博士論文

Chemical synthesis of ubiquitin derivatives to understand the ubiquitin code and its role with nucleosome stability

（ユビキチンコードおよびヌクレオソーム安定性における役割を理解するためのユビキチン誘導体の化学合成）

Shivani Dixit
シヴァニ　ディキシット

Preface and Acknowledgements

The study presented in this thesis has been carried out under the direction of Professor Akimitsu Okamoto at the Department of Advanced Interdisciplinary Studies, Graduate School of Engineering, The University of Tokyo, from September 2017 to September 2020. The study concerns the Quantitative analysis of Ubiquitin and the effect of Ubiquitin on nucleosome stability.

The author wishes to express her profound and foremost gratitude to Professor Akimitsu Okamoto for his inspiring guidance and great support throughout this work. The author is also profoundly grateful to the other members of the thesis committee: Professor Shinsuke Sando, Associate Professor Fumiaki Ohtake, Associate Professor Tsuyoshi Osawa, Associate Professor Satoshi Yamaguchi.

The author extends her sincere gratitude to Associate Professor Gosuke Hayashi for his valuable advice and encouragement.

The author expresses her appreciation to Professor Yasuhi Saeki and Associate Professor Fumaki Ohtake in the Tokyo Metropolitan Institute of Medical Science and Hoshi University for conducting the Mass Spectrometry analysis, discussion, and encouragement.

The author thanks the past and present members of Okamoto Laboratory for their suggestions and cooperation, and with whom the author shared an enjoyable time. In particular, the author wishes to thank Assistant Professor Kunihiko Morihiro, Assistant Porfessor Takaumi Furuhata, Masafumi Yanase, Kenta Koyama, Naoki Kamo, Yoshiki Konda, Iori Murayama, Koki Nakatsu for their help and encouragement.

The author would like to express her gratitude to Chie Shishido for the kind help with official business. The author would like to thank JICA (Japan International Cooperation Agency) fellowship and SEUT-RA for financial support that made it possible to complete this study.

Finally, the author wishes to express her most profound gratitude for her parents S.K. Dixit and Sadhna Dixit, and her brother Shubham Dixit. They have supported her education and encouraged her affectionately. The author dedicates this thesis to her family.

Shivani Dixit
Okamoto Laboratory
Department of Advanced Interdisciplinary Studies
School of Engineering

The University of Tokyo, September 2020

Table of Contents

Chapter 1

General Introduction

1. Introduction

In eukaryotes, genome DNA is packaged in a higher-order structure called chromatin, which is composed of DNA, histone proteins, and other nuclear proteins(figure1)[1]. Alternation of biophysical and biochemical properties of chromatin regulates the elaborate mechanism in nucleus such as transcription and replication. As a marker and driving force for the dynamic change of chromatin conformation, DNA and nuclear proteins are covalently modified by utilizing cellular substrates such as small molecular metabolites and small proteins. The chromatin modification mainly comprises post-translational modifications[2]. Post-translational modifications are essential for the proper coordination of the cellular events and for understanding the complexity of the eukaryotic genome, which is a repertoire of many modifications. The eukaryotic genome can give information such as structural properties of the protein, the binding properties of proteins, and the function of proteins[2]. Histones, the backbone of the chromatin undergoes several PTM such as acetylation, ubiquitination, phosphorylation, methylation(figure1)[3]. Among these, the methylation and acetylation are the widely studied modifications[4]. Recent advances in the techniques and discoveries have led to the identification of the new PTMs and enabled them to understand their role in gene regulation[5]. The patterning and crosstalk between different PTMs affect the properties of the histone proteins and the interaction between the DNA and protein[6]. They are overall affecting the accessibility of the DNA binding to various regulations such as transcription, DNA repair, replication[5].

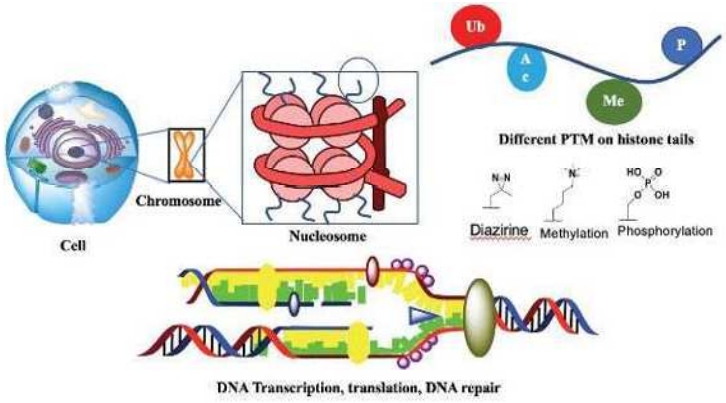


Figure1. Schematic representation of decoration of chromatin with post-translational modifications and gene regulation.

2. Ubiquitination

Ubiquitin is a 76 amino acid long protein of 8.5kDa, which attaches to intracellular proteins through a cascade of the enzyme complex[7]. Ubiquitination is an important recognition mark inducing protein degradation in the case of polyubiquitin chain formation and different biological processes in case of monoubiquitination and thus, affecting every pathway in the eukaryotic cell[8]. Ubiquitin regulates critical functions such as response to inflammation, degradation of proteins via proteasome-mediated degradation or lysosome mediated degradation, DNA repair, and functioning of certain transcription factors[9]. The fate of the ubiquitinated substrate is determined by the type of ubiquitin modification and chain length[10]. Ubiquitination also contributes to GPCR signaling, TGF-β, MAPK. It is also essential for the regulation of cellular homeostasis, cellular transport activity[11]. Various transcription factors such as E2F1, cell cycle progression, and DNA undergoes K11, K48, K63, K11/48 branched ubiquitination[12]. Another important protein Beclin1, which regulates autophagy, undergoes peculiar ubiquitin chains formation and controls autophagy[13]. Ubiquitination also plays a vital role in autoimmune and inflammatory responses such as allergy[14]. Ubiquitination also affects protein-protein interaction because of its large size, and it creates steric clashes, as seen in DIX-DIX interface[15].

Ubiquitination involves the bond formation between the C-terminus of Ubiquitin G76 and the ε-amino group of the target lysine residue. The reaction consists of a series of enzymes, i.e., an activating enzyme E1, which activates the carboxyl group of ubiquitin by the formation of the thiol ester for the nucleophilic attack(figure2)[16]. The E1 reactions begin with the binding of MgATP, and then the ubiquitin forms the ubiquitin adenylate, which serves as the intermediate donating the ubiquitin to the active site of E1. The structural analysis of the E1 enzyme suggests the ATP dependent conformation changes in the binding site of the enzyme enabling it to recognize the AMP-Ubiquitin complex[17]. Cysteine is the key residue present in the active site responsible for carrying out the catalysis. The E1 carries two forms of ubiquitin, and one is thiol form and other the adenylate form. The thiol form is transferred to the E2, the conjugating enzyme that transfers the activated ubiquitin. Although the number of E1 is less than the E2, the E1 has high catalytic activity in comparison with the E2 enzyme. The E2 binds tightly to the E1-Ubiquitin complex than free ubiquitin and free E1. E3 ligase plays an important role in transferring the activated ubiquitin from E2 to the target lysine on the substrate. The recognition of the substrate is determined by the sequence or structural motif, which is known as the "ubiquitination signals" and is recognised by the E3 ligase. Thus, the

E3 ligase plays an important role in guiding the ubiquitination process towards the target substrate. The E3 ligase has two main families which are known so far: Ubox/RING(Really Interesting New Gene)- and HECT(Homologous to the E6-AP Carboxyl Terminus)[18]. These families are categorized based on the domain present in the protein to conduct the ligase function. The HECT ligase has ~350 residues region with a conserved cysteine-rich region. These cysteine residues form the essential thiol ester essential for the transfer of ubiquitin from the E2 enzyme to the substrate. The RING group of E3 ligase is ~70 resides long, RING fingers have a set pattern of cysteine and histidine residues with a peculiar spacing forming a pattern needed as ligands for the coordination of the two Zinc ions which stabilizes the global conformation of the protein. Many diseases are associated with the miss regulation of these enzyme cascade[19].

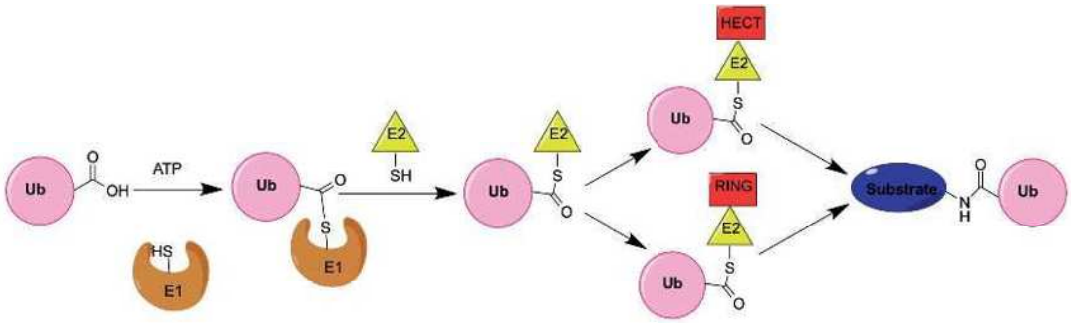


Figure2. Schematic representation of the ubiquitination process which involves a cascade of enzymes

2.1 Substrate of ubiquitination

The target substrates for ubiquitination can be histone proteins or ubiquitin itself. The substrates can undergo single ubiquitin modification or polyubiquitination. Histones H2A and H2B are the two most abundant substrates of ubiquitination in the nucleus. H2A corresponds to 5-15%, and H2B corresponds to 1-2% of ubiquitination in vertebrate cells and plays an important role in DNA repair, replication, and other important functions. H2A ubiquitination has been found at both N-terminal at K13 and K15 position and C-terminal K119 position[20]. Studies have suggested that the N-terminal methionine and lysine residues (M1, K6, K11, K29, K33, K48, K63) (figure3) on ubiquitin are also important targets for ubiquitination and play an important role in proteasome-mediated degradation and other cellular functions[21]. The size of the proteasomal substrate determines whether its degradation will be mediated by mono- or polyubiquitylation[21]. Other than lysine, some other amino acids targeted for ubiquitination are Ser, Thr, and Cys[22]. However, the biological role of these ubiquitinated amino acids is not known. Different polyubiquitin chains form different conformation, which is recognized by

different ubiquitin-binding domains[23]. Ubiquitin can form linear chains, branched chains which determine the fate of the target molecule[24]. The substrate carrying more ubiquitin molecules/longer chains is preferred over the substrate having fewer ubiquitin molecules/smaller chains[25]. There are two mechanisms for the ubiquitin assembly formation: 1. Sequential model. This is the most common and predominant method[25]. According to this method, only one ubiquitin molecule is transferred at a time to the substrate molecule. 2. En-bloc model. According to this method, a preassembled complex of ubiquitin chains on E2/E3 is transferred to the substrate[25]. Mass spectrometry could be used to determine the ubiquitin chain assembly pathway on the target substrate[25]. The mechanisms of ubiquitin chain assembly contribute to several factors, such as determining the rate of proteasome-mediated degradation. The rate of a substrate undergoing proteasome-mediated degradation is determined by the rate of attachment of the first ubiquitin molecule to the substrate[25].

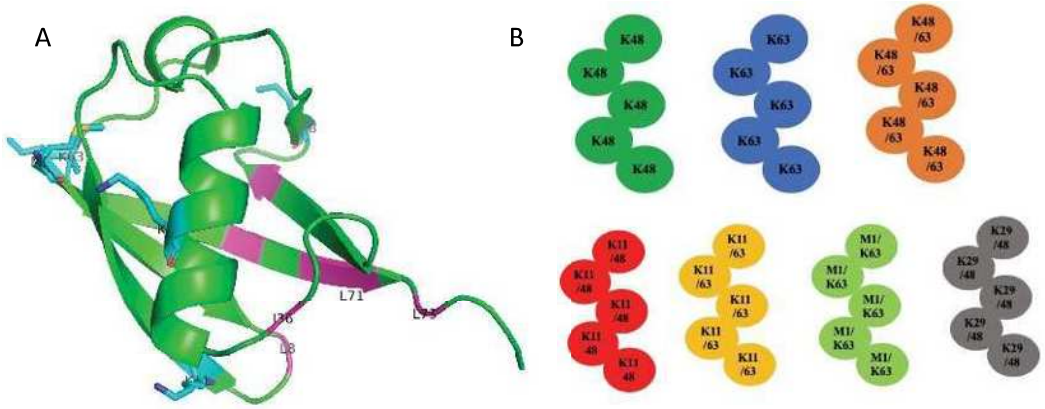


Figure3. A. Structure of ubiquitin (PDB:1UBQ). The target amino acid (M1, K11, K29, K48, K63) is shown in cyan color. The hydrophobic patches are shown in pink color. B. The target ubiquitin chains for this study are shown in the diagrammatic representation.

3. Chemical Protein synthesis

Total chemical synthesis benefits the preparation of proteins as it can introduce not only PTMs to specific sites but also other functional molecules to the proteins, such as fluorescent dyes, affinity tags, and photo crosslinkers. Chemical protein synthesis is mainly composed of the two processes, peptide synthesis, and peptide ligation.

Peptide chemistry has been widely used since solid-phase peptide synthesis (SPPS) was developed by Merrifield[26]. In general, SPPS resin beads with a 20-100um diameter couple with the C-terminal carboxylic acid of amino acid, and the peptide sequence elongates from C- to

N-terminus(figure4). An α-amino group in the amino acid is protected with *ter*-butyloxycarbonyl (Boc) or fluorenylmethyloxycarbonyl (Fmoc) group, which constitute the bedrock of SPPS. As for Boc-SPPS, trifluoroacetic acid (TFA) is required for the deprotection of the N-terminal Boc group, and hydrogen fluoride (HF) is responsible for the deprotection of side chains of amino acids as well as peptide cleavage from resin beads. These harsh reaction conditions make Boc-SPPS less ideal for the introduction of labile modifications (ex. *O*-glycosylation or phosphorylation) and highly reactive functional groups. In this case, Fmoc-SPPS is widely used in recent studies where the N-terminal Fmoc group is deprotected in basic conditions using piperidine. In the last step of SPPS, cleavage of a peptide from the resin and simultaneous deprotection of Boc, tert-butyl (tBu), trityl (Trt), and 2,2,4,6,7-pentamethyldihyobenzofuran-5-sulfonyl (Pbf) groups inside chains are achieved in acidic conditions using TFA[27].

