ABSTRACT

Title of Document: APPROACHES TO CHARACTERIZE UBIQUITIN CARBOXY-TERMINAL HYDROLASE-L1 BY PROTEOLYTIC MAPPING

NATASHA NICOLE SMITH, M.S., 2005

Directed By: Professor Catherine Fenselau
Department of Chemistry and Biochemistry

Ubiquitin carboxy-terminal hydrolase-L1 is a member of a class of deubiquitinating enzymes that functions to bind and stabilize ubiquitin, and, thereby, plays a role in the protein degradation of targeted substrates. It is of interest to the research communities in cancer and mental health because of it's suggested roles in neurological disorders and some carcinomas. Here, the recombinant form of this protein is sequenced and characterized using proteomic and mass spectrometry methods. Sixty-nine percent sequence coverage of recombinant UCH-L1 is achieved.

APPROACHES TO CHARACTERIZE UBIQUITIN CARBOXY-TERMINAL HYDROLASE-L1 BY PROTEOLYTIC MAPPING

By

Natasha Nicole Smith

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2005

Advisory Committee:
Professor Catherine Fenselau, Chair
Professor Neil Blough
Assistant Professor Sang Bok Lee

Dedication

For Mom, Grandma, Grandpa, and Nate,

for being the four corners of my world;

and to Him who is the center.

Acknowledgements

Though there are many people who helped make this accomplishment possible for me, I would like to extend very special thank you's to:

Mom, for not one moment of doubt;
Grandma and Grandpa for always being excited about whatever I try to do;

Dr. Fenselau for allowing me to do this work in your lab;

Kelli and Akilah for 12 years of support and good times… ***real*** *good times;*
Faith for answers, advice, Moe's, and laughter through tears;
Rachael and Sunny for support, friendship, and chicken-dip;
Steve for the happiest of happy hours;
Scott for putting up with the "banter";
Shawquia for understanding exactly what it's like;

And to Vince, my fourth-quarter motivation… glad you could make it.

Table of Contents

List of Figures

List of Tables

Chapter 1: INTRODUCTION

The small, yet highly conserved ubiquitin polypeptide can be found in all eukaryotes that have been studied thus far. Protein ubiquitination is a post-translational modification involved in many cellular processes including metabolic regulation and ATP-dependent protein degradation. The latter is of great importance as it allows removal of damaged or abnormal proteins whose accumulation could be detrimental to the cell.

The ubiquitin-proteasome system is responsible for the degradation of many cellular proteins. This non-lysosomal pathway consists of proteins involving a wide range of functions that act cooperatively on proteins that have been tagged by ubiquitin degradation. Ubiquitinating enzymes, E1 (ubiquitin activation), E2 (ubiquitin conjugation), and E3 (ubiquitin ligation), are responsible for the covalent attachment of ubiquitin to targeted cellular substrates in preparation for degradation via the 26S proteosomal pathway [1]. Comparably, deubiquitinating enzymes function to remove ubiquitin from these protein substrates so the ubiquitin may be regenerated and used for other condemned proteins [2, 3]. Deubiquitinating enzymes are grouped into two families: ubiquitin proteases (UBPs) and ubiquitin carboxy-terminal hydrolases (UCHs). The work presented here attempts to obtain sequence information of one UCH isozyme: UCH-L1.

Highly conserved and homologous within eukaryotes, the mammalian UCH family consists of three isozymes, L1, L2, and L3. Isozymes L1 and L3 have been found in human cells. These enzymes are cytoplasmic thiol

proteases that selectively hydrolyze the peptide bond at the carboxy-terminal glycine of ubiquitin molecules [4, 5]. UCH-L3 is found in the heart, lung and brain, while UCH-L1 is not as widely distributed. Alternatively labeled neuron cytoplasmic protein 9.5, or PGP9.5, UCH-L1 is localized in neuronal cytoplasm, comprising 1-2% of total soluble brain proteins, making it one of the most abundant brain proteins [6, 7]. It is also found in the testis [8-11]. *In vivo*, UCH-L1 is associated with monoubiquitin and elevates the level of monoubiquitin in neuron [5]. There are also reports of UCH-L1 acting as an ubiquitin ligase under dimerization conditions [12, 13]. According to the gene sequence, UCH-L1 consists of 223 amino acids, and gel electrophoresis measurements show the protein's migration corresponding to approximately 24.8 kDa.

Elucidation of the exon sequences of the gene encoding UCH-L1 was completed followed by publication of the intron sequences flanking those exons [14]. It was found that there are nine coding exons spanning a 10 kb range. The gene has been mapped to chromosome locus 4p14 by PCR analysis [15]. Mutagenesis and kinetic experiments identify cysteine 90 and histidine 161 as the active site nucleophile and general base for catalysis [16, 17]. The same experiments also found that UCH-L1 binds the intact ubiuqitin C-terminus, but not the amino terminus.

UCH-L1 is of growing interest in the proteomics research community because of its implications in various neurodegenerative diseases including Parkinson's disease (PD) and Alzheimer's disease (AD) [12, 13, 18-31]. The

protein is found in Lewy bodies and neurofibrillary tangles, which are pathogenic markers for PD and AD, respectively. The degradation of some protein aggregates by the ubiquitin proteasome system (UPS) is imperative for mental health. As a deubiquitinating enzyme, UCH L1 possesses an enzymatic activity that contributes to the stability of the UPS. UCH-L1 gene mutants have been linked to Parkinson's disease via an isoleucine-methionine mutation at residue 93. Research has shown that this I93M mutation decreases the catalytic activity of UCH-L1 *in vitro*. This in turn may lead to impairment of the UPS causing neuronal polyubiquitinated inclusions to form in excess. Without access to the degradation pathway, they are not subjected to normal processing, leading to the loss of functioning neurons and diminished mental health, as is the case for certain types of Parkinsonism and Alzheimer's disease. In contrast, an S18Y polymorphism has been linked to decreased susceptibility to PD in certain populations [19, 23, 32-35].

Researchers have also reported high UCH-L1 expression in different types of cancers. Hanash et. al [36] use a proteomic approach to detect UCH-L1 (referred to as PGP9.5 in their report) in small cell carcinoma, adenocarcinoma, squamous cell carcinoma, and neuroendocrine differentiated adernocarcinoma and propose it is a biomarker. Other studies suggest that UCH-L1 may be associated with lung cancers [37-41]. Hathout et al. [42] have also reported an increased expression of UCH-L1 protein in drug-resistant and metastatic cancer cell lines. These associations of UCH-

L1 with two very different diseases suggest that mutation, processing, and/or post-translational modifications may be important in directing its functions.

Post-translational Modifications

Protein modifications are widespread in protein expression. Some derivitizations arise as artifacts of sample preparation, such as oxidation; other derivitizations may be deliberately introduced, such as cysteine modifications to deter disulfide bridging. However, a majority of modifications occur after the translation process- after the ribosome has synthesized the protein. These modifications include enzymatic cleavage of a precursor protein or the attachment of biochemical functional groups to individual amino acid residues. Proteolytic cleavage can activate inactive proproteins, such as the activation of trypsin from trypsinogen.

Common covalent post-translational modifications that involve the attachment of functional groups include phosphorylation, glycosylation, acetylation, and methylation, among many others.

Phosphorylations, the addition of a phosphate group, are found on serine, tyrosine, and threonine residues. Phosphorylation is a reversible modification that is catalyzed by protein kinases, and removed by protein phosphatases. Protein phosphorylation has been found to be an activating switch for many cellular processes including signaling and cell growth. One example of great physiological relevance is the release of glucose, the major

source of energy in the body: glycogen phosphorylase is activated by phosphorylation to produce glucose from glycogen molecules.

Glycosylation, another dominant post-translational modification, is the attachment of saccharide groups to proteins. Glycosylations are found on an abundance of eukaryotic proteins. N-linked glycosylation is an attachment of oligosaccharide chains onto asparagine residues in a tripeptide sequence of Asn-X-Ser or Asn-X-Thr, where X is any amino acid except proline. O-linked glycosylation is found on the hydroxyl groups of serine and threonine resiudes. Glycosylation has a role in many different functions including protein folding and glycoprotein secretion.

Currently, there is evidence that UCH-L1 may be post-translationally glycoslyated. Cole and Hart [43] have reported N-acetylglucosamine (O-GlcNac) modifications in cytosolic fractions of proteins in nerve terminals. They use PNGase F-resistant galactosyltransferase labeling to identify O-GlcNAc modifications in synaptosome proteins. UCH-L1 was one of three proteins they identified as modified.

