Clin Endocrinol Metab 28*, 871-884.

[139] Turk, D., Turk, B., and Turk, V. (2003) Papain-like lysosomal cysteine proteases and their inhibitors: drug discovery targets?, *Biochem Soc Symp*, 15-30.

[140] Harper, S., and Speicher, D. W. (2011) Purification of proteins fused to glutathione S-transferase, *Methods Mol Biol 681*, 259-280.

[141] Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase, *Gene 67*, 31-40.

[142] Madadlou, A., O'Sullivan, S., and Sheehan, D. (2011) Fast protein liquid chromatography, *Methods Mol Biol 681*, 439-447.

[143] Hagn, F., Lagleder, S., Retzlaff, M., Rohrberg, J., Demmer, O., Richter, K., Buchner, J., and Kessler, H. (2011) Structural analysis of the interaction between Hsp90 and the tumor suppressor protein p53, *Nat Struct Mol Biol 18*, 1086-1093.

[144] B. Krishnarjuna, G. J., Anushikha Thakur, Patrick D'Silva,Hanudatta S. Atreya. (2011) Amino acid selective unlabeling for sequence specific resonance assignments in proteins, *J Biomol NMR 49*, 39–51.

[145] Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., and Dahlquist, F. W. (1989) Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance, *Methods Enzymol 177*, 44-73.

[146] Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, H. J. (2001) NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding, *Biochemistry 40*, 3561-3571.

[147] Teilum, K., Kragelund, B. B., and Poulsen, F. M. (2002) Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling, *J Mol Biol 324*, 349-357.

[148] Garcia, P., Serrano, L., Durand, D., Rico, M., and Bruix, M. (2001) NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation, *Protein Sci 10*, 1100-1112.

[149] Wishart, D. S., and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination, *Methods Enzymol 239*, 363-392.

[150] Frank, M. K., Clore, G. M., and Gronenborn, A. M. (1995) Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy, *Protein Sci 4*, 2605-2615.

[151] Bai, Y., Chung, J., Dyson, H. J., and Wright, P. E. (2001) Structural and dynamic characterization of an unfolded state of poplar apo-plastocyanin formed under nondenaturing conditions, *Protein Sci 10*, 1056-1066.

[152] Lin SC, C. J., Lamothe B, Rajashankar K, Lu M, et al. . (2008) Molecular basis for the unique deubiquitinating activity of the NF-kappaB inhibitor A20., *J Mol Biol 376*, 526–540.

[153] Glenn C. Capodagli, M. A. M., Joseph S. Brunzelle, Erica A. Baker, Scott D. Pegan Emily M. Masters,. (2011) Structural Analysis of a Viral Ovarian Tumor Domain Protease from the Crimean-Congo Hemorrhagic Fever Virus in Complex with Covalently Bonded Ubiquitin, *JOURNAL OF VIROLOGY 85*, 3621–3630

Appendix

Resonance Assignments for OTU $^{(1-169)}$

Table A: The resonance assignment list of OTU determined at pH 6.5 and 293 K.

Position	Residue Type	C' (ppm)	C^{α} (ppm)	C^{β} (ppm)	H^{N} (ppm)	H^{α}	N (ppm)
1	M (MET)	176.9	61.81	34.6	8.915	4.42	124
2	D (ASP)	177.4	56.53	40.84	8.233	4.41	118.1
3	F (PHE)	178	61.63	39.33	8.001	4.24	121.5
4	L (LEU)	178.9	56.71	41.82	7.523	3.43	117.6
5	R (ARG)	176.9	58.37	30.63	7.907	4.22	116.1
6	S (SER)	173.6	58.24	63.93	7.423	4.47	112.9
7	L (LEU)	175.8	55.76	42.36	6.484	3.57	121.5
8	D (ASP)	175	53.2	41.6	8.374	4.65	122.8
9	W (TRP)	176.3	56.06	30.76	8.37	4.76	125.3
10	T (THR)	174	62.09	70.47	9.541	4.61	121.8
11	Q (GLN)	175.4	57.08	29.76	9.353	4.15	129.4
12	V (VAL)	175.9	64.49	33.44	8.74	4.06	128.3
13	I (ILE)	174.1	59.81	42.07	7.475	4.24	116.5
14	A (ALA)	178.7	54.67	17.52	8.259	3.85	124.4
15	G (GLY)	173.2	45.84		8.41	-4	110
16	Q (GLN)	173.8	54.7	31.91	8.187	5.32	121.9
17	Y (TYR)	174.7	55.5	44.18	9.206	5.67	122.9
18	V (VAL)	174.9	59.76	36.48	9.358	5.59	114.9
19	S (SER)	173.4	58.17	65.9	8.508	4.48	115.9
20	N (ASN)	174.3	52.15	40.03	8.843	5.24	120.4
22	R (ARG)	175	55.46	30.62	8.26	4.84	116.4
23	F (PHE)	171.8	55.77	39.47	7.133	4.69	117.9
24	N (ASN)	176.7	52.31	38.87	9.058	5.08	119.6
25	I (ILE)	177.4	62.56	37.64	8.696	4.52	127
26	S (SER)	178.2	60.26	62.92	8.594	4.77	113.4
27	D (ASP)	175.8	56.14	40.96	7.739	4.31	119.8
28	Y (TYR)	173.2	59.16	43.55	7.09	4.52	111.9
29	F (PHE)	174.9	56.49	43.85	8.24	5.35	114.3