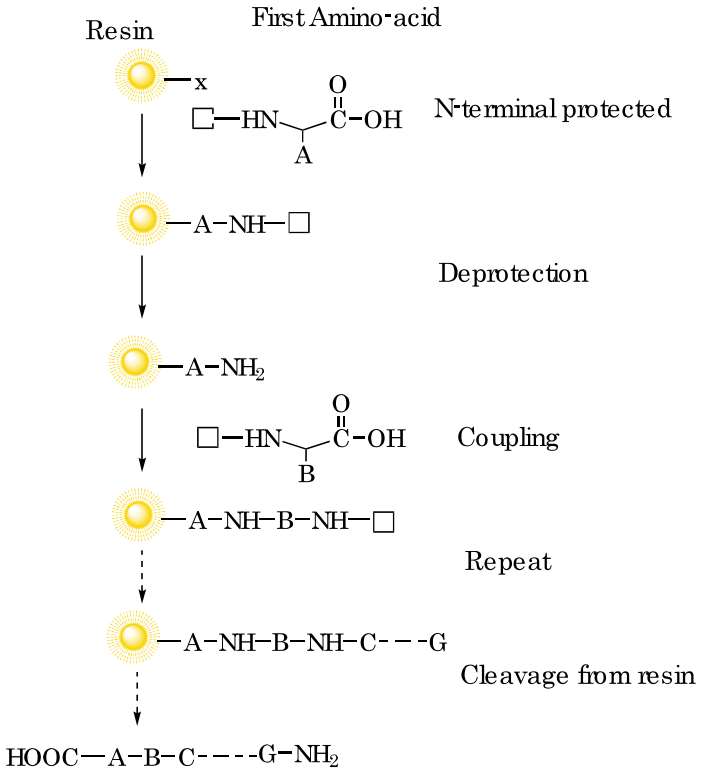


Figure 4. Solid Phase Peptide Synthesis

SPPS can routinely elongate a peptide sequence up to 50 amino acids, except for some problematic sequences. Although it is possible to synthesize several proteins over 100 amino acids all in once via SPPS, the potential of some longer sequence to form higher-order structure through the network of hydrogen bonds and hydrophobic effect increase significantly. This will result in undesired masking of the N-terminal group and stalling of the elongation.

Chemoselective ligation reaction between two peptide fragments has a great potential to synthesize long proteins. Thioester ligation developed by Aimoto and Hojo takes place between C-terminal carboxylic acid and N-terminal amine with the side chains protected[28]. However, this method cannot protect the side chains of Lys and Cys residues[29]. The most important breakthrough is native chemical ligation (NCL) reported by Kent and co-workers in 1995[30]. NCL is a selective reaction between C-terminal peptide thioester and N-terminal cysteine to form amide bind, through the exchange of the thioester and subsequent *S-N* acyl transfer (figure5). In this reaction, the protection of the side chains is not required, and the reaction proceeds in neutral conditions of an aqueous buffer.

On the other hand, the big limitation of NCL is the preparation of peptide thioester. To overcome this problem, a hydrazine derivative is used, which can be easily converted to the thioester form. Chlorotrityl resin is used for the introduction of the C-terminal hydrazine, and amino acids are coupled through standard Fmoc-SPPS. After the cleavage and purification of the peptide, sodium nitrite activates the hydrazide to form an azide, and subsequent addition of small thiol such as MPAA generates thioester *in situ* (figure6). Sequential NCL from N to C-terminus and one-pot ligation strategies have been performed because peptide hydrazine is stable under the NCL conditions.

The great efforts on protein chemistry greatly accelerated the synthesis of proteins in the last several years.

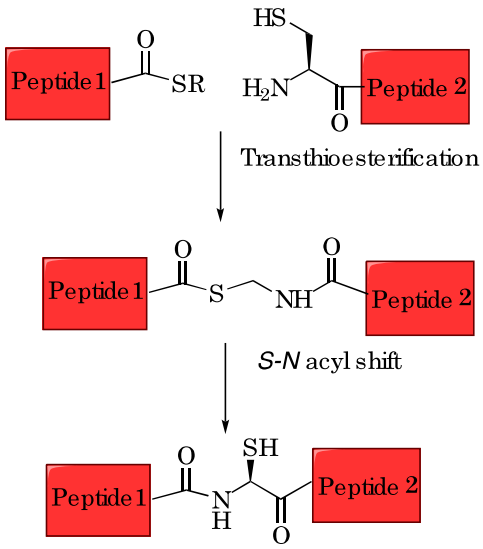


Figure5. Native chemical ligation

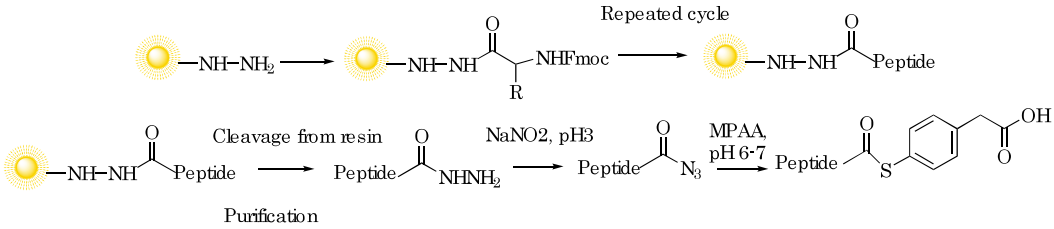
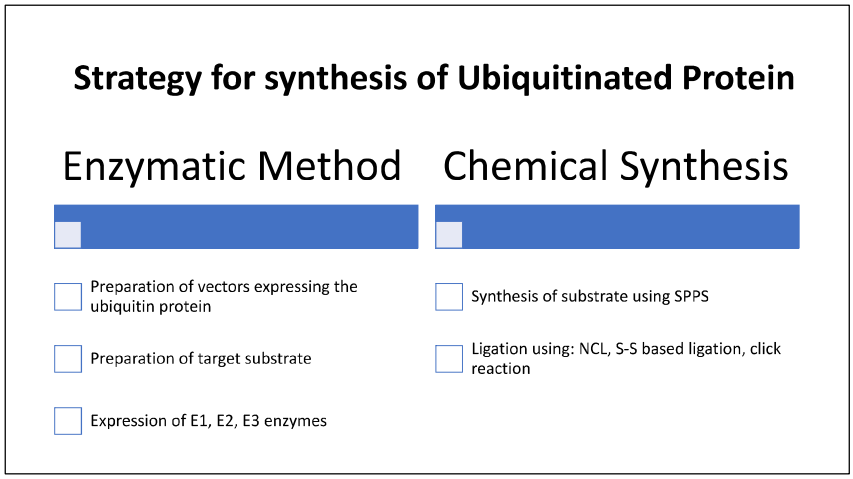


Figure 6. Synthesis of hydrazine peptide and peptide thioesterification

4. Ubiquitin Synthesis



4.1 Enzymatic synthesis of Ubiquitin chains

In 1997 by Pickart and coworkers reported the production of K48 chains of selected length[31] using the enzymatic method. This method was later used for the synthesis of K63 chains. This protocol involves the choice of E2, E3 enzymes to achieve the desired chain. E1 activates the Ub using ATP molecule, and E2 forms the specific isopeptide bond with the Ub molecules[17]. To conduct the enzymatic synthesis of ubiquitin chain, two Ub moiety are used one known as "proximal end" with its C terminal capped reversibly by adding extra amino acid D77 and the other known as "distal end" with the specific lysine residue been blocked by mutating the lysine with Cys or Arg, which can be recognized by the specific E2 enzyme[32]. After conducting the ubiquitination reaction, purification is needed to separate the reacted product from non-reacted Ub moiety. The use of mutant Ub moiety provides the reaction to be carried out in a controlled manner. However, the production of mutant Ub moiety is difficult then the production of wild type Ub and the mutant form might have some unknown effect on the properties of Ub. Yeast protease hydrolase-1(YUH1-1) can be used for the removal of D77,

and the treatment of ethyleneimine converts the Cys to *S*-aminoethylcysteine with amino group mimicking the lysine residue[32]. Later, a modified method using wild type Ub was used to generate the chains of K48, K63, K11[33]. These methods enabled the successful synthesis of different ubiquitin chains. The major limitation of this protocol was the use of multiple steps and blocking agents to obtain the ubiquitin chains with the desired length. To overcome this problem, Fushman's and Cropp's groups used pyrrolysyl-tRNA-synthetase to enable the incorporation of Lys(Boc) at the desired position[10]. Using this method several derivatives of ubiquitin chains have been synthesized till date[34].

4.2 Chemical synthesis of Ubiquitin chains

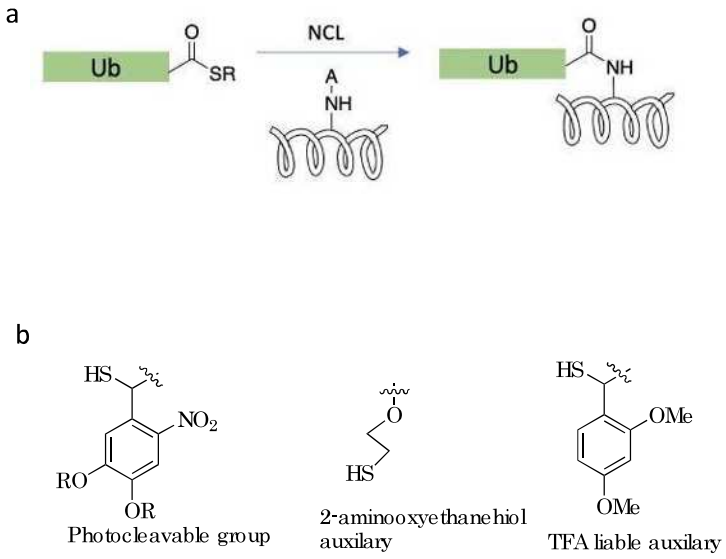
To study ubiquitination and its role, it is necessary to synthesis enough amount of each ubiquitin chain. Enzymatic method for production of the ubiquitin chain is difficult as all the enzymes known for each ubiquitin chain are not known, and it is difficult to control polyubiquitination and isolate enough amount of each chains[35]. The NCL was introduced by Dawson and co-worker in 1994[30](figure7a). The first chemical synthesis of ubiquitin was conducted by Briand *et al.* and Ramage *et al.* [36] Using the chemical synthesis method, ubiquitin can be conjugated with a small peptide such as Human influenza hemagglutinin (HA) or biotin and can be used for the identification of new target substrate or enzymes[37]. Many fluorogenic substrate have been conjugated to the ubiquitin using the chemical synthesis method for the identification of DUBs, understanding enzyme kinetics, identification of inhibitor molecules, synthesis of controlled ubiquitin chains, site-specific incorporation of thiolysine[37].

To assist in the native isopeptide backbone formation, some groups have used glycyl auxiliary δ-mercaptolysine analog or γ-mercaptolysine analogue[38](figure7b,c).These analogs provide the advantage of chemical synthesis over biological expression as they generate the native isopeptide bond, which is difficult to obtain for biologically expressed derivatives—using δ-mercaptolysine analog Brik group is successful in the synthesis of K48 linked tetra Ubiquitin α-Syn[39]. Studies have been conducted to increase the length in a controlled manner by using a less reactive alkylthioester for ligation. The major drawback of this method for the synthesis of long ubiquitin is the slow efficiency of the reaction.

To study and identification of new DUBs, the native isopeptide has been replaced by non-native bond by using disulfide, maleimide, oxime, triazole, thioester, Dha[40,41,42]. This non-native containing peptide has the major drawback of different chemical properties as of native isopeptide bond.

Glycyl auxiliary having thiol group are the most often used analog for the synthesis of ubiquitin moieties using the native chemical ligation. Using this auxiliary allows the site-specific ubiquitination with the formation of a native isopeptide bond. The auxiliary deprotection can be achieved by either irradiation of light in the case of photocleavable auxiliary or by treatment of acid in case of glycyl auxiliary(figure7b).

The first report of site-specific ubiquitination of peptides was conducted by Muir and co-workers, by using photocleavable auxiliary[43]. The photocleavable auxiliary helped in the site-specific introduction of the ubiquitin in the target peptide. They used this method for the synthesis of H2BK120ub[44]. Multistep and low ligation yield was the major drawback of using photocleavable group. Later Chatterjee and co-workers used 2-aminooxyethanethiol auxiliary for the synthesis[45]. Liu group synthesized a TFA liable auxiliary that overcomes the challenges of the previous method. The acid liable auxiliary is compatible with hydrazine mediated SPPS. Liu group was successful in the synthesis of diUb chains of K11, K48, K63, and other forms. Later they used this auxiliary for the synthesis of ubiquitinated histones for nucleosome study[40].



c

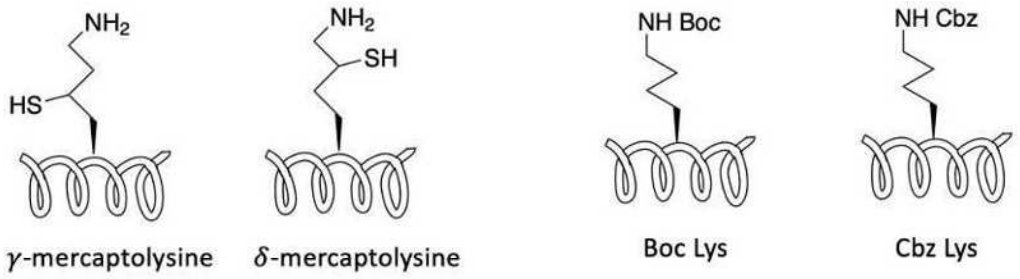


Figure 7. A, Schematic representation of the chemical synthesis of ubiquitin modified substrate. B, C, the list of auxiliary and lysine derivatives used for chemical synthesis of ubiquitin.

Using the method of NCL and SPPS, many ubiquitinated derivatives have been synthesized, such as p53, di-Ub, H2B, H2A[37]. The later derived ubiquitinated histones are precursors of ubiquitinated histones and play an important role in genome regulation. These synthesized derivatives have been used to understand the nucleosome structure and identification of the reader and erasers proteins.

To conduct the chemical synthesis of ubiquitin, the ubiquitin is divided into two fragments at A46 and mutated to Cys. Several Cys protecting group such as alloc cys, thz cys has been used to conduct the *N*-S acyl shift. The ligation is carried out at 6 M Gn 3 HCl, 200 mM phosphate buffer, pH 7.5, in the presence of MPAA. The reaction is monitored by HPLC and mass spectrometry, followed by purification of the product. Later, desulfurization is carried out under metal-free conditions. Using the scheme above mentioned derivatives such as K48 linked di-ub and others are synthesized with very high efficiency. This site-specific ligation, which a modified form of Cys based native chemical ligation, is known as isopeptide chemical ligation (IPL). Although the auxiliary mediated, ligation has increased the ability to achieve different target product but still has the drawback of low ligation yield and generates different side products due to hydrolysis.

1. Outlines of this thesis

In this research, we focused on the chemical synthesis of different ubiquitin derivatives. The target substrate for the synthesis is different ubiquitin chains and histone modified ubiquitin.
These ubiquitin chains would be used further to conduct different quantification analysis. This thesis mainly consists of two chapters, the contents of which are briefly summarized as follows:

In chapter 2, the author described the chemical synthesis of the ubiquitin derivatives. These ubiquitin derivatives have the linear and branched ubiquitin chains. These derivatives consist of the signature "GG-motif," which corresponds to the ubiquitin chain after trypsin digestion. After the successful synthesis of the ubiquitin chains, Mass spectrometry analysis was done. This analysis gave us the reference spectra, which could be used for conducting the quantification study of ubiquitin chains in the cell. Using these reference molecules, we were able to differentiate between linear chains and branched chains.