Mass Spectrometry

Mass spectrometry (MS) is a powerful and useful tool for analyzing all types of biomolecules over a vast range of masses. The extreme versatility of mass spectrometry makes it applicable to a broad range of samples including microorganisms, drugs, and proteins. Mass spectrometry provides the mass of a sample analyte by determining the mass-to-charge ratio (m/z) of its ionic

species. Samples subjected to MS must be vaporized and charged. A mass spectrometer consists of four basic components: sample inlet, ionization source, mass analyzer, and the detector. The sample inlet is the area in which the sample is introduced to the instrument. The ionization source generates analyte ions by inducing either the loss or gain of charge. Ions are then electrostatically directed to the mass analyzer where they are separated according to the m/z ratio. The ion detector then converts the ion energy into electrical signals. The resulting mass spectrum displays peaks corresponding to the m/z ratio of all charged species in the analyte. The spectrum plots the m/z on the x-axis against the relative intensity of each peak on the y-axis. (The relative intensity shows the abundance of an individual m/z in relation to the most intense peak in that particular spectrum).

There are many ionization techniques for mass spectrometry. Two very common sources include matrix–assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), both of which have helped to broaden the limits of mass spectrometric applications, especially to include proteins and peptides.

MALDI incorporates the use of a multi-functional chromophoric organic matrix in excess. As a chromophore, it absorbs the ultraviolet radiation from the laser beam; and as an acid, it provides protons for ionization of the analyte. The matrix is applied to the sample and allowed to dry, thus co-crystallizing with the analyte. A laser beam eradicates the crystals, causing the matrix phase to readily shift from solid to compressed gas, carrying with it

the analyte molecules. The analyte molecules are ionized in this process and are directed toward the mass analyzer to be separated.

Electrospray ionization also occurs through a series of processes that produce gaseous, ionized molecules directly from the liquid sample. First, a voltage of approximately 2 kV is applied to the ES capillary tip, producing charged droplets. Second, as heat is applied to the droplets, the solvent evaporates, causing an increase in the charge density. As coulombic repulsions grow stronger, surface tension forces grow weaker, and the sample ions are released and fly through to the vacuum of the mass analyzer.

Peptide mass mapping is a useful MS technique for protein identification. For this, the protein is subjected to proteolytic digestion to produce peptides that can be analyzed more readily than the more hydrophobic intact protein. Choice of digestion enzyme is fundamental to producing useful peptides. Trypsin is one of the most common digestion enzymes, and is used in the present study. It cleaves peptide bonds on the C-terminal side of lysine and arginine residues, at an optimal pH between 7.5 and 9.0. Depending on the desired properties of the resulting peptides, there are many other proteases which can be used. The other enzyme used for this work is endoproteinase Glu-C, which cleaves peptide bonds on the C-terminal side of glutamic acid at pH of 7.8.

The resulting peptide mixture can be analyzed by mass spectrometry. Analyzing proteins at the peptide level provides more detailed look at the protein under study, and allows localization of modifications.

Objective

The objective of this thesis is to develop methods to characterize the UCH-L1 protein structure and develop an approach to identify post-translational modifications by peptide mass fingerprint through mass spectrometry.

Chapter 2: EXPERIMENTAL

Gel Electrophoresis

Human recombinant UCH-L1 [1µg/µl], expressed in E. coli, was purchased from Boston Biochem (Cambridge, MA). DL-Dithiothreitol (DTT; Sigma, St. Louis, MO) was always prepared fresh immediately before its use. β-mercaptoethanol (Sigma), a reducing agent, was added to Laemmli sample buffer (Bio Rad, Hercules, CA) immediately before use. Tris-HCl polyacrylamide Criterion Gels with 10-20% linear gradient (Bio-Rad) were used for electrophoresis. Stock solution of 5X running buffer for electrophoresis consisted of 25mM Tris (Sigma), 192 mM glycine (Sigma), and 0.1% sodium dodecyl sulfate (SDS) (Sigma). Low-range pre-stained SDS-PAGE standards were purchased from Bio Rad. Fixing solution consisted of 35% methanol (J.T. Baker, Phillipsburg, NJ), 15% glacial acetic acid (J.T. Baker), and 50% deionized water (Millipore, Billerica, MA). Bio-Safe^{TM} Coomasie was obtained from Bio Rad. Densitometry was performed on the GS-800 calibrated densitometer (Bio Rad).

Two microliters of human recombinant UCH-L1 was added to a microcentrifuge tube. Two microliters 10 mM DTT was added. The sample was incubated for 15-20 minutes at room temperature. To prepare the sample buffer, 12.5 µl of β-mercaptoethanol were added to 237.5 µl Laemmli sample buffer and mixed well by vortex. Eight microliters of sample buffer were then added to the protein sample and mixed by pipette. The sample

was placed in a boiling water bath for five minutes. The Criterion gel was placed in the cell and filled with approximately 400 µl 1X running buffer. A small syringe was filled with running buffer and used to flush any residual SDS from each well. After cooling to room temperature, 10 µl of sample and the molecular weight markers were each loaded onto separate lanes on the gel. Running buffer was added to the sample chamber. The gel was run at 200 volts for approximately one hour. After electrophoresis, the gel was incubated with fixing solution for 45 minutes, followed by deionized water for 15 minutes, and finally stained in Coomasie blue overnight. The gel was destained in deionized water for 5 hours. Calibrated densitometry was used to capture an image of the gel.

In-Solution Digestion

Immobilized L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK) trypsin and Handee™ microcentrifuge tubes were purchased from Pierce (Rockford, IL). Endoproteinase Glu-C was purchased from Sigma. The digestion buffers for the trypsin and Glu-C digestions consisted of 0.1 M NH_4HCO_3, pH 8.0 and 7.8, respectively. The pH was adjusted with 6M sodium hydroxide (NaOH) (Fisher) as needed.

Five microliters of immobilized TPCK trypsin was added to a microcentrifuge tube and washed by adding 10 µl of digestion buffer. The tube was centrifuged for 15 sec at 4.0 rpm. The supernatant was removed by pipette and the wash was repeated two times. The trypsin beads were

resuspended in 5 µl of digestion buffer. Five microliters of human recombinant UCH-L1 was added to a microcentrifuge tube. Five microliters of 0.1 M DTT was added to the protein and incubated for 15 minutes. Three microliters of digestion buffer was added to the reduced protein sample. The protein sample was transferred to the tube containing the trypsin and mixed by pipetting. The sample was incubated overnight at 37ºC. Following overnight enzymatic digestion, the supernatant was recovered from the trypsin and transferred to a new centrifuge tube. The tryptic peptides were concentrated in the speed vac and dissolved in 0.1% TFA. The peptide solution was subjected to desalting with Zip Tip C_{18} as already described. The peptides were dried completely in the speed vac and reconstituted in electrospray buffer (methanol, water, glacial acetic acid, 49:49:2, v/v/v).

The endoproteinase Glu-C digestion was performed according to the manufacturer's instructions. The lyophilized Glu-C product was reconstituted in 250 µl of deionized water, resulting in a final concentration of 0.1 µg/µl. A 5 µl aliquot of human recombinant UCH-L1 was dried down in the speed vac and reconstituted in 5 µl of digestion buffer. To achieve an enzyme-to-substrate ratio of 1:20, 2.5 µl of Glu-C solution was added to the protein sample. The sample was incubated for 14 hours at 37ºC. The Glu-C peptides were dried down in the speed vac and reconstituted in 5 µl 0.1% TFA. The peptides were desalted with a Zip Tip C_{18} column as described above, except the peptides were eluted in 5 µl electrospray buffer.

Mass Spectrometry and Bioinformatics

The intact protein and proteolytic peptides were analyzed by mass spectrometry. The intact protein was first analyzed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-ToF MS) on an AXIMA-CFR instrument (Kratos Analytical, Chestnut Ridge, NY) equipped with pulsed extraction and a nitrogen laser operating at 337 nm. The matrix was saturated 3,5-dimethoxy-4-hydroxycinnamic acid in 70% ACN / 0.1% TFA. A sterile petri dish was filled with fifty microliters of deionized water. A dialysis membrane (Millipore) was placed on top of the water. Four microliters of human, recombinant UCH-L1 was placed directly on the surface of the membrane and dialyzed for 30 minutes for desalting. One microliter of protein sample was placed between one-microliter layers of matrix on the MALDI plate. For MALDI-MS, a three-point calibration was first performed. The calibrants and their corresponding weights are as follows: cytochrome c, 12,361.96 Da; apomyoglobin, 16,952.27 Da, aldolase, 39,212.28 Da. The mass range scanned was from 1,000-70,000 Da, in linear mode.

The intact protein and peptides from the in-solution digests were analyzed by electrospray ionization mass spectrometry (ESI-MS) on the API QSTAR Pulsar Qq-TOF (Applied Biosystems, Foster City, CA) utilizing a nanospray source (Protana A/s, Odense, Denmark). All scans were

performed with the following parameters: positive ion mode; ion spray voltage = 0.9 kV; curtain gas = 25 L/min; resolution = 2400 CEM.

For peptide analysis with ESI-MS, a two-point calibration was performed with Angiotensin II at m/z 1046.5 and 523.0. Two microliters of the peptide mixture were introduced into the ionization source via nanoelctrospray disposable needles (Proxeon Biosystems, Denmark). The TOF-MS scan was measured first to identify the m/z values of the peptide molecular ions. This precursor scan range was set to examine m/z 350-1500 and the instrument generated spectra accordingly. Collision induced dissociation was used for MS/MS analysis. Here, multiply charged molecular ions of interest from the MS spectra were isolated in Q1, the mass filter quadrupole cell, and solely passed into Q2, the collision cell. In these product ion scans, fragmentation was induced by nitrogen gas transmitted into Q2. During acquisition, the collision energy was adjusted as necessary to produce fragments, typically b- and y-ions, depending on the charge retention site following fragmentation.