Position	C' (ppm)	Cα (ppm)	Cβ (ppm)	HN (ppm)	H^N (ppm)	Hα	N (ppm)
30	E (GLU)	175.5	54.8	31.99	9.501	4.48	120.6
31	I (ILE)	175.8	61.83	38.82	8.654	4.26	125.8
32	V (VAL)	176.6	61.59	30.48	9.226	4.25	131.6
33	R (ARG)	176.6	57.14	29.97	8.688	4.14	128.1
34	Q (GLN)	173.6	52.27	27.75	9.476	4.67	123.6
35	P (PRO)	177.1	61.87	32.59		4.16	
36	G (GLY)	175.6	46.82		8.641	-4	111.6
37	D (ASP)	177.4	51.86	40.36	7.937	4.97	121.3
38	G (GLY)	174.9	44.76		8.295	3.17	106
39	N (ASN)	177.4	56.6	38.63	8.752	4.97	117.4
40	C (CYS)	175.5	54.84	32.8	7.581	4.04	113.5
41	F (PHE)	173.4	63.12	40.24	9.388	5.31	126.3
42	Y (TYR)	178.8	60.69	39.16	6.717	3.55	117.8
43	H (HIS)	177.7	59.79	32.17	8.212	4.41	118.4
44	S (SER)	174.6	63.6	65.65	7.814	3.97	113.9
45	I (ILE)	180.2	63.79	41.58	7.054	3.42	117.9
46	A (ALA)	178.3	56.12	18.49	8.4	3.76	125
47	E (GLU)	176.8	58.52	30.5	8.307	1.08	116.2
48	L (LEU)	178.1	54.81	44.41	6.721	0.83	112.6
49	T (THR)	174.6	62.23	71.83	7.702	4.67	106.7
50	M (MET)	174.4	62.21	42.23	8.099	5.01	122.4
52	N (ASN)	174	53.28	37.59	8.704	4.42	116.9
53	K (LYS)	177	57.5	32.75	7.758	4.13	120.3
54	T (THR)	176.5	60.32	73.26	9.374	4.92	118.5
55	D (ASP)	175.9	56.64	39.98	8.92	4.43	119.7
56	H (HIS)	177.1	55.49	31.33	8.317	5	115.1
57	S (SER)	175.7	50.97	61.26	7.979	2.39	120.8
58	H (HIS)	176.1	60.48	35.31	9.52	4.43	118.4
59	H (HIS)	176.2	60.35	36.54	9.526	4.46	118.4
60	Y (TYR)	178.3	57.43	29.53	6.37	3.54	119.7
61	I (ILE)	175.4	61.21	38.1	7.998	4.24	120.3
62	K (LYS)	173.9	55.26	33.89	8.038	4.2	122.4
63	R (ARG)	178	56.99	32.64	7.812	4.23	121.4
64	L (LEU)	178.7	57.61	42.26	7.87	4.08	121.6
65	T (THR)	176	44.37	67.65	8.089	3.72	118.7
66	E (GLU)	177.4	59.47	28.86	8.036	3.75	124.5
67	S (SER)	177.8	61.68	62.39	7.789	4	113.8
68	A (ALA)	179.7	54.26	20.26	8.746	4.18	123.3