In chapter 3, the author described the chemical synthesis of H2AK119ubiquitin. The chemical synthesis of ubiquitinated histone is achieved using the native chemical ligation method. H2AK119 ubiquitination is one of the well-known gene repression marks. The synthesized ubiquitinated histone will be used to understand the stability of nucleosome bearing ubiquitinated histone H2A. These gene repression marks play an important role in regulating chromatin compaction and thus affect the down signaling processes. We hypothesized that these modifications together will affect the thermal stability of the nucleosome and would affect the binding of linker histone.

2. References:

1. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
2. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3**, 662–673 (2002).
3. Bowman, G. D. & Poirier, M. G. Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.* **115**, 2274–2295 (2015).
4. An, W. Histone Acetylation And Methylation BT - Chromatin and Disease. in (eds. Kundu, T. K. et al.) 355–374 (Springer Netherlands, 2007).
5. Kinam Park. The correlation between histone modifications and gene expression. *Bone* **23**, 1–7 (2014).
6. Nightingale, K. P. *et al.* Cross-talk between histone modifications in response to histone deacetylase inhibitors: MLL4 links histone H3 acetylation and histone H3K4 methylation. *J. Biol. Chem.* **282**, 4408–4416 (2007).
7. Swatek, K. N. & Komander, D. Ubiquitin modifications. *Cell Res.* **26**, 399–422 (2016).
8. Rape, M. Ubiquitylation at the crossroads of development and disease. *Nat. Rev. Mol. Cell Biol.* **19**, 59–70 (2018).
9. Nakamura, N. Ubiquitin system. *Int. J. Mol. Sci.* **19**, 1–5 (2018).
10. Marx, A. A., Stengel, F., Scheffner, M., Lutz, J. & Rolf, P. The length of a ubiquitin chain is a general determinant for selective recognition by ubiquitin-binding proteins. *Angew. Chemie Int. Ed.* **59**, 12371 (2020)
11. Medina, A. B., Banaszczak, M., Ni, Y., Aretz, I. & Meierhofer, D. ρ^0 Cells Feature De-Ubiquitination of SLC Transporters and Increased Levels and Fluxes of Amino Acids. *Int. J. Mol. Sci.* **18**, 879 (2017).
12. Dubrez, L. Regulation of E2F1 Transcription Factor by Ubiquitin Conjugation. *Int. J. Mol. Sci.* **18**, 2188 (2017).
13. Boutouja, F., Brinkmeier, R., Mastalski, T., El Magraoui, F. & Platta, H. W. Regulation of the Tumor-Suppressor BECLIN 1 by Distinct Ubiquitination Cascades. *Int. J. Mol. Sci.* **18**, 2541 (2017).

14. Wang, J. & Maldonado, M. A. The Ubiquitin-Proteasome System and Its Role in Inflammatory and Autoimmune Diseases. **3**, 255–261 (2006).
15. Schwarz-Romond, T. *et al.* The DIX domain of Dishevelled confers Wnt signaling by dynamic polymerization. *Nat. Struct. Mol. Biol.* **14**, 484–492 (2007).
16. Hershkos, A., Heller, H., Elias, S. & Ciechanover, A. Components of Ubiquitin-Protein Ligase System. **258**, 8206–8214 (1983).
17. Hershko, A., Leshinsky, E., Ganoth, D. & Heller, H. ATP-dependent degradation of ubiquitin-protein conjugates. *Proc. Natl. Acad. Sci.* **81**, 1619 LP – 1623 (1984).
18. Rotin, D. & Kumar, S. Physiological functions of the HECT family of ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **10**, 398–409 (2009).
19. Mattiroli, F. Understanding the mechanisms of histone 2A ubiquitination. 156, (2013).
20. Vissers, J. H. A., Nicassio, F., van Lohuizen, M., Di Fiore, P. P. & Citterio, E. The many faces of ubiquitinated histone H2A: Insights from the DUBs. *Cell Div.* **3**, 1–14 (2008).
21. Shabek, N. *et al.* Article The Size of the Proteasomal Substrate Determines Whether Its Degradation Will Be Mediated by Mono- or Polyubiquitylation. *Mol. Cell* **48**, 87–97 (2012).
22. McDowell, G. S. & Philpott, A. Non-canonical ubiquitylation: Mechanisms and consequences. *Int. J. Biochem. Cell Biol.* **45**, 1833–1842 (2013).
23. Husnjak, K. & Dikic, I. Ubiquitin-Binding Proteins: Decoders of Ubiquitin-Mediated Cellular Functions. *Annu. Rev. Biochem.* **81**, 291–322 (2012).
24. Shabek, N., Iwai, K. & Ciechanover, A. Ubiquitin is degraded by the ubiquitin system as a monomer and as part of its conjugated target. *Biochem. Biophys. Res. Commun.* **363**, 425–431 (2007).
25. Deol, K. K., Lorenz, S. & Strieter, E. R. Enzymatic Logic of Ubiquitin Chain Assembly. *Front. Physiol.* **10**, 1–14 (2019).
26. Meienhofer, J. *et al. SOLID PHASE SYNTHESIS WITHOUT REPETITIVE ACIDOLY SIS Preparation of Leucyl-Alanyl-Glycyl-Valine Using 9-Fluorenylmethyloxycarbonylamino Acids*. *Int. J. Peptide Protein Rex* vol. 13 (1979).
27. Isidro-Llobet, A., Álvarez, M. & Albericio, F. Amino Acid-Protecting Groups. *Chem. Rev.* **109**, 2455–2504 (2009).
28. Hojo, H. & Aimoto, S. Polypeptide synthesis using the S-alkyl thioester of a partially protected peptide segment. synthesis of the DNA-binding domain of c-Myb protein (142-193)-NH2. *Bulletin of the Chemical Society of Japan* vol. 64 111–117 (1991).

29. Hojo, H., Kawakami, T., Hiroyama, Y. & Aimoto, S. An N-protection free ligation of the peptide thioester and the peptide with an N-alkoxy- or N-aryloxyamino group at its N-terminus. *Tetrahedron Lett.* **58**, 4638–4641 (2017).
30. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science (80-.).* **266**, 776 LP – 779 (1994).
31. Piotrowski, J. *et al.* Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. *J. Biol. Chem.* **272**, 23712–23721 (1997).
32. Raasi, S. & Pickart, C. M. Ubiquitin chain synthesis. *Methods Mol. Biol.* **301**, 47–55 (2005).
33. Cook, W. J., Jeffrey, L. C., Carson, M., Chen, Z. & Pickart, C. M. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *J. Biol. Chem.* **267**, 16467–16471 (1992).
34. Faggiano, S., Alfano, C. & Pastore, A. The missing links to link ubiquitin: Methods for the enzymatic production of polyubiquitin chains. *Anal. Biochem.* **492**, 82–90 (2016).
35. Faggiano, S. & Pastore, A. The Challenge of Producing Ubiquitinated Proteins for Structural Studies. *Cells* **3**, 639–656 (2014).
36. Ramage, R. *et al.* Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem. J.* **299 (Pt 1**, 151–158 (1994).
37. Hameed, D. S., Sapmaz, A. & Ovaa, H. How chemical synthesis of ubiquitin conjugates helps to understand ubiquitin signal transduction. **28**, 3, 805–815 (2016)
38. Kumar, K. S. A., Spasser, L., Ohayon, S., Erlich, L. A. & Brik, A. Expeditious chemical synthesis of ubiquitinated peptides employing orthogonal protection and native chemical ligation. *Bioconjug. Chem.* **22**, 137–143 (2011).
39. Haj-Yahya, M. *et al.* Synthetic polyubiquitinated α-Synuclein reveals important insights into the roles of the ubiquitin chain in regulating its pathophysiology. *Proc. Natl. Acad. Sci.* **110**, 17726 LP – 17731 (2013).
40. Qi, Y. K. *et al.* Recent advances in the chemical synthesis and semi-synthesis of poly-ubiquitin-based proteins and probes. *Sci. China Chem.* **62**, 299–312 (2019).
41. Morimoto, D., Walinda, E., Sugase, K. & Shirakawa, M. Biological and physicochemical functions of ubiquitylation revealed by synthetic chemistry approaches. *Int. J. Mol. Sci.* **18**, (2017).
42. Long, L., Furgason, M. & Yao, T. Generation of nonhydrolyzable ubiquitin-histone mimics. *Methods* **70**, 134–138 (2014).
43. Chatterjee, C., McGinty, R. K., Pellois, J.-P. & Muir, T. W. Auxiliary-Mediated Site-

Specific Peptide Ubiquitylation. *Angew. Chemie Int. Ed.* **46**, 2814–2818 (2007).

44. McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G. & Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **453**, 812–816 (2008).
45. Weller, C. E., Huang, W. & Chatterjee, C. Facile synthesis of native and protease-resistant ubiquitylated peptides. *Chembiochem* **15**, 1263–1267 (2014).
46. Darwin, K. H. Prokaryotic ubiquitin-like protein (Pup), proteasomes and pathogenesis. *Nat. Rev. Microbiol.* **7**, 485–491 (2009).
47. Geis-Asteggiante, L., Lee, A. E. & Fenselau, C. *Analysis of the topology of ubiquitin chains*. *Methods in Enzymology* vol. 626 (Elsevier Inc., 2019).
48. Goldberg, A. L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–899 (2003).
49. Huang, X. & Dixit, V. M. Drugging the undruggables: exploring the ubiquitin system for drug development. *Cell Res.* **26**, 484–498 (2016).
50. Voorhees, P. M. & Orlowski, R. Z. THE PROTEASOME AND PROTEASOME INHIBITORS IN CANCER THERAPY. *Annu. Rev. Pharmacol. Toxicol.* **46**, 189–213 (2006).
51. Ozawa, T. Analysis of a multiqueue model for an ISDN access interface. *Perform. Eval.* **15**, 65–75 (1992).
52. Rose, A. & Mayor, T. Exploring the rampant expansion of ubiquitin proteomics. *Methods Mol. Biol.* **1844**, 345–362 (2018).
53. Wilkinson, K. D. Ubiquitination and deubiquitination: Targeting of proteins for degradation by the proteasome. *Semin. Cell Dev. Biol.* **11**, 141–148 (2000).
54. Rajalingam, K. & Dikic, I. SnapShot: Expanding the Ubiquitin Code. *Cell* **164**, 1074-1074.e1 (2016).
55. Okatsu, K. *et al.* Phosphorylated ubiquitin chain is the genuine Parkin receptor. *J. Cell Biol.* **209**, 111–128 (2015).
56. Chau, V. *et al.* A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science (80-).* **243**, 1576 LP – 1583 (1989).
57. Valkevich, E. M., Sanchez, N. A., Ge, Y. & Strieter, E. R. Middle-Down mass spectrometry enables characterization of branched ubiquitin chains. *Biochemistry* **53**, 4979–4989 (2014).
58. Bezstarosti, K., Wal, L. Van Der & Demmers, J. A. A. Detection of Protein Ubiquitination Sites by Peptide Enrichment and Mass Spectrometry. *J. Vis. Exp.* **157**,

1–9 (2020).

59. Bignell, G. R. *et al.* Identification of the familial cylindromatosis tumour-suppressor gene. *Nat. Genet.* **25**, 160–165 (2000).

60. Sylvestersen, K. B., Young, C. & Nielsen, M. L. Advances in characterizing ubiquitylation sites by mass spectrometry. *Curr. Opin. Chem. Biol.* **17**, 49–58 (2013).

61. Xiao, W. *et al.* Ac-LysargiNase Complements Trypsin for the Identi fi cation of Ubiquitinated Sites. **91**, 24, 15890–15898 (2019).

62. Swatek, K. N. *et al.* Insights into ubiquitin chain architecture using Ub-clipping. *Nature* **572**, 533–537 (2019).

63. Williamson, E. A., Wray, J. W., Bansal, P. & Hromas, R. Overview for the histone codes for DNA repair. *Prog. Mol. Biol. Transl. Sci.* **110**, 207–227 (2012).

64. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

65. Meas, R. & Mao, P. Histone ubiquitylation and its roles in transcription and DNA damage response. *DNA Repair (Amst).* **36**, 36–42 (2015).

66. Meas, R. & Mao, P. Histone ubiquitylation and its roles in transcription and DNA damage response. *DNA Repair (Amst).* **36**, 36–42 (2015).

67. Whitcomb, S. J. *et al.* Histone monoubiquitylation position determines specificity and direction of enzymatic cross-talk with histone methyltransferases Dot1L and PRC2. *J. Biol. Chem.* **287**, 23718–23725 (2012).

68. Shukla, A., Chaurasia, P. & Bhaumik, S. R. Histone methylation and ubiquitination with their cross-talk and roles in gene expression and stability. *Cell. Mol. Life Sci.* **66**, 1419–1433 (2009).

69. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).

70. Zhou, W. *et al.* Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. *Mol. Cell* **29**, 69–80 (2008).

71. Uckelmann, M. & Sixma, T. K. Histone ubiquitination in the DNA damage response. *DNA Repair (Amst).* **56**, 92–101 (2017).

72. Simon, J. A. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49**, 808–824 (2013).

73. Rougeulle, C. *et al.* Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol. Cell. Biol.* **24**, 5475–5484 (2004).

74. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science (80-.).* **293**, 1074 LP – 1080 (2001).
75. Tamburri, S. *et al.* Histone H2AK119 Mono-Ubiquitination is Essential for Polycomb-Mediated Transcriptional Repression. *bioRxiv* 690461 (2019)
76. Pengelly, A. R., Kalb, R., Finkl, K. & Müller, J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev.* **29**, 1487–1492 (2015).
77. Kasinath, V., Beck, C., Sauer, P., Poepsel, S. & Kosmatka, J. JARID2 and AEBP2 regulate PRC2 activity in the presence of H2A ubiquitination or other histone modifications. 1–20 (2020).
78. van Mierlo, G., Veenstra, G. J. C., Vermeulen, M. & Marks, H. The Complexity of PRC2 Subcomplexes. *Trends Cell Biol.* **29**, 660–671 (2019).
79. Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445–1459 (2014).
80. Cao, R. *et al.* Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science (80-.).* **298**, 1039 LP – 1043 (2002).
81. Cooper, S. *et al.* Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat. Commun.* **7**, 13661 (2016).
82. Cao, R., Tsukada, Y. I. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and hox gene silencing. *Mol. Cell* **20**, 845–854 (2005).
83. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
84. Nakagawa, T. *et al.* Deubiquitylation of histone H2A activates transcriptional initiation via. *Genes Dev.* 37–49 (2008)
85. M. Christopher, A. M. L. S. Recruiting Polycomb to chromatin. *Physiol. Behav.* **176**, 100–106 (2016).
86. Richly, H. *et al.* Transcriptional activation of polycomb-repressed genes by ZRF1. *Nature* **468**, 1124–1128 (2010).
87. Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
88. Yuan, W. *et al.* Dense chromatin activates polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science (80-.).* **337**, 971–975 (2012).