For analysis of the intact protein with, 2 microliters of sample were dialyzed against water and subjected to ESI-MS. The scan range was m/z 600-4000.

Theoretical peptides were generated using GPMAW software (Lighthouse Data, Denmark). This computer program has the ability to provide masses of the singly and multiply charged ions from various proteolytic digests. Furthermore, it provides the masses of the fragment a-, b-

, x-, and y-ions from MS/MS analysis. The SwissProt accession number for human UCH-L1, P09936, was used for GPMAW.

Chapter 3: RESULTS

The molecular weight of intact human, recombinant UCH-L1 was obtained on MALDI-MS and ESI-MS, as well as gel electrophoresis. Figure 1 shows the densitometry image of the SDS-PAGE experiment. The intact protein migrates to a position at approximately 27 kDa, as seen by its position relative to the adjacent molecular weight standards.

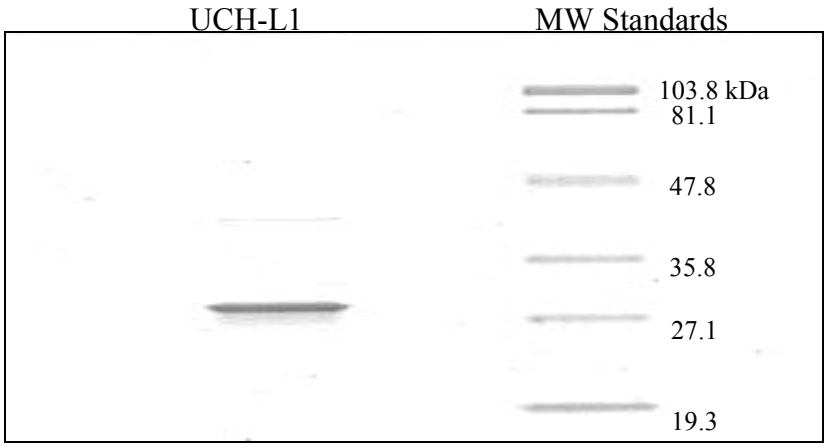

Figure 1: SDS-PAGE image of intact UCH-L1

Figure 2 shows the MALDI-MS spectrum of the intact protein with high abundance peaks at m/z = 8,284; 12,420; and 24,830 Da, representing the triply, doubly and singly charged ions, respectively. The lower intensity peak m/z 49,617 Da is due to an artifactual dimerized cluster of the protein.

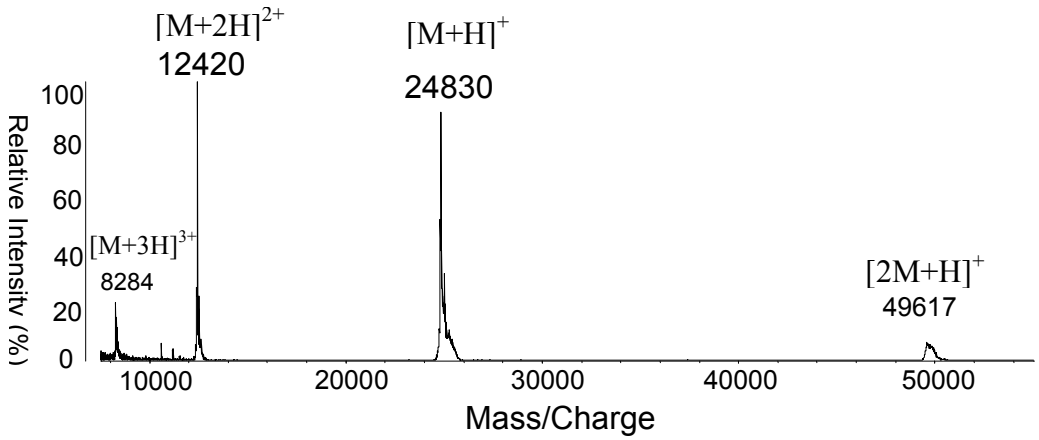

Figure 2: Matrix-assisted laser desorption ionization mass spectrum of intact UCH-L1

The ESI-MS spectrum shows a characteristic charge distribution of the molecular ions of the intact protein, as seen in Figure 3. The charged species within this distribution range from $M+34H^+$ to $M+17H^+$, at m/z = 731.61 Da and 1,462.35 Da, respectively. Using the Bayesian Peptide Reconstruct feature on the Analyst 1.1 software, the molecular ions from the TOF-MS scan were deconvoluted to show the neutral species. In Figure 4, the reconstructed chromatogram shows a single peak at m/z 24,840.00. No dimer was observed.

The ESI-TOF-MS deconvoluted spectra from the proteolytic digests are shown in Figures 5 and 6. The monoisotopic masses from each peak are listed, along with the theoretical peptide masses and the differences between these values, in Tables 1 and 2.

For further analysis and verification, a select group of ions from the TOF-MS scan were fragmented by MS/MS for sequence verification. The product ion scans from selected TOF-MS ions are shown in Figure 7 and 8. The b- and y-ions of the fragment sequences are displayed in the spectrum window.

The sequence coverage is a ratio of the number of amino acids in the identified peptides to the total number of amino acids in the sequence, defined as follows:

$$\% \text{ Sequence coverage} = \frac{\text{\# of amino acids in the identified proteins}}{\text{total \# of amino acids in sequence}} \times 100\%$$

When treated with immobilized TPCK trypsin alone, the sequence coverage was 58% of the protein. The Glu-C digest achieved 48% sequence coverage.