Position	Residue Type	C' (ppm)	Cα (ppm)	Cβ (ppm)	HN (ppm)	Hα	N (ppm)
69	A (ALA)	180.1	54.75	18.87	9.129	4.78	121.6
70	R (ARG)	177.4	60.25	29.42	8.053	3.93	115.4
71	K (LYS)	178.8	58.9	33.59	7.062	4.05	115.5
72	Y (TYR)	177.4	59.38	40.8	8.664	4.83	115.1
73	Y (TYR)	176.5	63.55	40.54	9.122	3.91	122.1
74	Q (GLN)	176.7	57.97	28.46	8.511	3.84	112.1
75	E (GLU)	176.7	55.29	30.92	7.509	4.44	115.3
76	E (GLU)	176.6	54.52	29.37	7.489	4.2	123.1
77	P (PRO)	180	65.56	32.08		4.19	
78	E (GLU)	178.3	59.01	29.56	9.047	4.02	115.3
79	A (ALA)	180.2	54.77	18.55	7.321	3.79	121.5
80	R (ARG)	177.7	57.72	30.01	7.222	3.97	116.3
81	L (LEU)	178.5	56.51	41.57	7.132	4.13	117.5
82	V (VAL)	178.5	64.68	32.29	7.525	3.88	118.4
83	G (GLY)	174.1	46.61		7.724	-3	105
84	L (LEU)	178.9	53.6	44.63	6.961	4.78	117.7
85	S (SER)	173.8	57.84	64.49	9.656	4.53	122.3
86	L (LEU)	177.8	58.54	40.89	9.018	3.66	122.4
87	E (GLU)	177.8	60.25	28.83	8.645	3.85	115.1
88	D (ASP)	178.9	56.82	40.3	7.644	4.42	120.4
89	Y (TYR)	176.8	61.78	37.03	8.937	3.85	124.1
90	L (LEU)	178	57.98	42.01	8.564	3.21	119.1
91	K (LYS)	178.9	59.08	32.53	7.203	4.13	115.4
92	R (ARG)	177.8	59.2	30.4	7.15	3.92	118.9
93	M (MET)	179.5	58.93	30.7	8.676	2.52	120.8
94	L (LEU)	177.8	56.12	42	8.018	3.7	115.2
95	S (SER)	173.8	59.42	63.69	7.576	4.27	118.6
96	D (ASP)	176.8	55.61	41.07	8.601	2.5	123.5
97	N (ASN)	173.5	56.64	39.95	8.581	2.64	116.4
98	E (GLU)	176.4	63.09	32.04	8.31	4.7	114.9
99	W (TRP)	178	55.21	42.68	8.405	4.31	122.2
100	G (GLY)	172.4	43.38		8.377	-3	109.2
101	S (SER)	173.4	57.05	65.72	8.478	4.22	116.3
102	T (THR)	176.4	65.18	67.33	8.159	4.76	110.6
103	L (LEU)	177	58.43	41.03	7.979	4.2	125.8
104	E (GLU)	178.5	57.32	42.53	7.499	3.63	118.3
105	A (ALA)	176.4	54.63	19.44	7.169	4.58	116.6
106	S (SER)	179	60.41	62.62	7.09	3.94	109.7