Chapter 2

Synthesis of GG-Ub derivatives to understand the Ubiquitin Code

Abstract

Ubiquitin has emerged as an important post-translational regulator that regulates numerous processes for the proper function of the cell. Recent advances in the detection techniques have made it possible to better understand this post-translational modification. Mass spectrometry and then the use of antibodies is one of the widely used methods for quantifying ubiquitin. In this study, we used Middle-Down Mass Spectrometry to elucidate the MS/MS patterns of each of the ubiquitin chains using the chemically synthesized ubiquitin derivatives. These chemically synthesized ubiquitin derivatives are the model molecules that represent the monoubiquitin and polyubiquitin chains as in linear form and branched form having the signature "GG motif." The MS/MS spectrum of each derivative will contribute to conducting the quantification study of the ubiquitin chains.

Introduction

Ubiquitin is a 76 amino acid long protein attached to various substrates with the help of a multi-enzyme cascade and regulates cellular activity in cells. In prokaryotic cells, Pup performs the same function as ubiquitin[1]. In eukaryotic cells, SUMO, Rub1, Need8 performs the complementary function as ubiquitin. These proteins known as ubiquitin-like proteins (Ubls) have limited sequence similarity to ubiquitin but consist of similar globular 3D structure the beta grasp fold, also known as ubiquitin fold[2]. Some of the important functions performed by ubiquitin chains include DNA repair, proteasome-mediated degradation, endocytosis, tumor morphology determination[2]. However, there are numerous functions linked to ubiquitin, which remains to be undetermined until now. Ubiquitin-Proteasome System regulates the degradation of defective proteins. Failure in the ubiquitin-proteasome system can result in several diseases such as neurodegenerative diseases, cancer, and is the target area for the development of new drugs[3]. Many drugs based on this system are in clinical trials, such as the use of inhibitors of thalidomide to treat cancers[4]. Thus it is necessary to understand the ubiquitin protease system to develop potential drugs and their role in cancer[5], autoimmunity other neurodegenerative diseases[6].

The different cellular function linked with ubiquitin signaling is determined by the chain topology and linkages of the ubiquitin chain diversity at a glance. The ubiquitin-protein attaches to the substrate via the action of a group of proteins: E1 Ub activating an enzyme, E2 Ub conjugating enzyme, E3 Ub ligating enzyme. These group of proteins acts together leading to the covalent attachment of the Ub to the target molecule via the formation of an isopeptide bond between the C-terminus(G76) of Ub and ε-amine of lysine amino acid[7]. Among these enzymes, the E3 enzyme plays an important role in the proper functioning of the Ubiquitin-Proteasome System. Ubiquitination is a reversible process and can be reversed back by the action of deubiquitinase (DUBs)[8]. In humans, only 2E1, over 30 E2, more than 600 E3, and around 90 DUBs are known to date[9]. Unlike other protein modifications, ubiquitination is a chain reaction that can result in the formation of monoubiquitin and polyubiquitin chain (branched chains or linear chains and anchored or unanchored and homotypic or heterotypic). Some of the widely studied ubiquitin chains are M1(Attachment of Ubiquitin to the N-terminus), K11, K29, K33, K48, K63 with each governing a particular function (Figure1)[9]. However, the precise function of all the ubiquitin chains present in eukaryotes is still unknown and need techniques to decipher the role. These polyubiquitin chains can be further modified,

such as phosphorylation of polyubiquitin chains is related to the genuine Parkin receptor, which is involved in the depolymerization of mitochondria[10].

Among the above-mentioned substrate, the most widely studied are K48, K63 linked polyUb. K48 Ub chains act as a principal signal for proteasome degradation[11]. At the same time, K63 polyUb chains act as nondegradative signals during DNA damage response and cytokine signalling[12].

To understand the relation between the chain topology with a particular biological function, there are three most common methods used for the analysis. The first includes the expression of Ub lysine to arginine mutant in different cell lines; second is the use of antibodies specific for each chain to identify the deep cellular ubiquitin, and third is the use of Mass Spectroscopy to analyze different chains[13,14]. These methods have the drawback of not been able to conduct the conformational analysis and hence show biased results. Ubiquitin being a bulkier post-translational modification than others such as phosphorylation is difficult to be analyzed by MS, and the amount of ubiquitin present in the cell is low as it is been degraded by the proteasome or used for the cell signalling process. Moreover, the modified lysine positions are also limited in cell.

In this study, we aim to overcome the drawback of the available methods used for the analysis of ubiquitin chains. For this, we have used the Middle Down Mass Spectroscopy based analysis of the chemically synthesized derivatives of the ubiquitin models.

Middle Down Mass Spectrometry involves the controlled digestion of ubiquitin by trypsin following the analysis of the fragments. Trypsin cleaves the bond between the Arg74 and Gly75 and creates the signature "GlyGly" motif. The modified lysine is resistant to trypsin cleavage and thus leaves the "GG-motif" at the modification site with the mass of +114.0429Da observed during MS/MS analysis. In place of trypsin, several other enzymes have been used, such as LysC, which leaves long signature peptide and LyargiNase, a mirrored trypsin protease[15,16,17]. This protease has been used in the studies of methylation and phosphorylation. A modified and high-performance Ac-LyargiNase showed high specificity for K63 chains in comparison to the other modification sites[17]. The use of trypsin or any other protease leads to internal fragmentation of the ubiquitin chain and thus leaving partial peptides of ubiquitin chains. Thus, leading to the loss of chain architecture and information of the branching information.

In a recent study to understand the ubiquitin architecture, an engineered viral protease Lbpro has been used[18]. This method overcomes the drawbacks of the previously available methods to characterize the ubiquitin chains such as reduction of background noise, robust,

simple, and retaining the chain architecture. Lbpro being specific to the C-terminal Arg and Gly bond helps in retaining the ubiquitin chain architecture. Although the use of Lbpro protease provides a better analysis of ubiquitin than trypsin mediated digestion, this method cannot be used to differentiate between different polyubiquitin chains and quantification of chains.

In this protocol, we use Middle Down Mass Spectrometry to conduct the quantification study. Middle Down Mass Spectrometry involves controlled trypsin digestion of the protein fragments. Using such conditions, it overcomes the limitation of Top-Down Mass Spectrometry, which is to evaluate the complex data because the intact protein is used for the analysis and loss of the chain architecture the limitation of Bottom-Up Mass Spectrometry. However, the Middle Down Mass Spectrometry has an advantage over the other available methods for conducting the proteomic study. The protocol of using the controlled trypsin digestion is complicated to be carried out. Therefore, we overcome this difficulty with the use of chemically synthesized ubiquitin derivatives.

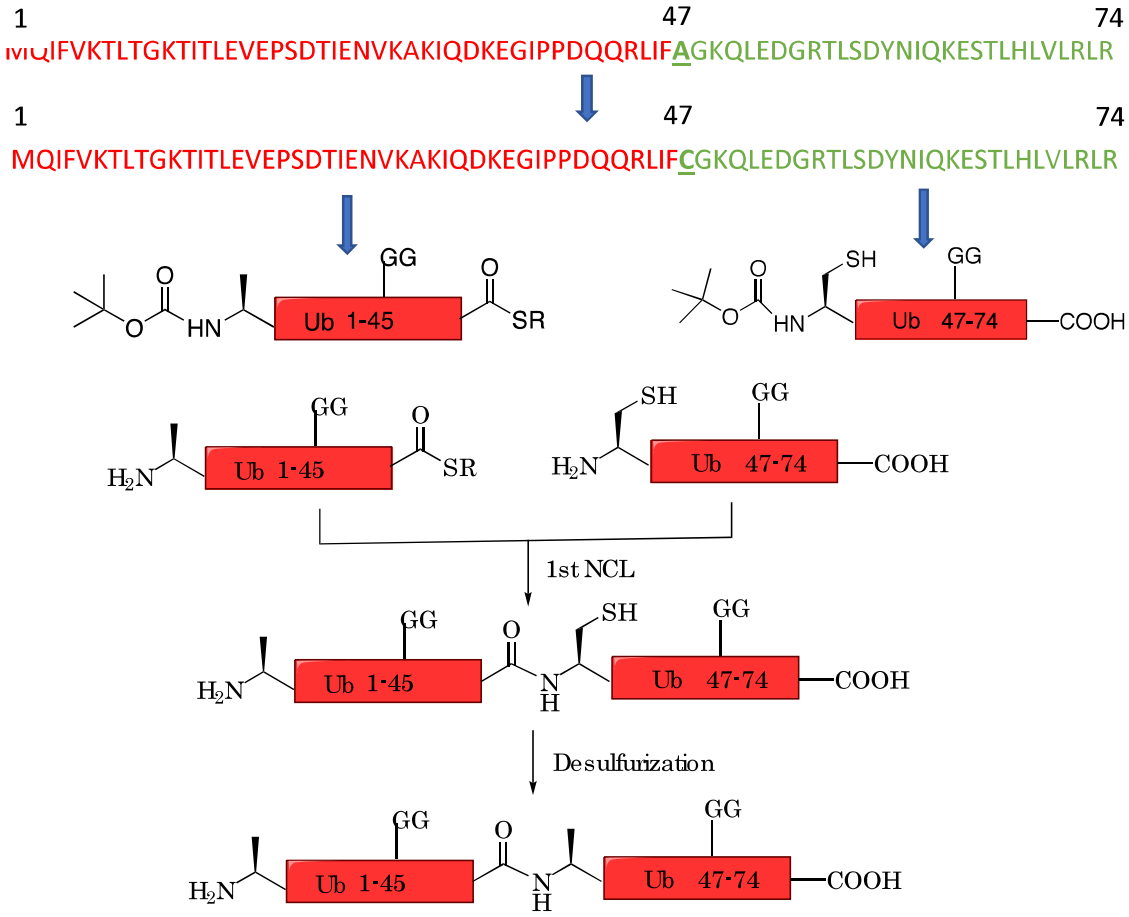
Many different forms of ubiquitin chains such as di-UbK48 have been synthesized earlier, but the ubiquitin molecule which mimics the trypsinized product of linear and branched chains has not been synthesized. Therefore, this is the first study of the synthesis of trypsinized ubiquitin chains and the use of Middle Down Mass Spectrometry to conduct the quantification analysis. The mass data obtained represents the relative intensity of each ubiquitin chain linear and branched-chain known to date. This data will help in determining how much the unmodified and modified chains of ubiquitin are present in the cell.

For conducting the quantification study, we have used the chemically synthesized ubiquitin as a reference molecule, which would help in the generation of the distinguished MS/MS spectra. The MS/MS spectra of the model would correspond to the individual chains of ubiquitin, and thus after isolating the chains from the cells, we would use these spectra as a reference and differentiate and quantify different ubiquitin chains present in the cell.

Results and Discussion

Ubiquitin consists of 76 amino acids, does not consist of Cys residues, which are both indispensable for NCL reaction and suitable for use as peptide cleavage sites. However, desulfurization reactions, in which Cys residues are transformed into Ala residues through radical reactions, have been developed to enable the design of peptide fragments at Ala residues. We envisaged dividing the Ubiquitin sequence into two fragments at A47(Scheme1). Each fragment was synthesized through Fmoc SPPS, purified by HPLC, and identified by MALDI-TOF MS(figure1). The isolated yield of each fragment is mentioned in table1. The N-terminal

and C-terminal were assembled to generate 74 amino acid long ubiquitin derivatives in the presence of 4-mercatophenylacetic acid (MPAA), which accelerates (figure2,3) NCL. The isolated yield of different ubiquitin derivatives generated is mentioned in table2 after purification, and lyophilization of the ubiquitin derivatives desulfurization was initiated by VA-044 (2,2`-azobis[2-(2-imidazoline-2-yl)-propane] dihydrochloride) producing ubiquitin derivatives which were purified by HPLC and identified by MALDI-TOF MS.



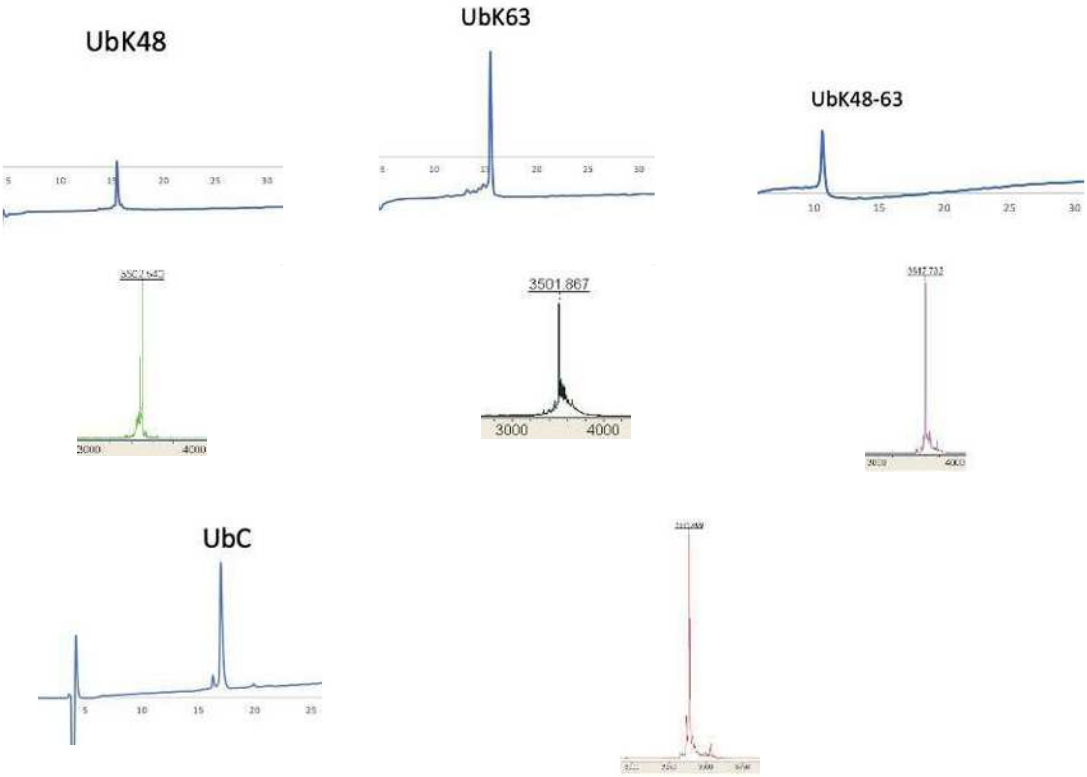
Scheme 1. Synthetic route for the synthesis of the Ubiquitin derivatives. The GG amino acid is attached to the desired position, such as M1, K11, K29, K48, K63.

Table 1. List of fragments used for the synthesis of ubiquitin chains with the terminology used, the isotope-labeled amino acid is in red color.