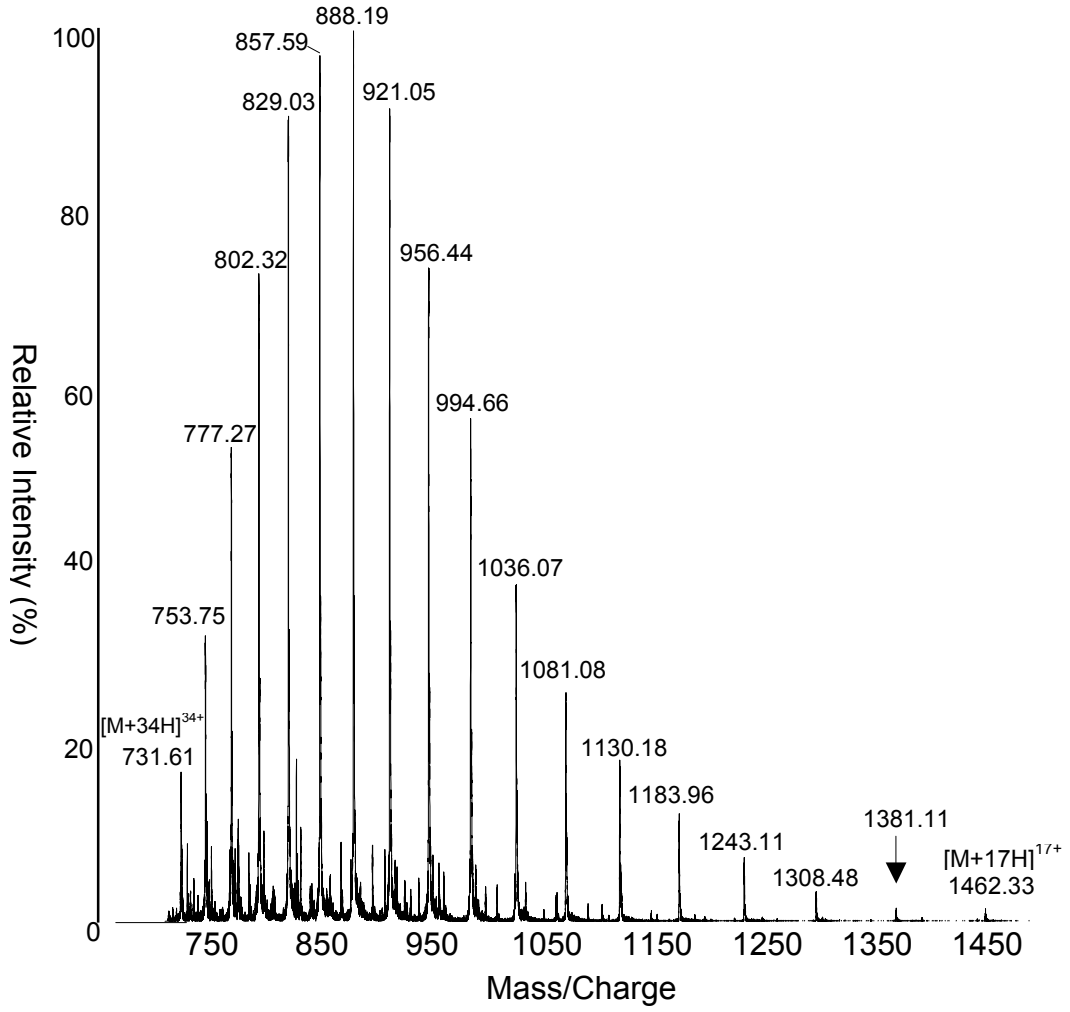

Figure 3: Electrospray ionization mass spectrum of intact UCH-L1

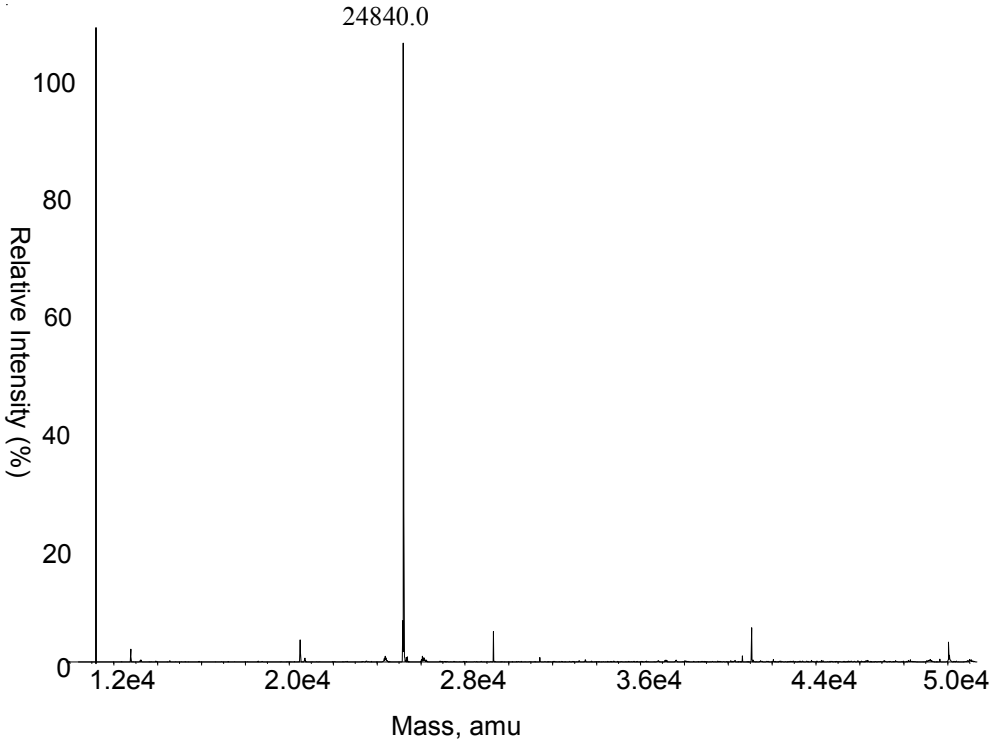

Figure 4: Deconvoluted mass spectrum from Fig. 3

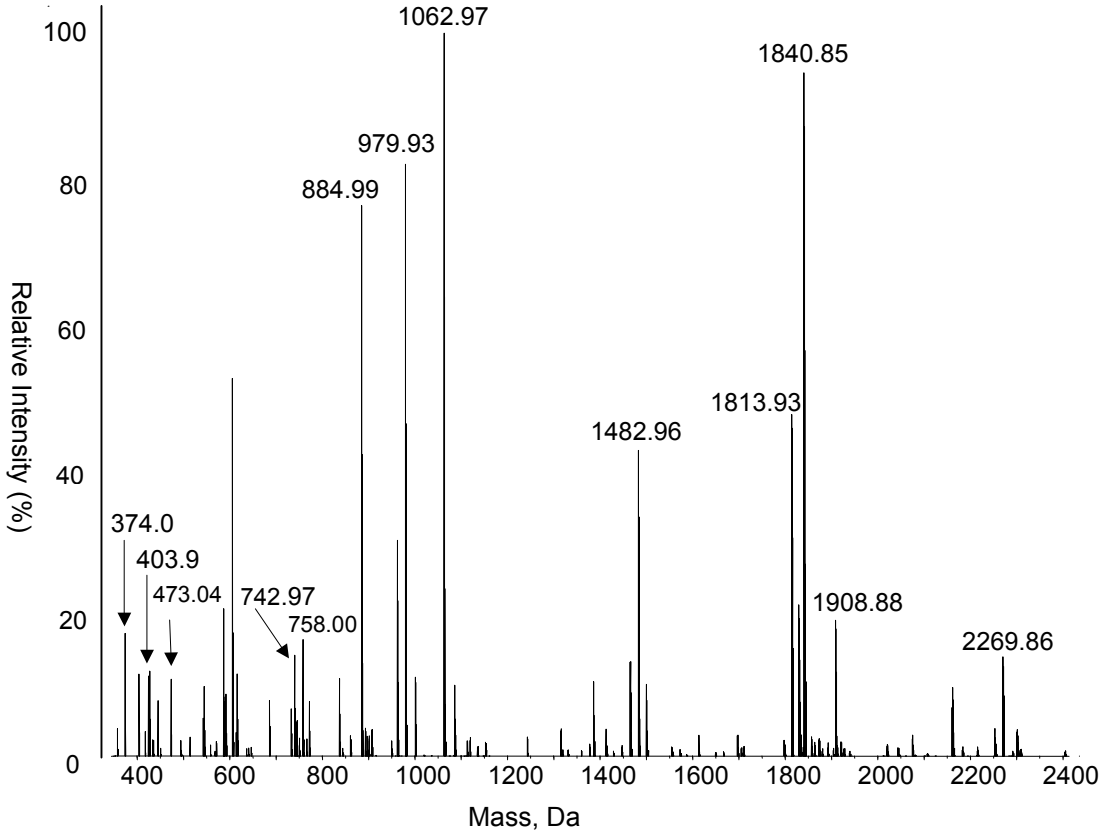

Figure 5: Deconvoluted electrospray ionization mass spectrum of trypsin digest

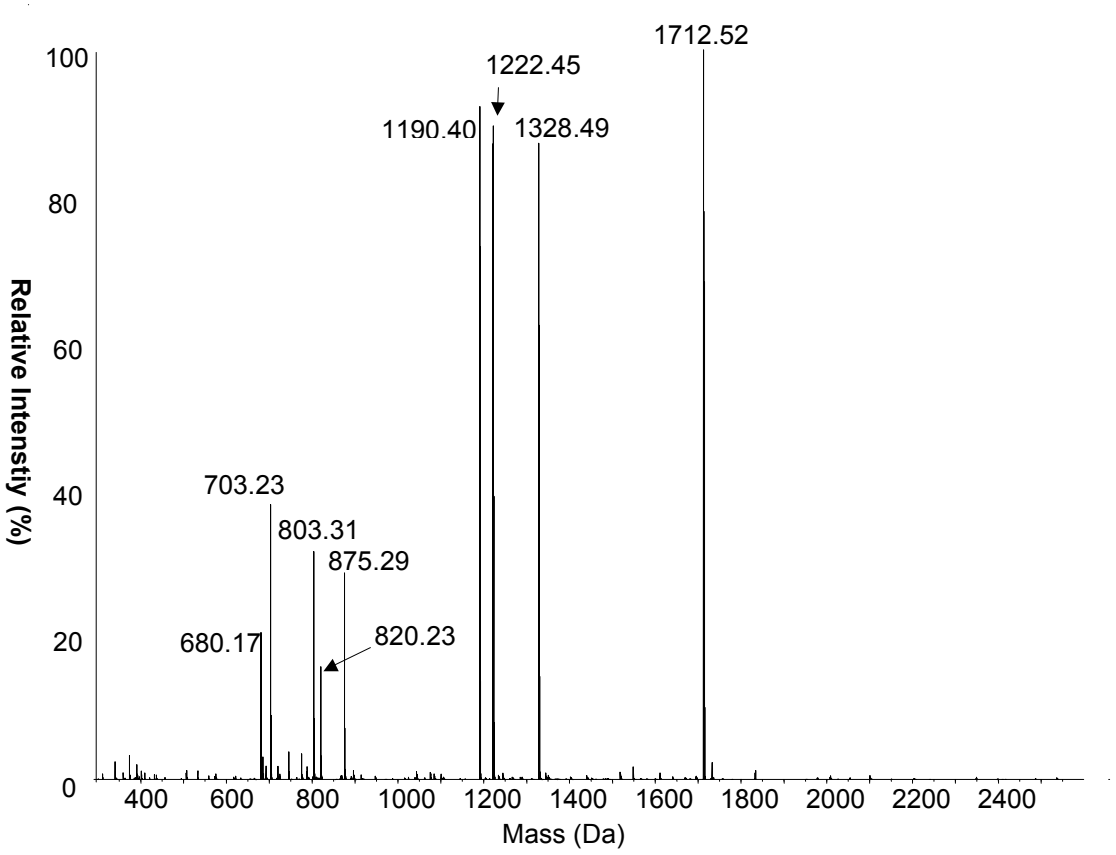

Figure 6: Deconvoluted electrospray ionization mass spectrum of endoproteinase Glu-C digest