Position	Residue Type	C' (ppm)	C^{α} (ppm)	C^{β} (ppm)	H^{N} (ppm)	H^{α}	N (ppm)
107	M (MET)	177.8	60.32	33.91	7.431	3.75	115.9
108	L (LEU)	178.8	57.87	42.72	8.911	3.85	119.5
109	A (ALA)	180.3	56.12	18.02	8.225	4.29	122.2
110	K (LYS)	179.1	59.32	32.49	6.519	3.92	114
111	E (GLU)	178.3	58.38	31.92	8.22	4.55	121
112	M (MET)	176.8	56.93	35.06	9.122	4.24	113.7
113	G (GLY)	175.1	47.13		7.604	-3	110
114	I (ILE)	174.3	59.34	41.72	7.859	4.69	113.1
115	T (THR)	173.6	61.73	70.08	8.274	5.37	117.5
116	I (ILE)	174.3	59.8	42.07	8.802	4.92	126.6
117	I (ILE)	174.5	59.42	40.49	8.983	4.76	127.1
118	I (ILE)	175.4	60.38	40	9.025	4.52	124
119	W (TRP)	175.6	55.13	31	9.278	5.51	132.6
120	T (THR)	174.3	63.03	70.62	9.232	5.27	121.4
121	V (VAL)	175	59.34	34.87	8.185	4.76	120.2
122	A (ALA)	177	51.88	19.45	8.738	4.56	127.2
123	A (ALA)	177	52.93	19.09	8.228	4.31	122.4
124	S (SER)	173.7	57.77	63.59	8.123	4.27	113.1
125	D (ASP)	174.6	54.78	40.72	8.328	4.48	118.1
126	E (GLU)	175.7	56.11	30.73	7.763	4.38	119.5
127	V (VAL)	175.6	63.64	33.01	8.541	3.81	126.4
128	E (GLU)	176.1	56.54	32.23	8.733	4.67	122.3
129	A (ALA)	175	52.42	22.95	7.682	4.95	119.4
130	G (GLY)	172.1	45.3		8.536	-3	106.6
131	I (ILE)	171	59.9	41.54	8.341	4.69	123.3
132	K (LYS)	172.9	54.17	37.51	8.956	4.68	128.2
133	F (PHE)	174.1	56.45	39.62	9.242	5.01	126.6
134	G (GLY)	174.8	43.44		8.147	-4	109.4
135	D (ASP)	176.7	55.26	41.7	8.354	4.24	116.8
136	G (GLY)	172.3	44.3		8.422	-3	110.4
137	D (ASP)	176.6	51.86	43.42	7.704	4.56	117.5
138	V (VAL)	175.5	64.62	32.48	8.199	1.81	115.8
139	F (PHE)	177.4	60.76	38.54	7.248	4.21	115.6
140	T (THR)	174.3	60.92	69.88	7.694	4.64	108.2
141	A (ALA)	176.5	51.47	21.4	7.275	4.43	124.5
142	V (VAL)	173.7	62.83	32.31	8.713	4.07	121.3
143	N (ASN)	171.8	53.83	43.35	9.126	5.27	127.5
144	L (LEU)	174.9	52.61	45.81	9.507	4.97	120.1

Position	Residue Type	C' (ppm)	C^{α} (ppm)	C^{β} (ppm)	H^{N} (ppm)	H^{α}	N (ppm)
145	L (LEU)	175.2	52.82	45.72	8.695	5.1	122.4
146	H (HIS)	172.6	52.74	31.4	9.862	0.93	130.8
147	S (SER)	173.4	56.69	66	8.652	4.73	121.1
148	G (GLY)	174.8	46.33		8.5	-3	107.6
149	Q (GLN)	179	56.64	29.64	8.455	5.55	118.2
150	T (THR)	173.1	60.88	42.84	8.187	4.76	108.1
151	H (HIS)	176.4	53.26	70.85	7.842	5.17	120
152	F (PHE)	172.6	56.6	32.61	7.717	4.15	127.9
153	D (ASP)	175.5	53.35	43.45	8.451	5.04	113.8
154	A (ALA)	174.8	51.27	47.02	8.884	5.2	125.3
155	L (LEU)	175.3	52.87	21.84	7.85	5.43	118.8
156	R (ARG)	176	54.81	44.73	8.744	4.98	118.3
157	I (ILE)	176	62.43	33.74	9.294	4.08	127.6
158	L (LEU)	172.4	54.21	38.84	8.299	4.43	128.8
160	Q (GLN)	175.1	57.08	40.65	8.34	3.56	117.1
161	F (PHE)	174.8	57.47	27.77	7.961	4.85	119
162	E (GLU)	176.6	56.9	39.42	7.644	4.37	119.4
163	T (THR)	174	61.82	30.64	8.068	4.24	114.4
164	D (ASP)	176.5	55.53	69.85	8.368	4.67	123.1
165	T (THR)	174.8	61.88	41.27	8.11	4.23	115.2
166	R (ARG)	176.3	56.35	69.37	8.267	4.26	123.2
167	E (GLU)	175.8	56.54	30.61	8.391	4.1	121.8
168	A (ALA)	176.9	52.84	30.39	8.217	4.31	125.9
169	L (LEU)	173.5	57.64	19.17	7.983	4.32	115.9