Light labeled	Sequence	Terminology	Yield
	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIF	UbN	8%
	MQIFVKTLTGK(G1G2)TITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIF	UbK11	11%
	MQIFVKTLTGKTITLEVEPSDTIENVKAK(G1G2)IQDKEGIPPDQQRLIF	UbK29	7%
	(G1G2)MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIF	UbM1	5%
	CGK(G1G2)QLEDGRTLSDYNIQKESTLHLVLRLR	UbK48	14.2%
	CGKQLEDGRTLSDYNIQK(G1G2)ESTLHLVLRLR	UbK63	14.5%
	CGK(G1G2)QLEDGRTLSDYNIQK(G1G2)ESTLHLVLRLR	UbK48-63	15%
	CGKQLEDGRTLSDYNIQKESTLHLVLRLR	UbC	25%

Heavy labeled	Sequence	Terminology	Yield
	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIF	UbN	6.3%
	MQIFVKTLTGK(G1G2)TITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIF	UbK11	8.4%
	MQIFVKTLTGKTITLEVEPSDTIENVKAK(G1G2)IQDKEGIPPDQQRLIF	UbK29	4.7%
	(G1G2)MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIF	UbM1	5.6%

Purity check of Ubiquitin C-terminal fragments



Purity check of Ubiquitin N-terminal fragments

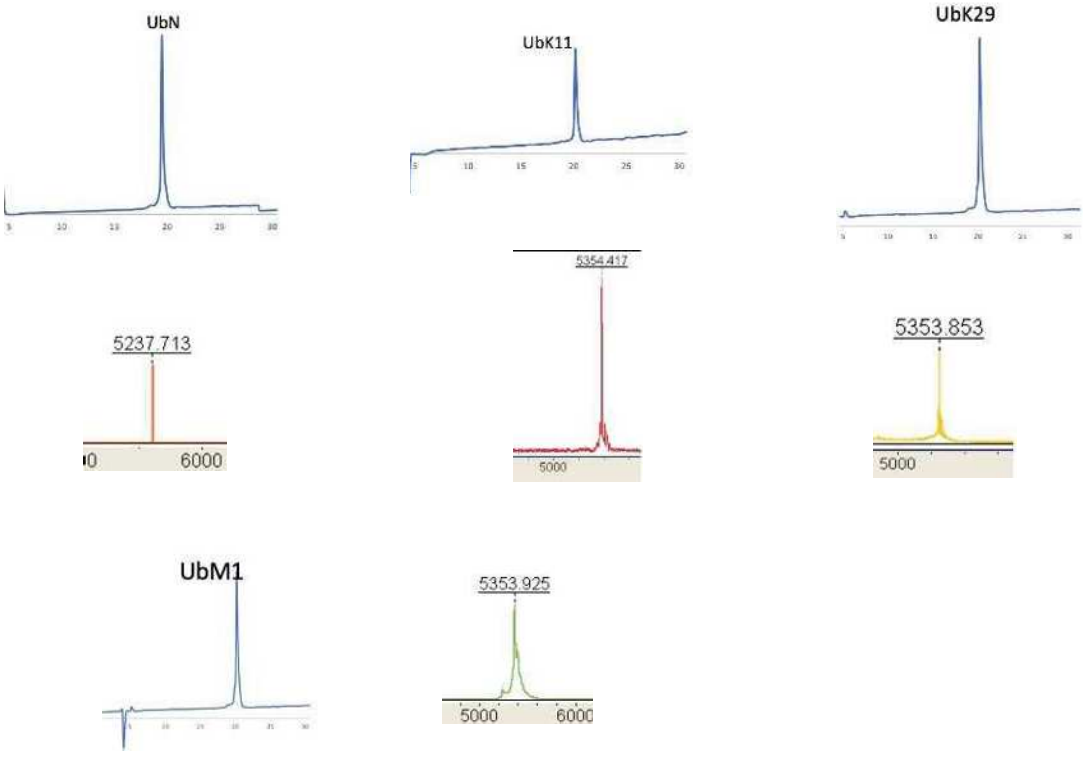


Figure 1. Synthesis of N-terminal and C-terminal peptide fragments. HPLC charts of the purified fragments were shown in blue. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1% TFA. Gradient 20-45%, 30min. MALDI-TOF MS of each fragments are as follows: UbN, observed mass 5237.71 (calcd 5237.05), UbK11 observed mass 5354.41 (calcd 5352.15), UbK29 observed mass 5353.85 (calcd 5352.15), UbM1 observed mass 5353.92 (calcd 5353.15), UbK48 observed mass 3502.54 (calcd 3501.94), UbK63 observed mass 3501.86 (calcd 3501.94), UbK48/63 observed mass 3617.73 (calcd 3617.04)

Purity check of isotope-labeled Ubiquitin N-terminal fragments

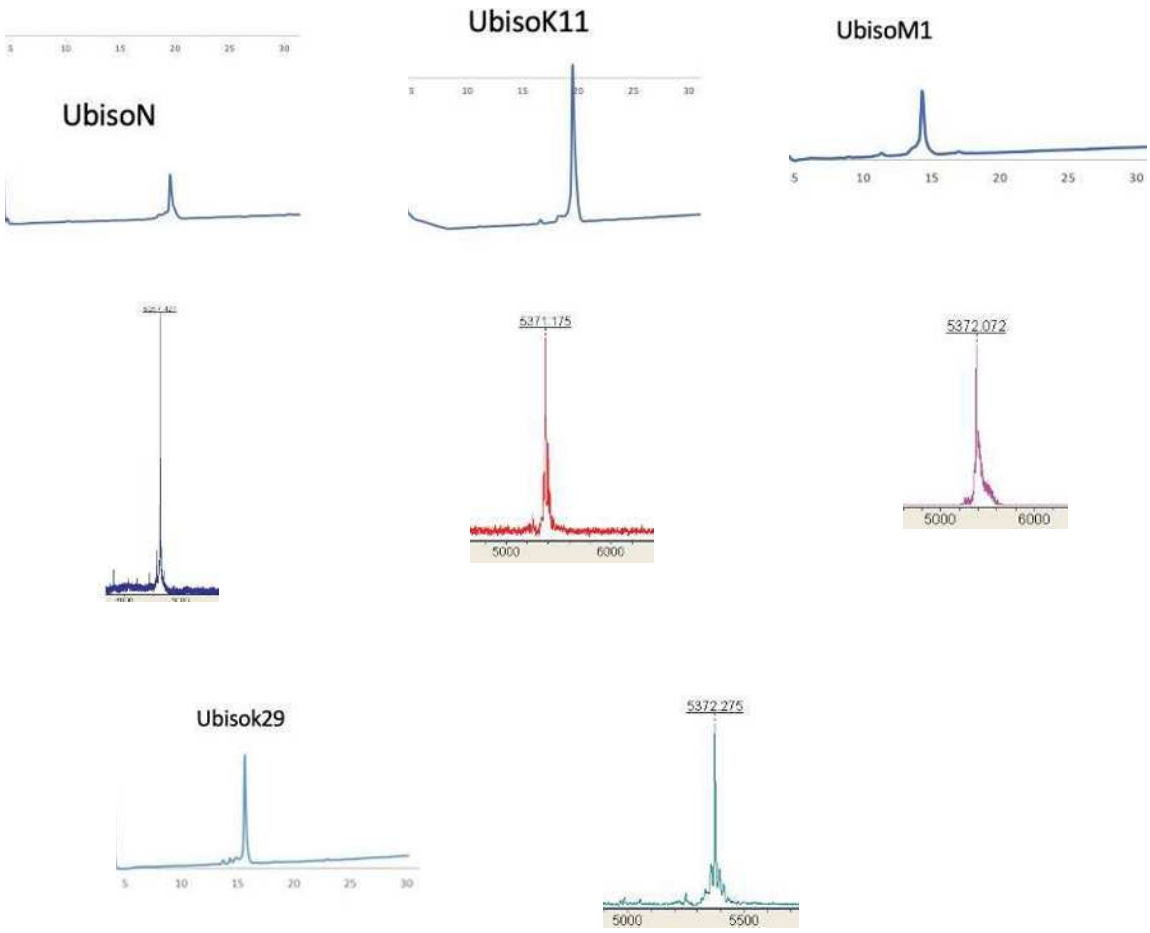
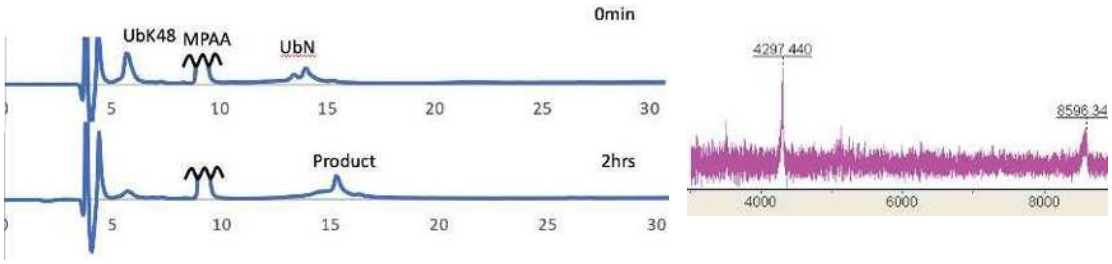


Figure 2. Synthesis of isotope-labeled N-terminal fragments. HPLC charts of the purified fragments were shown in blue. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1% TFA. Gradient 20-45%, 30min. MALDI-TOF MS of

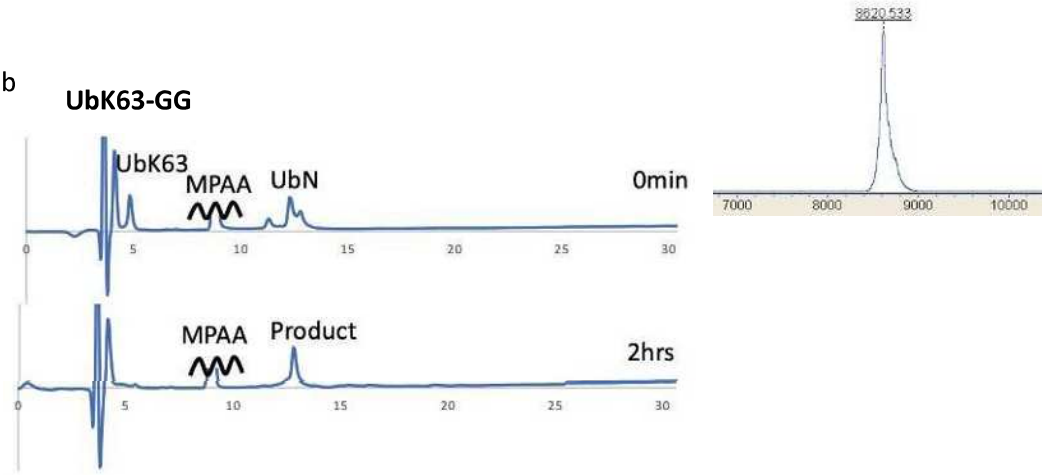
each fragment is as follows: UbisoN, observed mass 5257.42 (calcd 5256.05), UbisoK11 observed mass 5371.17 (calcd 5370.15), UbisoK29 observed mass 5372.27 (calcd 5370.15), UbisoM1 observed mass 5372.07 (calcd 5370.15).

Synthesis of different ubiquitin derivatives Ub(1-74)-GG as linear chains and Ub(1-74)-2GG as branched chains