Table 1: Theoretical trypsin digest peptides

Theoretical Mass (Da)	Observed Mass (Da)	Δ M	Amino Acid #	Peptide Sequence
4761.4867*			20-63	LGVAGQWRFVDVLGLEEE SLGSVPAPACALLLLFPLTA QHENFR
4342.1270*			158-195	VNFHFILFNNVDGHLYELD GRMPFPVNHGASSEDTLLK
4022.1101*			28-64	FVDVLGLEEESLGSVPAPA CALLLLFPLTAQHENFRK
3894.0152			28-63	FVDVLGLEEESLGSVPAPA CALLLLFPLTAQHENFR
3297.6538*			84-115	QTIGNSCGTIGLIHAVANNQ DKLGFEDGSVLK
2975.4566*			154-178	VDDKVNFHFILFNNVDGHL YELDGR
2920.4450*			79-105	VYFMKQTIGNSCGTIGLIHA VANNQDK
2518.2393			158-178	VNFHFILFNNVDGHLYELD GR
2415.0696*			132-153	CFEKNEAIQAAHDAVAQEG QCR
2366.0795*			136-157	NEAIQAAHDAVAQEGQCR VDDK
2270.1949*	2269.86	0.33	1-19	MQLKPMEINPEMLNKVLSR
2252.1093			84-105	QTIGNSCGTIGLIHAVANNQ DK
2227.0943*			179-199	MPFPVNHGASSEDTLLKDA AK
2026.0259*			106-123	LGFEDGSVLKQFLSETEK
1908.8622	1908.88	0.03	136-153	NEAIQAAHDAVAQEGQCR
1841.8982	1840.85	1.05	179-195	MPFPVNHGASSEDTLLK
1814.9093	1813.93	0.73	1-15	MQLKPMEINPEMLNK
1695.7774*	1696.97	1.19	116-129	QFLSETEKMSPEDR
1534.7688*			208-221	EQGEVRFSAVALCK
1483.7882*	1482.96	0.83	66-78	QIEELKGQEVSPK
1411.7170*	1412.91	1.19	72-83	GQEVSPKVYFMK
1378.6477*	1377.91	0.74	203-213	EFTEREQGEVR
1340.7677*			16-27	VLSRLGVAGQWR
1063.5550	1062.97	0.59	106-115	LGFEDGSVLK
1037.4838*			200-207	VCREFTER

(*) indicates missed cleavage

Table 1: Theoretical trypsin digest peptides (continued)

Theoretical Mass (Da)	Observed Mass (Da)	Δ M	Amino Acid #	Peptide Sequence
980.4815	979.93	0.55	116-123	QFLSETEK
978.5083*			214-223	FSAVALCKAA
932.4386*			124-131	MSPEDFRAK
886.5124*			65-71	KQIEELK
885.4821	884.99	0.49	20-27	LGVAGQWR
836.4340	836.98	0.55	214-221	FSAVALCK
760.3776*			196-202	DAAKVCR
758.4174	758.00	0.42	66-71	QIEELK
743.3814	742.97	0.41	72-78	GQEVSPK
733.3065	732.91	0.40	124-129	MSPEDR
723.3500*			130-135	AKCFEK
716.3453			208-213	EQGEVR
686.3462	685.96	0.39	79-83	VYFMK
680.3130			203-207	EFTER
524.2179			132-135	CFEK
475.2278			154-157	VDDK
473.2962	473.04	0.26	16-19	VLSR
403.2067	403.98	0.77	196-199	DAAK
375.1814	374.02	0.16	200-202	VCR
274.2005*			64-65	KK
217.1426			130-131	AK
160.0848			222-223	AA
146.1055			64-64	K
146.1055			64-64	K

(*) indicates missed cleavage

Table 2: Theoretical Glu-C digest peptides

Theoretical Mass (Da)	Observed Mass (Da)	Δ Mass	Amino Acid #	Peptide Sequence
5285.8223*			12-60	MLNKVLSRLGVAGQWRFV DVLGLEEESLGSVPAPACA LLLLFPLTAQHE
4981.5232*			75-120	VSPKVYFMKQTIGNSCGTI GLIHAVANNQDKLGFEDGS VLKQFLSE
4673.1890*			150-190	GQCRVDDKVNFHFILFNNV DGHLYELDGRMPFPVNHG ASSE
4333.2113*			70-109	LKGQEVSPKVYFMKQTIGN SCGTIGLIHAVANNQDKLGF E
4181.9892*			138-174	AIQAAHDAVAQEGQCRVD DKVNFHFILFNNVDGHLYE
3777.9097			75-109	VSPKVYFMKQTIGNSCGTI GLIHAVANNQDKLGFE
3518.8834*			38-69	SLGSVPAPACALLLLFPLTA QHENFRKKQIEE
3410.8020*			8-37	INPEMLNKVLSRLGVAGQW RFVDVLGLEEE
3154.5414*			175-203	LDGRMPFPVNHGASSEDT LLKDAAKVCRE
2977.4055			150-174	GQCRVDDKVNFHFILFNNV DGHLYE
2957.5797			12-37	MLNKVLSRLGVAGQWRFV DVLGLEEE
2346.2532			38-60	SLGSVPAPACALLLLFPLTA QHE
1836.9166*			191-206	DTLLKDAAKVCREFTE
1745.9424*			61-74	NFRKKQIEELKGQE
1712.7941	1712.66	0.13	175-190	LDGRMPFPVNHGASSE
1593.7747*			135-149	KNEAIQAAHDAVAQE
1547.8004*	1548.51	0.71	209-223	QGEVRFSAVALCKAA
1459.7579			191-203	DTLLKDAAKVCRE
1451.7144*	1451.53	0.18	110-122	DGSVLKQFLSETE
1438.6459*	1440.46	1.81	123-134	KMSPEDRAKCFE
1328.6468*	1328.49	0.15	1-11	MQLKPMEINPE
1237.5635*			128-137	DRAKCFEKNE

(*) indicates missed cleavages

Table 2: Theoretical Glu-C digest peptide mass list (continued)

Theoretical Mass (Da)	Observed Mass (Da)	Δ Mass	Amino Acid #	Peptide Sequence
1234.6778	1234.44	0.24	212-223	VRFSAVALCKAA
1222.5942			138-149	AIQAAHDAVAQE
1221.6241	1221.46	0.16	110-120	DGSVLKQFLSE
1190.6408	1190.41	0.23	61-69	NFRKKQIEE
875.4245	875.294	0.13	1-7	MQLKPME
866.3831			128-134	DRAKCFE
820.3637*	820.24	0.12	121-127	TEKMSPE
680.3130*	681.1904	0.88	204-208	FTERE
617.2769*			207-211	REQGE
590.2734			123-127	KMSPE
573.3122			70-74	LKGQE
471.2329			8-11	INPE
395.1692			204-206	FTE
389.1910			135-137	KNE
332.1332			209-211	QGE
303.1543			207-208	RE
248.1008			121-122	TE

(*) indicates missed cleavage

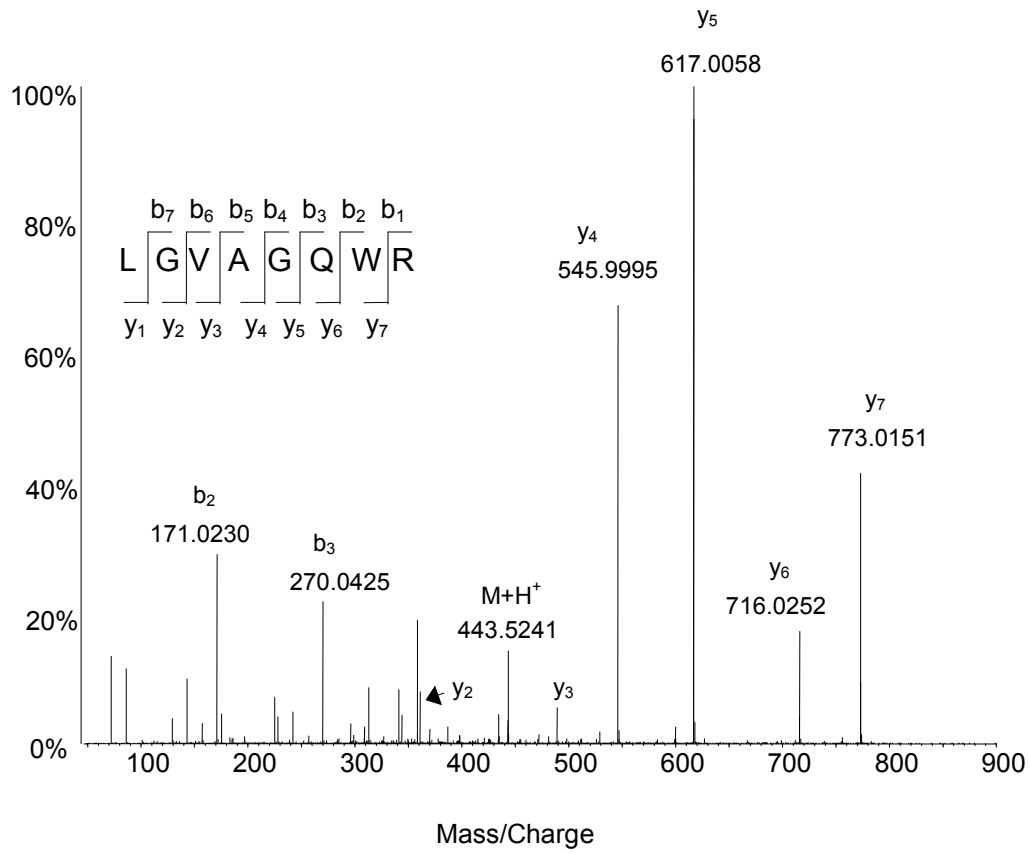

Figure 7: Tandem mass spectrum of molecular ion 443.5.

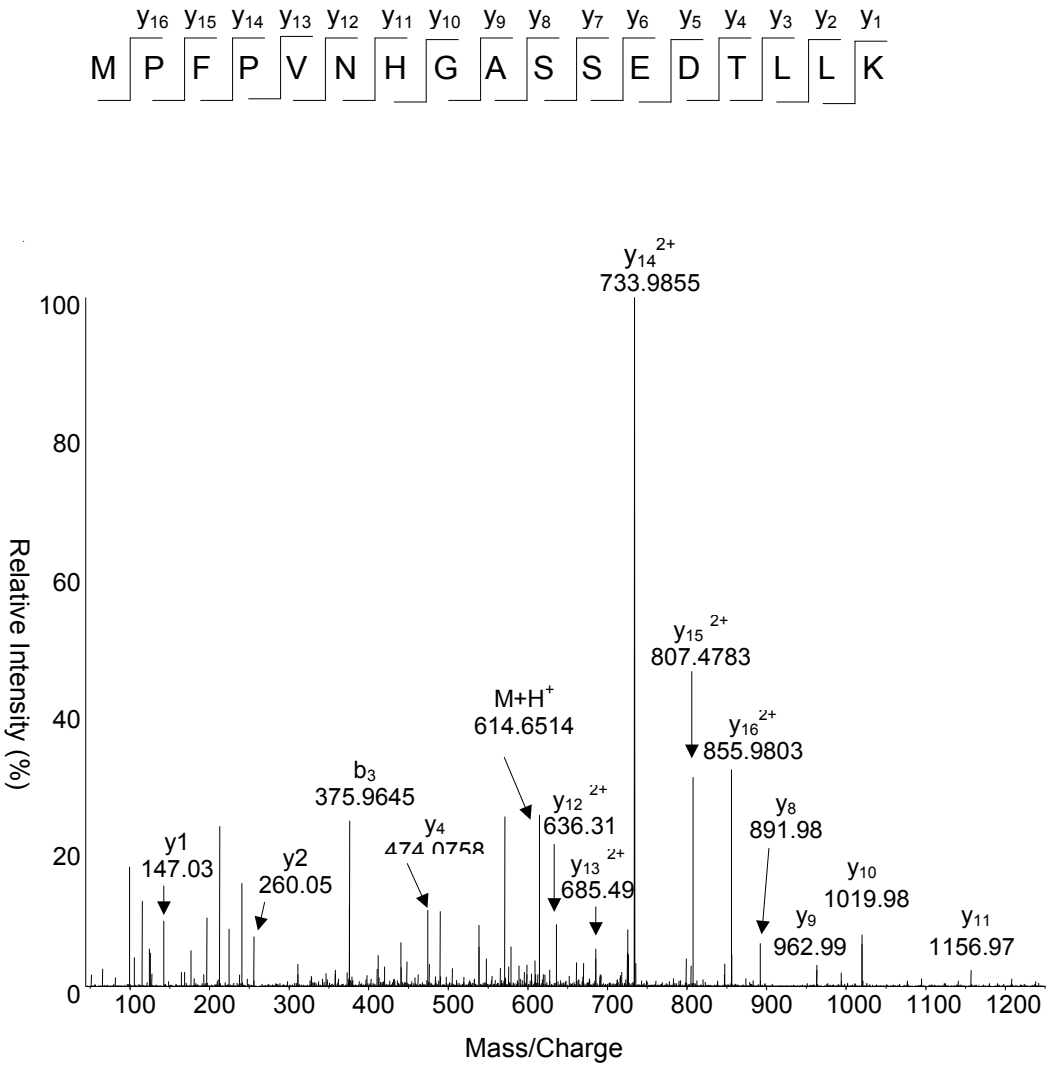

Figure 8: Tandem mass spectrum of m/z 614.6.

The "additive sequence coverage" refers to combined sequence coverage from the two enzymes with different enzymatic specificity. The additive sequence coverage of the recombinant UCH-L1 is 69%. This slight increase is due to a small number of amino acids recovered in the Glu-C digest that are not identified in the tryptic digest. The sequence coverages of both digests are displayed in Figures 9 and 10.

MQLKPMEINP EMLNKVLSRL GVAGQWRFVD VLGLEEESLG SVPAPACALL

LLFPLTAQHE NFRKKQIEEL KGQEVSPKVY FMKQTIGNSC GTIGLIHAVA

NNQDKLGFED GSVLKQFLSE TEKMSPEDRA KCFEKNEAIQ AAHDAVAQEG

QCRVDDKVNF HFILFNNVDG HLYELDGRMP FPVNHGASSE DTLLKDAAKV

CREFTEREQG EVRFSAVALC KAA

Figure 9: Sequence coverage of tryptic peptides. Red indicates matched peptides.

MQLKPMEINP EMLNKVLSRL GVAGQWRFVD VLGLEEESLG SVPAPACALL

LLFPLTAQHE NFRKKQIEEL KGQEVSPKVY FMKQTIGNSC GTIGLIHAVA

NNQDKLGFED GSVLKQFLSE TEKMSPEDRA KCFEKNEAIQ AAHDAVAQEG

QCRVDDKVNF HFILFNNVDG HLYELDGRMP FPVNHGASSE DTLLKDAAKV

CREFTEREQG EVRFSAVALC KAA

Figure 10: Sequence coverage of Glu-C peptides; red indicates matched peptides.

Chapter 4: DISCUSSION

If the gene sequence for the protein under study is known, experimental data can be compared to the theoretical sequence. The sample used in this study was UCH-L1 protein isolated from human cells and grown in an E. coli host. Therefore, the recombinant sample is not expected to be modified much, because E. coli lacks the enzyme systems to carry out that processing. However, using recombinant sample is an effective approach for primary analysis, as it provides a model for analysis, and a baseline for comparisons for future studies of cells containing abnormal UCH-L1 expression.

The molecular masses from the electrophoresis results vary from the mass spectrometry values. The SDS-PAGE image shows the monomeric protein's migration at approximately 27 kDa, approximately 3,000 Da greater than both mass spectrometry values. Mass spectrometry has greater mass accuracy than electrophoresis. Other proteins are reported to behave in a similar manner in gel electrophoresis measurements [45]. To address the discrepant results between the two methods, it was important to analyze the proteolytic peptides of UCH-L1.

Using mass spectrometry and peptide mass mapping techniques a total of 69% of the UCH-L1 human recombinant protein was recovered. In the tryptic digest analysis, there are three peptides that were not recovered in the MS spectra (Table 3). A common challenge in proteomics is retrieving peptides with higher masses (>2.5 kDa) after the digest. Higher mass

peptides consist of longer amino acid sequences and tend to be more hydrophobic. Increased hydrophobicity affects the solubility of these peptides making it difficult to resuspend them in solution after digestion and desalting. The first peptide listed in Table 3 contains two hydrophobic clusters, [FVDVL] and [ACALLLLF], which may have contributed to its insolubility and subsequent absence from the mass spectrum.

Peptide	MW (Da)	% Hydrophobicity	%Charged
FVDVLGLEEESLGSV PAPACALLLLFPLTAQ HENFR	3894.02	53	17
VNFHFILFNNVDGHLY ELDGR	2518.24	48	29
QTIGNSCGTIGLIHAV ANNQDK	2252.11	41	14

Table 3: Unmatched peptides from trypsin digest

The Glu-C digest produced fewer peptides that corresponded to the values listed in the theoretical digest. Many of these peptides included missed cleavage sites. For many of the peptides with missed cleavages, there is a corresponding mass value of one of the two included peptides. This pattern is indicative of incomplete digestion and may be alleviated by longer incubation of the protein with the digestion enzyme or increased enzyme to protein concentration ratio.

It is important to note that the enzymatic activity of endoproteinase Glu-C is pH-dependent. In addition to cleaving the C-terminal peptide bonds of glutamic acid residues at pH 7.8, this endoproteinase will also cleave C-terminal peptide bonds of aspartic acid and glutamic acid residues at pH 4. There are two relatively intense peaks on the Glu-C digest spectrum, at m/z 703.23 Da and 803.31 Da, which result from an aspartic acid residue cleavage. Since these peptides are legitimate products of the Glu-C digest, they have been included in the sequence coverage calculations.