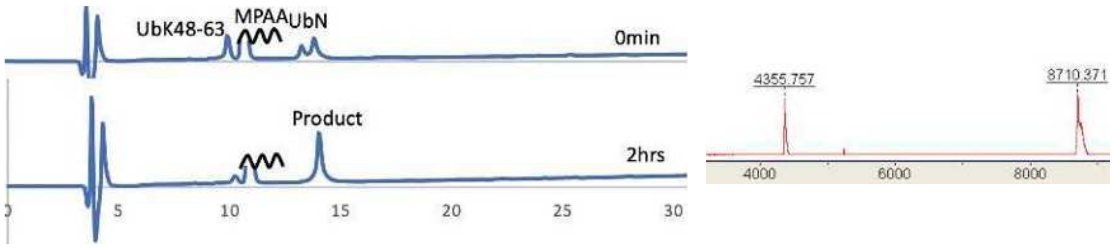
a **UbK48-GG**



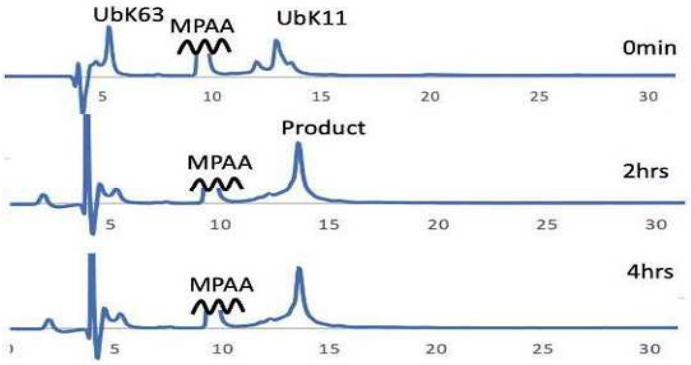
b **UbK63-GG**

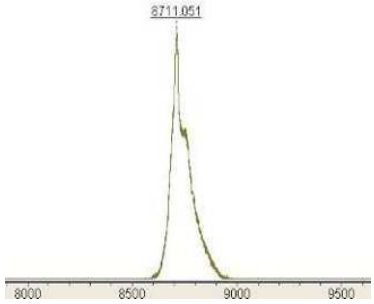


c **UbK48/63-2GG**

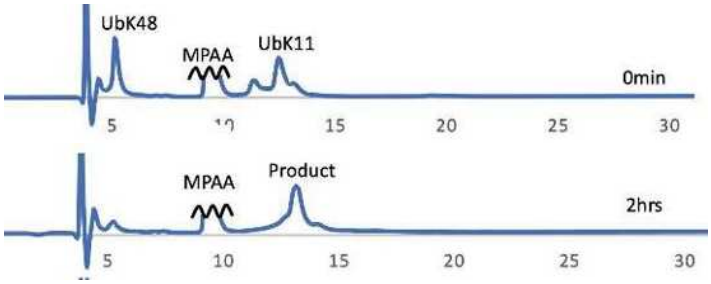


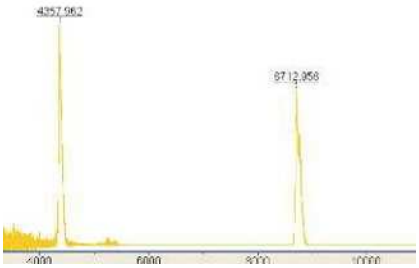
d **UbK11/63-2GG**



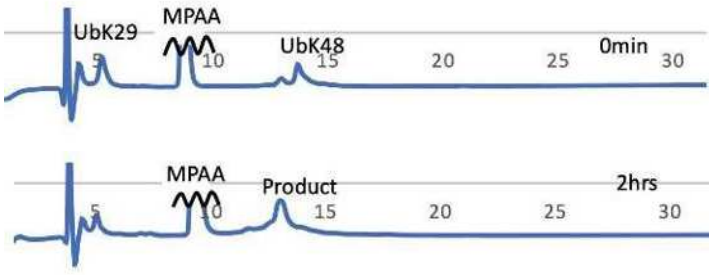


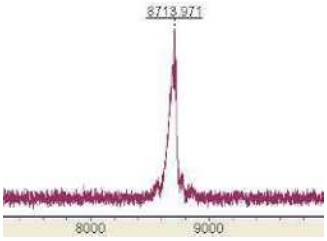
e **Ub11/K48-2GG**





f **UbK29/48-2GG**





g **UbM1/63-2GG**

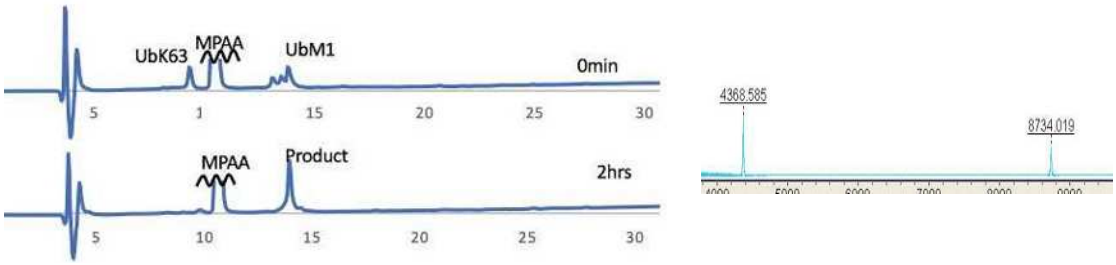


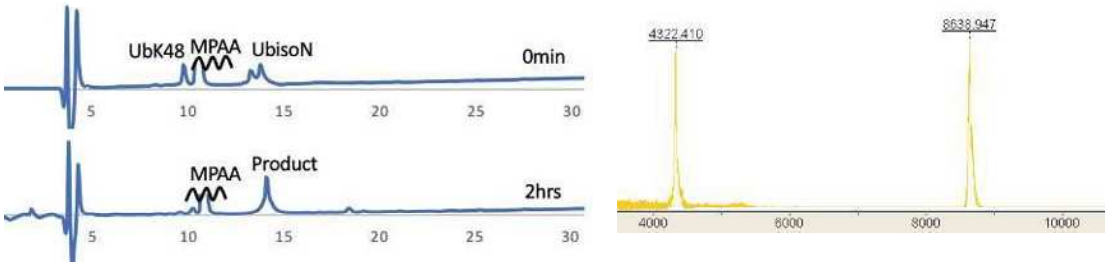
Figure 3. HPLC charts for the NCL of N-terminal and C-terminal to synthesize the ubiquitin derivatives with desired modifications. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1% TFA gradient 30-40%, 30min. MALDI-TOF MS of each spectrum was shown.

Linear chains: UbK48-GG[Peptide+H]$^+$,UbK63-GG [Peptide+Na]$^+$ observed: 8596.34(calcd 8596.81), 8620.53(calcd 8619.81) respectively.

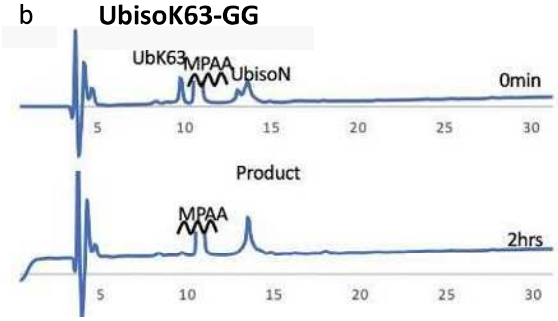
Branched chains: UbK48/63-2GG[Peptide+H]$^+$, UbK11/K63-2GG[Peptide+H]$^+$, UbK11/K48-2GG[Peptide+H]$^+$, UbK29/K48-2GG[Peptide+H]$^+$, UbM1/K63-2GG[Peptide+Na]$^+$ observed: 8710.37(calcd 8711.91), 8711.05(calcd 8711.91), 8712.95(calcd 8711.91), 8713.97(calcd 8711.91), 8734.01(calcd 8734.91) respectively.

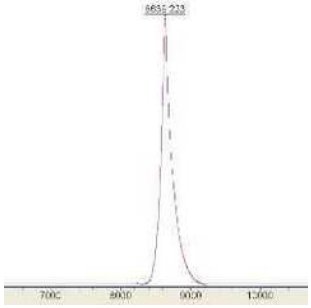
Synthesis of isotope-labeled ubiquitin derivatives Ubiso(1-74)-GG as linear chains and Ubiso(1-74)-2GG as branched chains

a **UbisoK48-GG**

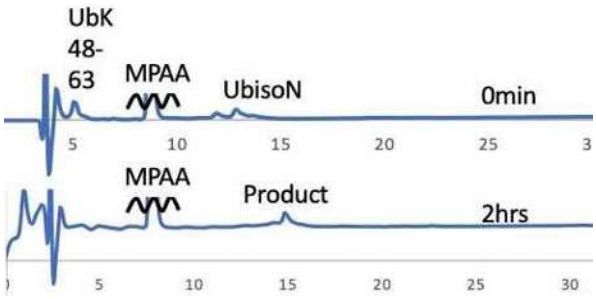


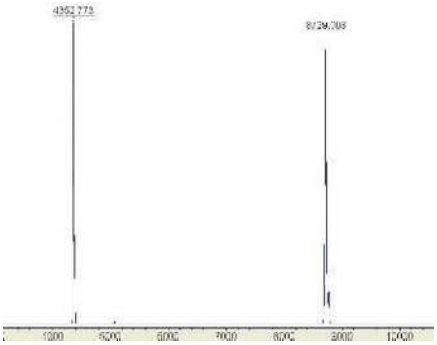
b **UbisoK63-GG**



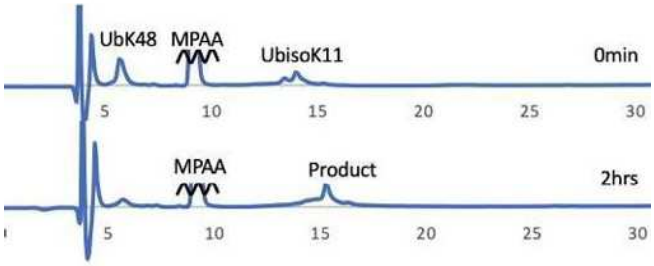


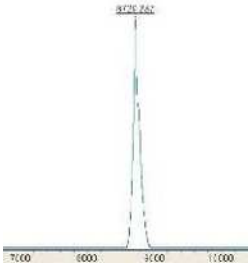
c **UbisoK48/K63-2GG**



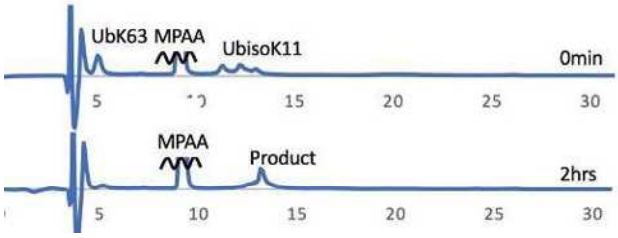


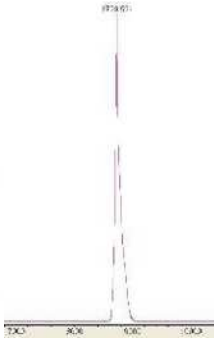
d **UbisoK11/48-2GG**





e **UbisoK11/63-2GG**





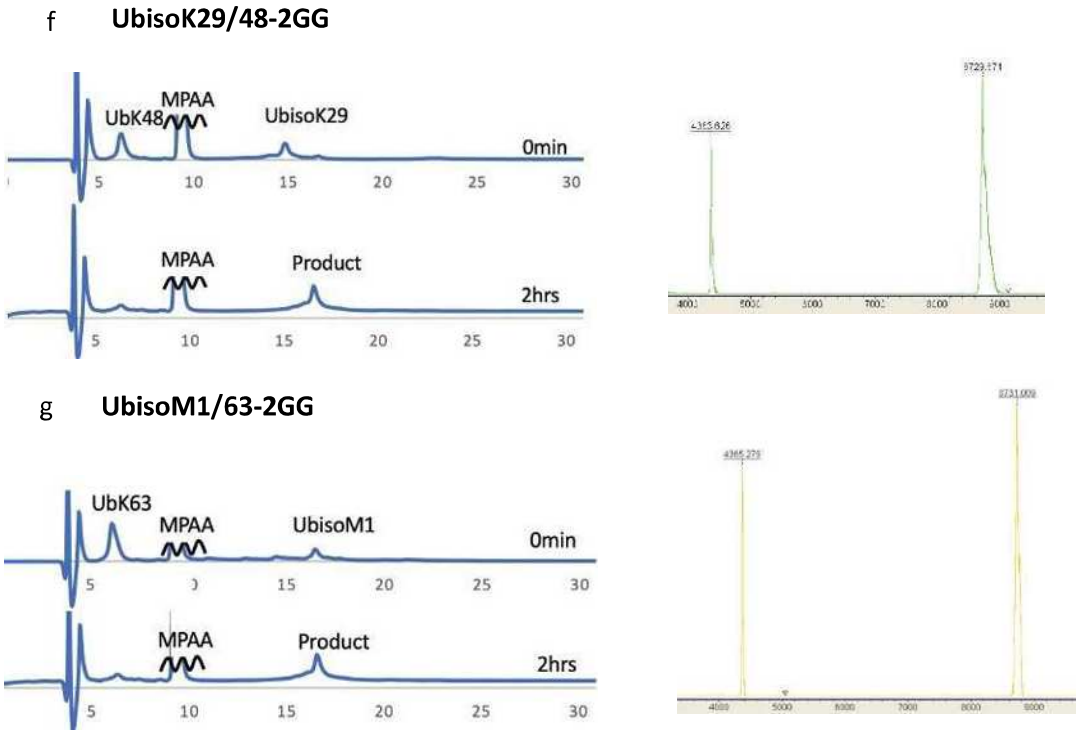


Figure 4. HPLC charts for the NCL of N-terminal and C-terminal to synthesize the ubiquitin derivatives with desired modifications. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1% TFA gradient 30-40%, 30min. MALDI-TOF MS of each spectrum was shown.

Linear chains: A,UbK48-GG[Peptide+Na]$^+$, B,UbK63-GG [Peptide+Na]$^+$ observed: 8638.94(calcd 8638.81), 8638.24(calcd 8638.81) respectively.

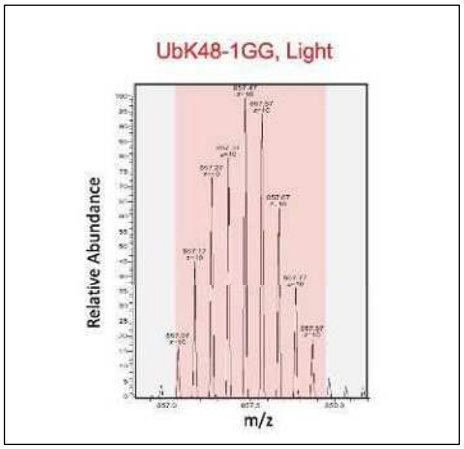
Branched chains: C,UbK48/63-2GG[Peptide+H]$^+$, D,UbK11/K48-2GG[Peptide+H]$^+$, E, UbK11/K63-2GG[Peptide+H]$^+$, F, UbK29/K48-2GG[Peptide+H]$^+$, G,UbM1/K63-2GG[Peptide+H]$^+$ observed: 8729.00(calcd 8728.91), 8729.33(calcd 8728.91), 8729.52(calcd 8728.91), 8729.69(calcd 8728.91), 8731.00(calcd 8728.91) respectively.

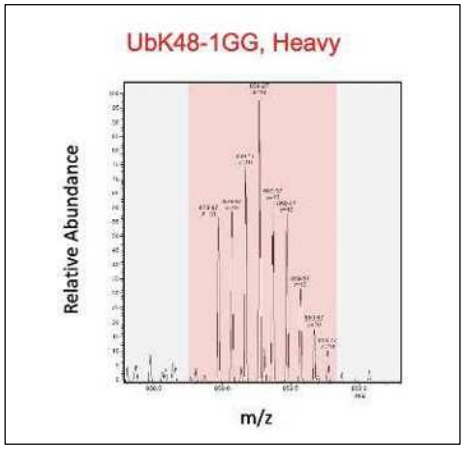
Table 2. The isolated yield of each ubiquitin derivatives synthesized by NCL

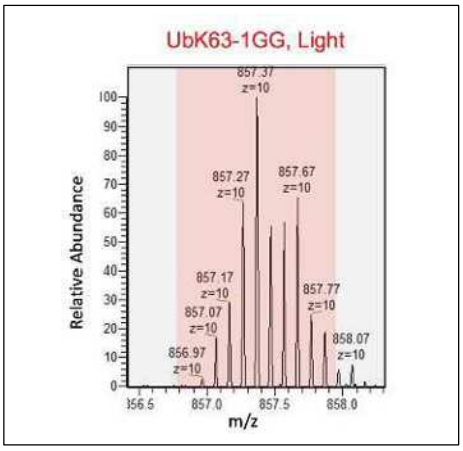
Light labeled	**Ubiquitin derivatives**	**Isolated yield**
Ub(1-74)-GG	UbK48-GG	15%
	UbK63-GG	14%
Ub(1-74)-2GG	UbK48/K63-2GG	12%
	UbM1/K63-2GG	9%
	UbK29/K48-2GG	17%
	UbK11/K48-2GG	10%
	UbK11/K63-2GG	10%

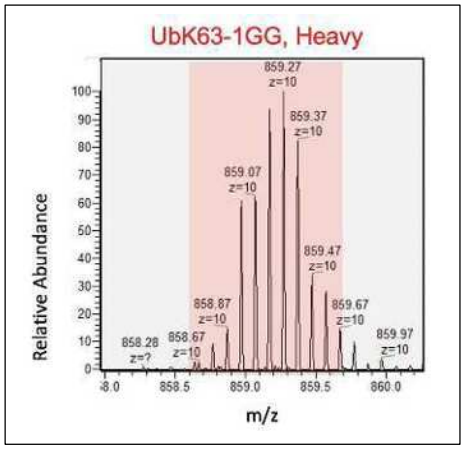
Heavy labeled	**Ubiquitin derivatives**	**Isolated yield**
Ubiso(1-74)-GG	UbK48-GG	15%
	UbK63-GG	13%
Ubiso(1-74)-2GG	UbK48/K63-2GG	11%
	UbM1/K63-2GG	8%
	UbK29/K48-2GG	15%
	UbK11/K48-2GG	7%
	UbK11/K63-2GG	8%