In both enzymatic digestions, the observed mass values are within 1.8 Da of the theoretical mass values, with an average Δ mass of 0.5 ± 0.4 Da. It can be concluded that the close agreements of these values indicate that there are no post-translational modifications present on these particular peptides of the recombinant protein.

Another significant observation is the retention of the N-terminal methionine that serves as the translation initiation signal. E. coli will commonly remove the N-terminal methionine in many proteins and/or formylate the N-terminus [46]. The N-terminal proteolytic peptides are seen in both digests with the methionine residue intact, and without variation in masses.

High sequence coverage is not necessarily required for identification of unknown proteins. Many proteins can be identified by matching as few as three or four peptide masses [47]. However, when attempting to characterize any post-translational modifications, sequence coverage becomes significant.

Individual amino acids in the sequence may provide essential information to the overall structure or be a site for modification.

One approach to increasing the sequence coverage of UCH-L1 would be through sequential proteolysis. In the work presented here, two aliquots of recombinant UCH-L1 sample were digested separately, and their sequence coverages were compared. With sequential proteolysis, one aliquot of UCH-L1 can be digested with a primary protease; the resulting peptides would then be subjected to a second round of digestion with a different protease. This technique would be beneficial for the high-mass peptides that are not being recovered in the initial digestion, such as the unmatched peptides mentioned above. Maintaining the appropriate pH levels is imperative for this method, as most proteolytic enzymes only perform their enzymatic role in specific pH ranges.

Liquid chromatography-MS may also be used to identify peptides from a proteolytic digest. Liquid chromatography would provide peptidic separation prior to MS. Adjustable gradients and flow rates allow analysis of peptides that pose challenges in one-dimensional mass spectrometry. Greater resolution of components in peptide mixtures is one advantage obtained when LC is coupled with MS. Initial separation of a peptide mixture by LC-MS is sometimes warranted to reduce ionization suppression effects. Other advantages of LC-MS include enhanced peak intensity, greater number of identifiable peaks in the mass spectra, and greater sequence coverage.

The recombinant UCH-L1 protein contains many potential sites for modifications as highlighted in Figure 11. Phosphorylations are known to occur on the hydroxyl groups of serine, tyrosine, and threonine residues. Glycosyl groups are added to the side chains of asparagine, serine, and threonine residues. There are reports of a reciprocal relationship between phosphorylation and glycosylation- more specifically, O-GlcNAc modification- in proteins [48, 49]. With the evidence of the presence of O-GlcNAc on UCH-L1, it is of great interest to pursue the search for phosphorylations.

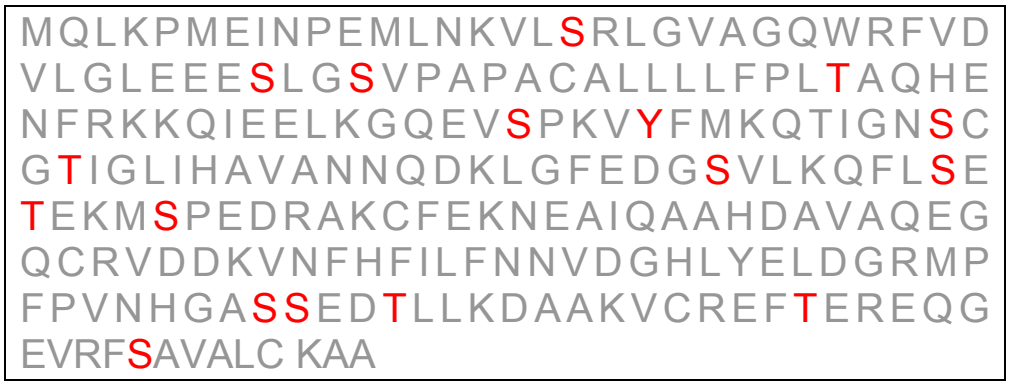

Figure 11: Some potential post-translational modification sites

Protein phosphorylations can be quantitatively detected by the proteomic approaches used in this work. One approach to mapping phosphorylations is to isolate UCH-L1 from the cell and treat part of the sample with protein phosphatases. These enzymes work to remove phosphate groups from amino acid residues. Depending on the amount of phosphorylation, if any, the protein may be initially analyzed by two-

dimensional gel electrophoresis to determine a difference in migration patterns between the phosphorylated and dephosphorylated proteins' isoelectric points. Localization of the modification is more likely to occur at the peptide level. Any phosphorylations detected in pairs of peptides can be mapped directly to the site of modification by MS and tandem MS (MS/MS). Tandem MS provides more detailed sequence information by fragmenting molecular ions into the constituent amino acids. In this way, modification sites can be located by monitoring mass shifts between the unmodified and modified proteins. In the case of phosphorylation, Δ m/z = 80 Da, for each phosphate group present.

Carbohydrate groups tend to be more massive, and mass differences between glycosylated protein and its unmodified form can be more than 1,000 daltons, which may be determined on high-resolution gels. Like phosphorylations, glycosylations can also be enzymatically removed depending on the nature of the modification. For example, peptide N-glycosidase F will deglycosylate N-linked glycoproteins. Chemical methods are used to remove O-linked glycans such as treatment with alkaline solutions (e.g., NaOH, Na_2CO_3) to promote hydrolysis via a β–elimination mechanism. Once the carbohydrate groups are removed, any mass differences in the peptide pairs can also be identified by MS/MS.

Chapter 5: CONCLUSION

Mass spectrometry is an efficient method for analyzing biomolecules in proteomics. Protein characterization involves many steps, and should provide a thorough understanding of the protein's structure which will subsequently lead to a better understanding of its function. Characterization can be done through "bottom-up" proteomics where the protein is enzymatically cleaved into its consitutent peptides. Bottom-up proteomics allows for the protein to be analyzed and reconstructed one peptide at a time. Because peptides are smaller and less hydrophobic than proteins, they are more readily analyzed by analytical techniques such as HPLC and mass spectrometry.

In this study, the recombinant form of human ubiquitin carboxy-terminal hydrolase L1 protein has been mapped. After being treated with two proteolytic enzymes, the resulting peptides are recovered and analyzed by mass spectrometry. As the name suggests, peptide mass mapping allows identification of proteins by matching theoretical peptide mass values with observed values from mass spectrometry data. Peptides from two proteolytic digests of UCH-L1 were used to gather structural information of UCH-L1 and provided 69% sequence coverage of the recombinant protein.

The portion of the protein sample that was not detected by these methods may be obtained by liquid chromatography coupled with mass spectrometry by providing greater peptidic separation and reducing ion suppression effects. Sequence coverage may also be increased by

generating smaller peptides that will more likely be detected by mass spectrometry. This can be done by sequential proteolysis using different enzymes to produce smaller sequences.

Searching for post-translational modifications is the final step to protein structure characterization. Abnormal proteins are sometimes due to mutations in the gene sequence or PTMs. Elucidation of modifications can lead to understanding the protein functions, or lack thereof. Previous studies of UCH-L1 have found it to be modified with O-GlcNAc, a common O-linked glycosylation. Other glycosylations or phosphorylations may also be a part of it's structure.

UCH-L1's abundance in brain protein and it's role in protein degradation makes it of interest to mental health researchers studying neurodegenerative disorders. Furthermore, it has been implicated in metastasis because of expression detected outside of brain proteins. By mapping the recombinant form of UCH-L1, primary structural information has been obtained and can be used as a model for comparison for cells in which UCH-L1 has altered expression.