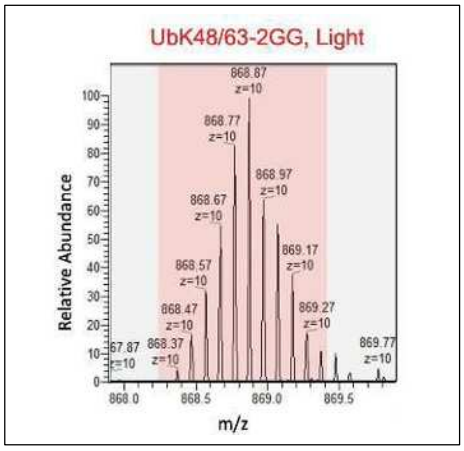
After the chemical synthesis of the desired ubiquitin derivatives mimicking as linear and branched ubiquitin chains as Ub(1-74)-GG and Ub(1-74)-2GG respectively, they were used for conducting the Middle down mass spectrometry analysis. For doing so, the lyophilized peptides were dissolved in a mixture of ACN/TFA and injected to Orbitrap Fusion LUMOS mass spectrometer. The simultaneous analysis was done for the light labeled and heavy labeled derivatives. The MS spectra showed the successful synthesis of the reference model for conducting the quantification study of ubiquitin chains

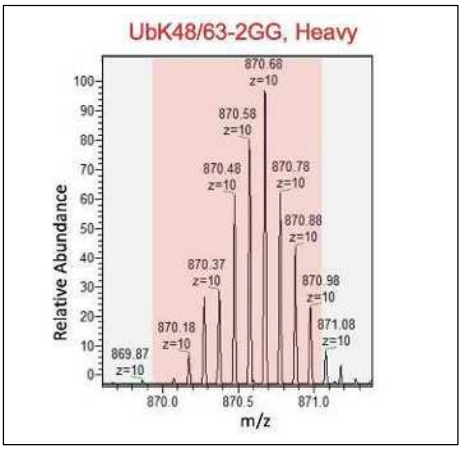

UbK48-1GG, Light
Relative Abundance
m/z

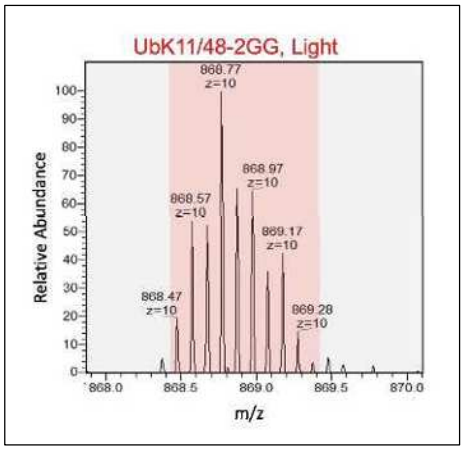

UbK48-1GG, Heavy
Relative Abundance
m/z

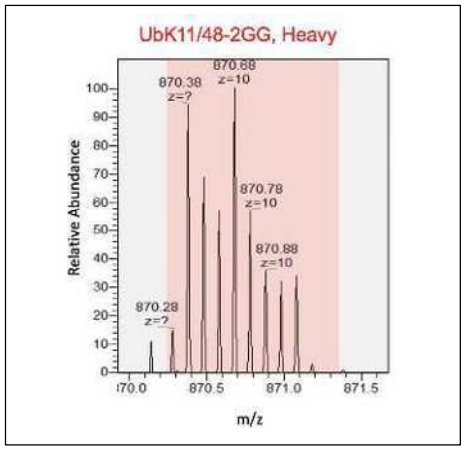

UbK63-1GG, Light
Relative Abundance
857.37 z=10
857.27 z=10
857.67 z=10
857.17 z=10
857.77 z=10
857.07 z=10
856.97 z=10
858.07 z=10
m/z

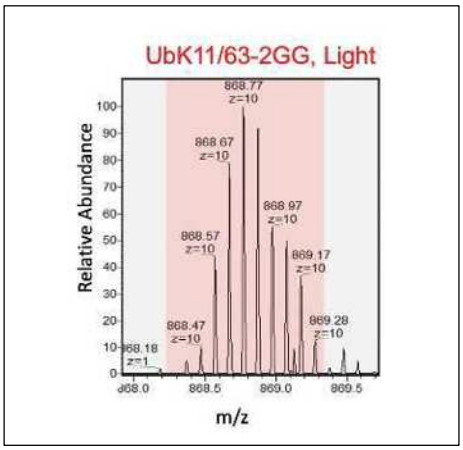

UbK63-1GG, Heavy
Relative Abundance
859.27 z=10
859.37 z=10
859.07 z=10
859.47 z=10
858.87 z=10
859.67 z=10
858.28 z=?
858.67 z=10
859.97 z=10
m/z

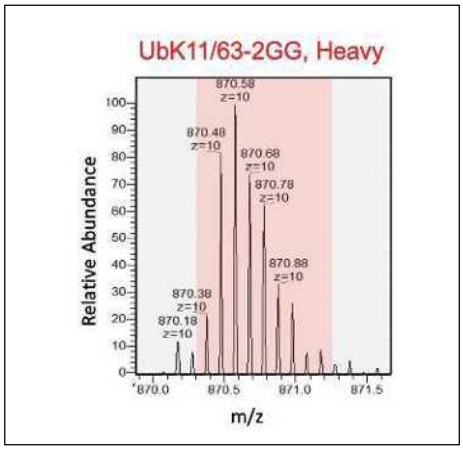

UbK48/63-2GG, Light
Relative Abundance
868.87 z=10
868.77 z=10
868.97 z=10
868.67 z=10
869.17 z=10
868.57 z=10
868.47 z=10
869.27 z=10
868.37 z=10
869.77 z=10
m/z

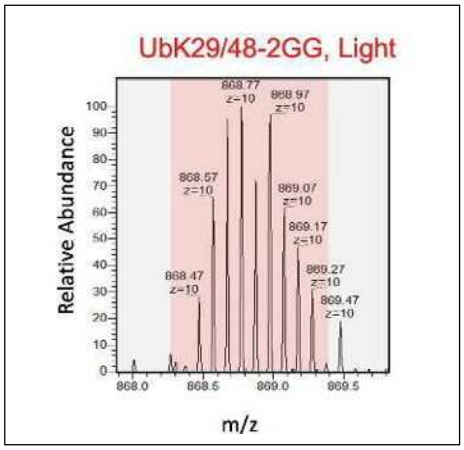

UbK48/63-2GG, Heavy
Relative Abundance
870.68 z=10
870.58 z=10
870.48 z=10
870.78 z=10
870.88 z=10
870.37 z=10
870.98 z=10
870.18 z=10
871.08 z=10
869.87 z=10
m/z

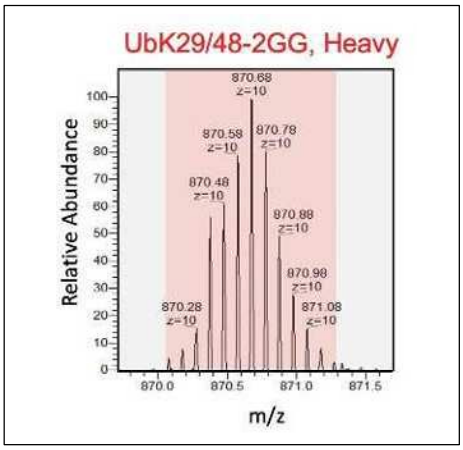

UbK11/48-2GG, Light
Relative Abundance
m/z
868.77 z=10
868.97 z=10
868.57 z=10
869.17 z=10
868.47 z=10
869.28 z=10


UbK11/48-2GG, Heavy
Relative Abundance
m/z
870.68 z=10
870.38 z=?
870.78 z=10
870.88 z=10
870.28 z=?


UbK11/63-2GG, Light
Relative Abundance
m/z
868.77 z=10
868.67 z=10
868.97 z=10
868.57 z=10
869.17 z=10
868.47 z=10
869.28 z=10
868.18 z=1


UbK11/63-2GG, Heavy
Relative Abundance
m/z
870.58 z=10
870.48 z=10
870.68 z=10
870.78 z=10
870.88 z=10
870.38 z=10
870.18 z=10


UbK29/48-2GG, Light
Relative Abundance
m/z
868.77 z=10
868.97 z=10
868.57 z=10
869.07 z=10
869.17 z=10
868.47 z=10
869.27 z=10
869.47 z=10


UbK29/48-2GG, Heavy
Relative Abundance
m/z
870.68 z=10
870.58 z=10
870.78 z=10
870.48 z=10
870.88 z=10
870.98 z=10
870.28 z=10
871.08 z=10

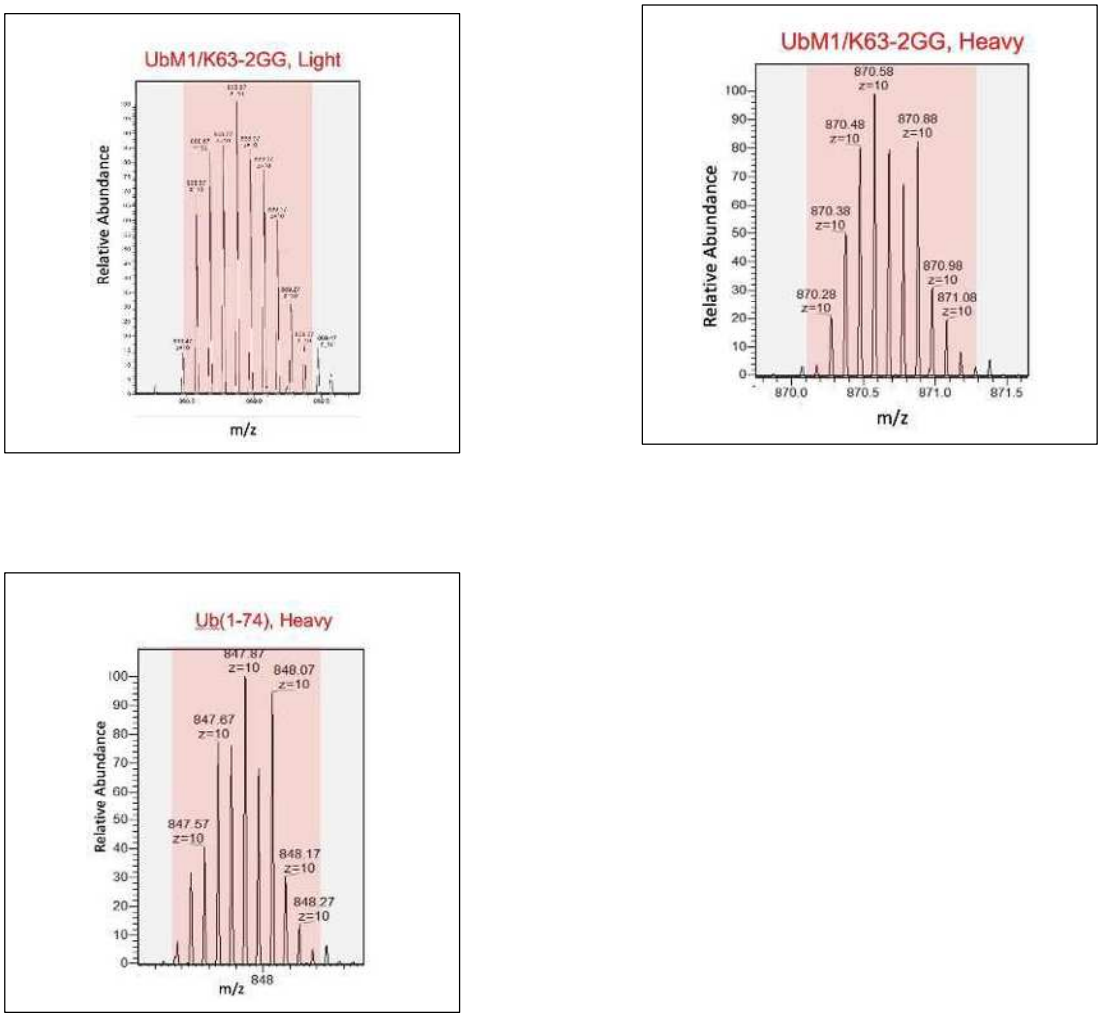
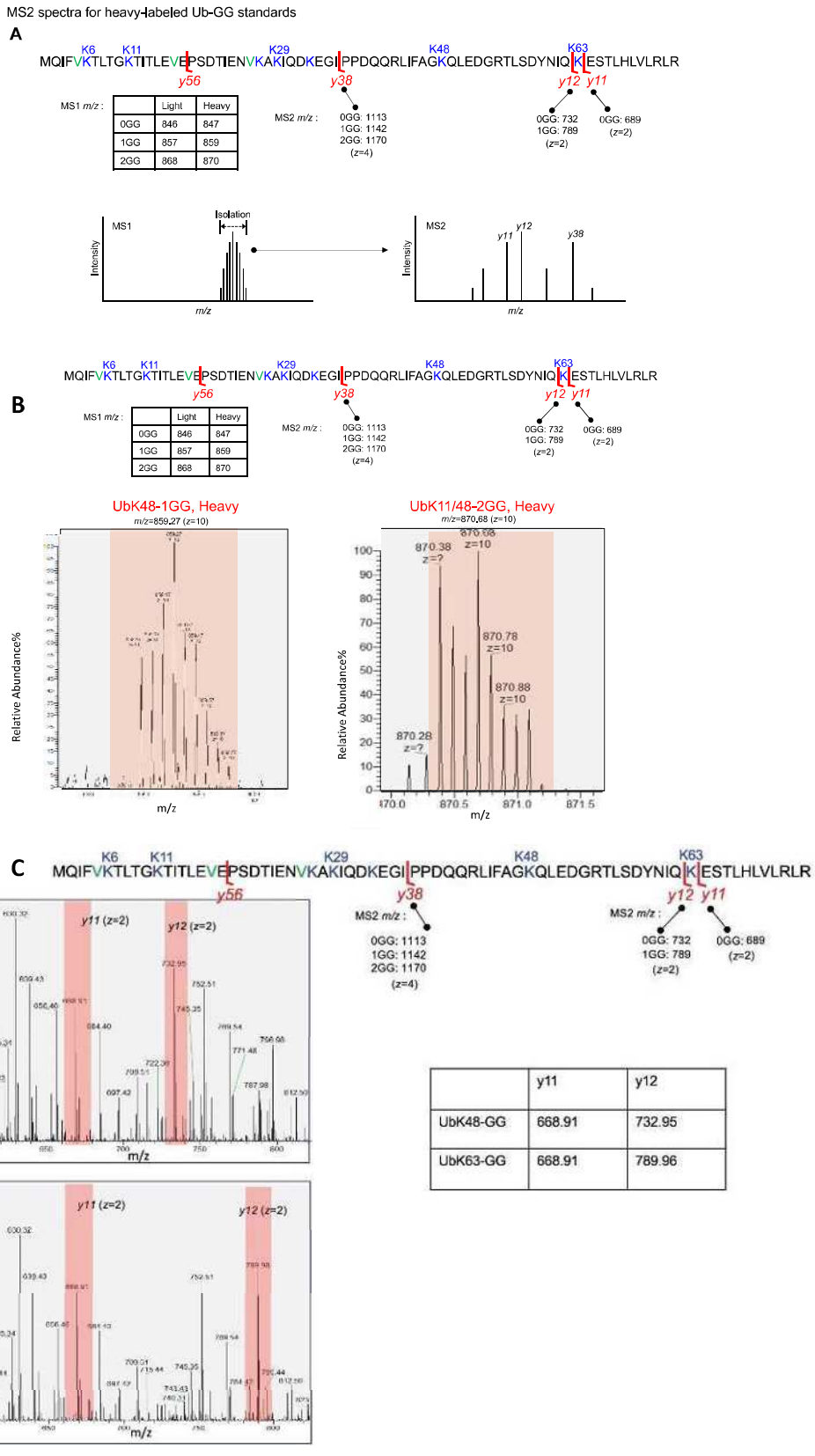
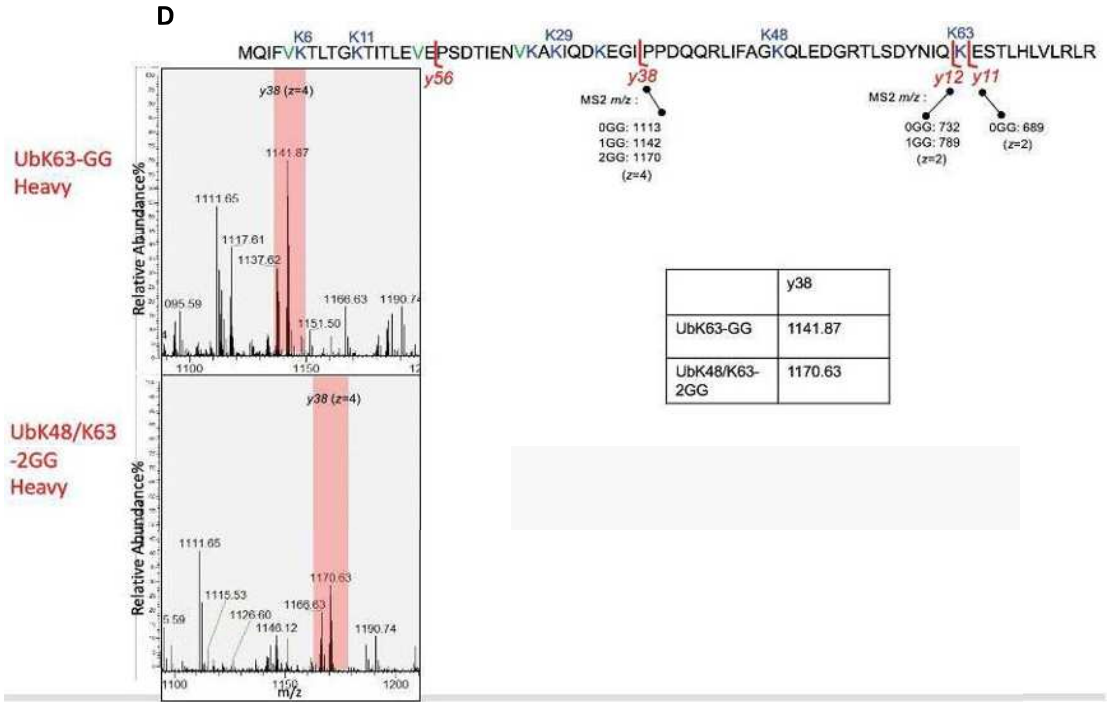