References

1. Hochstrasser, M., *Ubiquitin-dependent protein degradation.* Annual Review of Genetics, 1996. **30**: p. 405-439.
2. Wilkinson, K.D., *Regulation of ubiquitin-dependent processes by deubiquitinating enzymes.* Faseb Journal, 1997. **11**(14): p. 1245-1256.
3. D'Andrea, A. and D. Pellman, *Deubiquitinating enzymes: A new class of biological regulators.* Critical Reviews in Biochemistry and Molecular Biology, 1998. **33**(5): p. 337-352.
4. Wilkinson, K.D., et al., *The Neuron-Specific Protein Pgp-9.5 Is a Ubiquitin Carboxyl-Terminal Hydrolase.* Science, 1989. **246**(4930): p. 670-673.
5. Osaka, H., et al., *Ubiquitin carboxy-terminal hydrolase L1 binds to and stabilizes monoubiquitin in neuron.* Human Molecular Genetics, 2003. **12**(16): p. 1945-1958.
6. Thompson, R.J., et al., *Pgp 9.5 - a New Marker for Vertebrate Neurons and Neuroendocrine Cells.* Brain Research, 1983. **278**(1-2): p. 224-228.
7. Doran, J.F., et al., *Isolation of Pgp 9.5, a New Human Neuron-Specific Protein Detected by High-Resolution Two-Dimensional Electrophoresis.* Journal of Neurochemistry, 1983. **40**(6): p. 1542-1547.
8. Kwon, J., et al., *Ubiquitin C-terminal hydrolase L-1 is essential for the early apoptotic wave of germinal cells and for sperm quality control during spermatogenesis.* Biology of Reproduction, 2005. **73**(1): p. 29-35.
9. Kwon, J., et al., *Two closely related ubiquitin C-terminal hydrolase isozymes function as reciprocal modulators of germ cell apoptosis in cryptorchid testis.* American Journal of Pathology, 2004. **165**(4): p. 1367-1374.
10. Kwon, J., et al., *Developmental regulation of ubiquitin C-terminal hydrolase isozyme expression during spermatogenesis in mice.* Biology of Reproduction, 2004. **71**(2): p. 515-521.
11. Kwon, J., et al., *Characterization of the testis in congenitally ubiquitin carboxy-terminal hydrolase-1 (Uch-L1) defective (gad) mice.* Experimental Animals, 2003. **52**(1): p. 1-9.
12. Liu, Y.C., et al., *Atypical ubiquityl ligase activity of UCH-L1 may explain its link to Parkinson's disease susceptibility.* Neurobiology of Aging, 2002. **23**(1): p. S175-S175.
13. Liu, Y.C., et al., *The UCH-L1 gene encodes two opposing enzymatic activities that affect alpha-synuclein degradation and Parkinson's disease susceptibility.* Cell, 2002. **111**(2): p. 209-218.
14. Leroy, E., Boyer, R., Polymeropoulos, M., *Intron-exon Structure of Ubiquitin C-Terminal Hydrolase-L1.* DNA Research, 1998. **5**(6): p. 397-400.
15. Edwards, Y.H., et al., *The Gene for Human Neuron Specific Ubiquitin C-Terminal Hydrolase Maps to Chromosome-4p14.* Cytogenetics and Cell Genetics, 1991. **58**(3-4): p. 1886-1887.
16. Larsen, C.N., B.A. Krantz, and K.D. Wilkinson, *Substrate specificity of deubiquitinating enzymes: Ubiquitin C-terminal hydrolases.* Biochemistry, 1998. **37**(10): p. 3358-3368.

17. Larsen, C.N., J.S. Price, and K.D. Wilkinson, *Substrate binding and catalysis by ubiquitin C-terminal hydrolases: Identification of two active site residues.* Biochemistry, 1996. **35**(21): p. 6735-6744.
18. Zhang, J., et al., *Association between a polymorphism of ubiquitin carboxy-terminal hydrolase L1 (UCH-L1) gene and sporadic Parkinson's disease.* Parkinsonism & Related Disorders, 2000. **6**(4): p. 195-197.
19. Wang, J., et al., *ACT and UCH-L1 polymorphisms in Parkinson's disease and age of onset.* Movement Disorders, 2002. **17**(4): p. 767-771.
20. von Bohlen und Halbach, O., A. Schober, and K. Krieglstein, *Genes, proteins, and neurotoxins involved in Parkinson's disease.* Progress in Neurobiology, 2004. **73**(3): p. 151-177.
21. Vigouroux, S., M. Briand, and Y. Briand, *Linkage between the proteasome pathway and neurodegenerative diseases and aging.* Molecular Neurobiology, 2004. **30**(2): p. 201-221.
22. Snyder, H. and B. Wolozin, *Pathological proteins in Parkinson's disease.* Journal of Molecular Neuroscience, 2004. **24**(3): p. 425-442.
23. Satoh, J. and Y. Kuroda, *A polymorphic variation of serine to tyrosine at codon 18 in the ubiquitin C-terminal hydrolase-L1 gene is associated with a reduced risk of sporadic Parkinson's disease in a Japanese population.* Journal of the Neurological Sciences, 2001. **189**(1-2): p. 113-117.
24. Nishikawa, K., et al., *Alterations of structure and hydrolase activity of Parkinsonism-associated human ubiquitin carboxyl-terminal hydrolase L1 variants.* Biochemical and Biophysical Research Communications, 2003. **304**(1): p. 176-183.
25. Naze, P., et al., *Mutation analysis and association studies of the ubiquitin carboxyterminal hydrolase L1 gene in Huntington's disease.* Neuroscience Letters, 2002. **328**(1): p. 1-4.
26. Moss, A., et al., *A role of the ubiquitin-proteasome system in neuropathic pain.* Journal of Neuroscience, 2002. **22**(4): p. 1363-1372.
27. Moore, D.J., V.L. Dawson, and T.M. Dawson, *Role for the ubiquitin-proteasome system in Parkinson's disease and other neurodegenerative brain amyloidoses.* Neuromolecular Medicine, 2003. **4**(1-2): p. 95-108.
28. Maraganore, D.M., et al., *UCHL1 is a Parkinson's disease susceptibility gene.* Annals of Neurology, 2004. **55**(4): p. 512-521.
29. Lowe, J., et al., *Ubiquitin Carboxyl-Terminal Hydrolase (Pgp 9.5) Is Selectively Present in Ubiquitinated Inclusion-Bodies Characteristic of Human Neurodegenerative Diseases.* Journal of Pathology, 1990. **161**(2): p. 153-160.
30. Layfield, R., J.R. Cavey, and J. Lowe, *Role of ubiquitin-mediated proteolysis in the pathogenesis of neurodegenerative disorders.* Ageing Research Reviews, 2003. **2**(4): p. 343-356.
31. Leroy, E., et al., *The ubiquitin pathway in Parkinson's disease.* Nature, 1998. **395**(6701): p. 451-452.
32. Maraganore, D.M., et al., *Case-control study of the ubiquitin carboxy-terminal hydrolase L1 gene in Parkinson's disease.* Neurology, 1999. **53**(8): p. 1858-1860.

33. Momose, Y., et al., *Association studies of multiple candidate genes for Parkinson's disease using single nucleotide polymorphisms.* Annals of Neurology, 2002. **51**(1): p. 133-136.
34. Elbaz, A., et al., *The relation between Parkinson's disease and the S18Y polymorphism in the UCH-L1 gene is modified by age at onset.* Neurology, 2002. **58**(7): p. A110-A111.
35. Elbaz, A., et al., *S18Y polymorphism in the UCH-L1 gene and Parkinson's disease: Evidence for an age-dependent relationship.* Movement Disorders, 2003. **18**(2): p. 130-137.
36. Brichory, F., et al., *Proteomics-based identification of protein gene product 9.5 as a tumor antigen that induces a humoral immune response in lung cancer.* Cancer Research, 2001. **61**(21): p. 7908-7912.
37. Sasaki, H., et al., *Expression of the protein gene product 9.5, PGP9.5, is correlated with T-status in non-small cell lung cancer.* Japanese Journal of Clinical Oncology, 2001. **31**(11): p. 532-535.
38. Rosas, S.L.B., et al., *Methylation status in the promoter region of the human PGP9.5 gene in cancer and normal tissues.* Cancer Letters, 2001. **170**(1): p. 73-79.
39. Hibi, K., et al., *PGP9.5 as a candidate tumor marker for non-small-cell lung cancer.* American Journal of Pathology, 1999. **155**(3): p. 711-715.
40. Hibi, K., et al., *Serial analysis of gene expression in non-small cell lung cancer.* Cancer Research, 1998. **58**(24): p. 5690-5694.
41. Caballero, O.L., et al., *Interaction and colocalization of PGP9.5 with JAB1 and p27(Kip1).* Oncogene, 2002. **21**(19): p. 3003-3010.
42. Hathout, Y., et al., *Proteomic phenotyping: metastatic and invasive breast cancer.* Cancer Letters, 2004. **210**(2): p. 245-253.
43. Cole, R.N. and G.W. Hart, *Cytosolic O-glycosylation is abundant in nerve terminals.* Journal of Neurochemistry, 2001. **79**(5): p. 1080-1089.
44. Link, A.J., ed. *Sample Preparation Methods for Mass Spectrometric Peptide Mapping Directly from 2-DE Gels*. Methods in Molecular Biology. Vol. 112: 2-D Proteome Analysis Protocols. 1999, Humana Press Inc.: Totowa. 513-521.
45. Bjork, I., B.A. Petersson, J. Sjoquist, *Some physicochemical properties of Protein A from Staphylococcus aureus.* European Journal of Biochemistry, 1972. **29**(3): p. 579-584.
46. Hirel, P.H., et al., *Extent of N-Terminal Methionine Excision from Escherichia-Coli Proteins Is Governed by the Side-Chain Length of the Penultimate Amino-Acid.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(21): p. 8247-8251.
47. Pappin, D.J.C., P. Hojrup, and A.J. Bleasby, *Rapid Identification of Proteins by Peptide-Mass Fingerprinting.* Current Biology, 1993. **3**(6): p. 327-332.
48. Comer, F.I. and G.W. Hart, *O-glycosylation of nuclear and cytosolic proteins - Dynamic interplay between O-GlcNAc and O-phosphate.* Journal of Biological Chemistry, 2000. **275**(38): p. 29179-29182.

49. Chou, T.Y., C.V. Dang, and G.W. Hart, *Glycosylation of the C-Myc Transactivation Domain.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(10): p. 4417-4421.