Figure 5. MS spectra of the chemically synthesized derivatives consisting of the "GG-motif" mimicking the trypsin digestion of 74 amino acid long ubiquitin is analyzed by orbitrap.

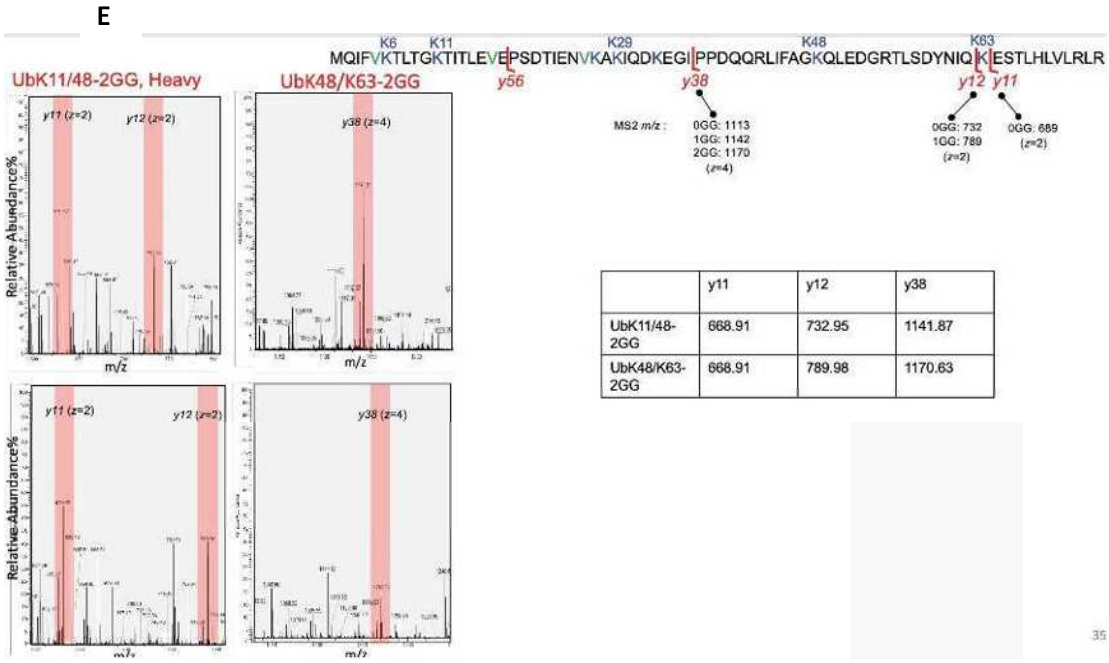
The MS data represents the MS1 spectra of linear and branched (light and heavy labeled) with the m/z=10 value. It is found that the calculated mass difference between Ub(1-74) and Ub(1-74)-GG of 114.04kDa and between Ub(1-74) and Ub(1-74)-2GG of 228.08kDa was observed as calculated. Further, the difference between light labeled and heavy labeled chains of 18kDa was observed. This showed that the chemically synthesized ubiquitin derivatives could be used to conduct the quantification analysis of the ubiquitin chains.

After conducting the MS1 analysis, MS2 analysis has been carried out. The MS2 analysis helped to differentiate between different ubiquitin chains on the basis of different ionization position. On the basis of y11, y12, y48, and y56 ions, the different linear and branched ubiquitin chains where characterized.


MS2 spectra for heavy-labeled Ub-GG standards
A
K6 K11 K29 K48 K63
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLR
y56 y38 y12 y11
MS1 m/z :
Light Heavy
0GG 846 847
1GG 857 859
2GG 868 870
MS2 m/z :
0GG: 1113
1GG: 1142
2GG: 1170
(z=4)
0GG: 732
1GG: 789
(z=2)
0GG: 689
(z=2)
MS1
Isolation
Intensity
m/z
MS2
y11
y12
y38
B
UbK48-1GG, Heavy
m/z=859.27 (z=10)
UbK11/48-2GG, Heavy
m/z=870.68 (z=10)
870.38 z=?
870.78 z=10
870.88 z=10
870.28 z=?
Relative Abundance%
C
UbK48-GG Heavy
UbK63-GG Heavy
y11 (z=2)
y12 (z=2)
m/z
y11 y12
UbK48-GG 668.91 732.95
UbK63-GG 668.91 789.96



	y38
UbK63-GG	1141.87
UbK48/K63-2GG	1170.63



	y11	y12	y38
UbK11/48-2GG	668.91	732.95	1141.87
UbK48/K63-2GG	668.91	789.98	1170.63

Figure 6: MS2 data for heavy labeled chemically synthesized ubiquitin derivatives linear UbK48-GG and UbK63-GG and branched UbK11/K48-2GG and UbK48/K63-2GG.

A. The ionization pattern of the ubiquitin protein. The y11, y12, and y38 ions are focused in this study to differentiate among the linear and branched-chain ubiquitin.

B. MS1 spectrum of UbK48-GG and UbK11/K63-2GG. The spectrum shows the mass shift with the presence of two GG-motif, through which the linear and branched chains can be differentiated.
C. MS2 spectrum shows the ionization pattern of UbK48-GG and UbK63-GG. There is a mass shift at the y12 ions, which shows the presence of GG-motif at the K63 position. Thus, through this MS spectrum, we can clearly differentiate between the linear ubiquitin chains.
D. MS2 spectrum shows the ionization pattern of UbK63-GG and UbK48/63-2GG. There is a mass shift at the y38 ions, which shows the presence of 2GG-motif at the K48 and K63 position. Thus, through this MS spectrum, we can clearly differentiate between the linear and branched ubiquitin chains.
E. MS2 spectrum shows the ionization pattern of UbK11/K48-2GG and UbK48/K63-2GG. There is a mass shift at the y12 ions and y38 ions. This shift enables the identification of the two-branched chains and the branching position. Thus, through this MS spectrum, we can clearly differentiate between the two-branched ubiquitin chains.

The Mass spectrometry data obtained for the chemically synthesized compounds, both linear and branched chains, can be validated from the data obtained in the recent research. Strieter *et al.* conducted the MS analysis for phosphor-Ser65-ubiquitin, which is a key activator for Parkin mediated chain assembly. They obtained a similar MS1 spectrum with the z=11 for the Ub(1-74), Ub(1-74)-GG, Ub(1-74)-2GG. [19]

Conclusion

In summary, we reported the chemical synthesis of the ubiquitin derivatives, representing the different ubiquitin chains. The ubiquitin chains synthesized in this study mimic the homogenous and heterogenous chains found in the cell. Homogenous chains correspond to the ubiquitin chain attachment at a distinct position. Similarly, heterogeneous chains are the ubiquitin chain attachment at the peculiar two lysine position of ubiquitin.

After being isolated from the cells, the ubiquitin chains are digested with trypsin for conducting the quantification study. Trypsin cleaves the bond between the Arg and Lys. After the trypsin digestion, a signature "GG-motif" is left on specific fragments. The signature "GG-motif" corresponds to the C-terminal 75-76 diGly. The presence of one "GG-motif" corresponds to the linear chain, and two "GG-motif" corresponds to the branched/forked chain structure. However, many researchers have used trypsin to study the different ubiquitin chains, but the main obstacle for the analysis was the loss of chain topology and identifications of

different chains. Therefore, for identification of the linkage position of ubiquitin chains, specific antibodies were used, such as M1 and K48-linked antibodies. These antibodies have the specificity for a singular position of the ubiquitin chain. After the antibody treatment, MS analysis further helped in the assessment of ubiquitin chains. Although, these analyses helped in the isolation and detection of ubiquitin chains but could not differentiate between different chain topology as linear and branched chains and quantify them, respectively.

Previous studies have conducted the characterization of ubiquitin chains using trypsin digestion and detected them using the middle down mass spectrometry. Due to the limitation of trypsin as it cleaves the internal Arg and Lys bond leaving the ubiquitin chains into several smaller peptide fragments. Thus, leading to loss of ubiquitin chain information, later limited trypsinosis has been used to avoid the loss of chain topology. Although with the use of limited trypsinosis, the ubiquitin chain topology was not a loss, the process of use of limited trypsinosis was difficult to be carried out because of its complex protocol and efficient handling of the controlled cleavage conditions. Therefore, another group, Komander *et al.,* used Lb pro protein to generate full-length ubiquitin fragments with the signature "GG-motif." In one of their previous studies, they found that the Lb pro protein has the characteristic of cleavage of Arg and Gly bond at the C-terminus of ubiquitin-like modifier ISG15. Using the Lb pro protein, Komander *et al.* were able to generate ubiquitin chains without losing the ubiquitin chains, as previously observed when trypsin was used. However, they could generate ubiquitin with the signature "GG-motif" but could not comprehend between the different branching position.

In this study, we focused on developing a new method to conduct the quantification analysis of the different ubiquitin chains present in the cell by overcoming the hindrance of loss of chain topology and chain branching position. To achieve this, we used the chemical synthesis approach and synthesized different ubiquitin chains having the signature "GG-motif" at the pertinent position. These structures defined ubiquitin chains are used as reference molecules to obtain the mass patterns for middle-down MS analysis towards cell-based assay. The ubiquitin chains are synthesized by first dividing the ubiquitin-protein into N-terminal fragment and C-terminal fragment, followed by the introduction of Fmoc-Lys(Alloc)-OH at the appropriate position such as K48, K63. The alloc group was deprotected using the palladium complex, followed by the manual introduction of the "GG-motif." This method enabled us to achieve the target using chains.

These synthesized ubiquitin chains resemble linear chains; for example, K48 linked chains and branched chains; for example, K48/63 linked chains. These chains endure the signature "GG-motif" at those positions, which signifies the position of trypsin cleavage. For

example, the UbK48-GG exemplifies with a ubiquitin linear chain formation at K48 position and after trypsin digestion leaving a "GG-motif" with confined chain topology.

Similarly, the UbK48/63-2GG imitates to branched ubiquitin chain with the branching position at K48 and K63 positions. The 2-GG represents the branched ubiquitin chain after being cleaved by trypsin, leaving the "GG-motif" attached at K48 and K63. The chemical synthesis method helped us in the incorporation of the signature "GG-motif" at the suggestive positions.

We conducted the MS analysis using the Middle down mass spectrometry approach after the synthesis of the different ubiquitin chains (linear and branched). The analysis showed the successful synthesis of the reference chains. The reference chains correspond to linear chains such as UbK48, UbK63, and branched chains such as UbM1/K63, UbK48/K63, and other chains. Similarly, isotope-labeled linear and branched reference chains are synthesized. Analyzing the MS data of the Ub(1-74), Ub(1-74)-GG and Ub(1-74)-2GG showed the expected mass difference of 114.04Da, respectively (figure 5). Moreover, K48/K63 branched linkages are distinguishable based on MS2 ions. This enables linkage-specific quantification of branched chains. K48/K63 is one of the most studied and important branched-chain, and with our synthesis strategy and analysis, we able to successfully distinguish these chains from other chains (figure 6).

Thus, in this research, we were able to successfully synthesis different ubiquitin chains as linear and branched (light and heavy labeled) using the chemical synthesis method. Further, using the Middle down mass spectrometry helped in the analysis of these chains and standardizing the MS data for individual ubiquitin chain. The chemically synthesized molecules help in overcoming the challenge of previous research by maintaining the chain topology and branching information. Hence, these molecules could be further used for conducting the quantification study and decoding the ubiquitin code.

Material and Methods

General methods and materials. MALDI-TOF mass spectra were recorded with micro flex (BRUKER), using Protein Calibration Standard 2 as an external standard. Reversed-phase HPLC was performed on a 5C18-Protein-R for analysis and purification, with a PU-2080 plus Intelligent HPLC Pump (JASCO) and MD-2018 plus Photodiode Array Detector (JASCO) at 195 to 650 nm. All solvents and reagents were commercially available and used without further purification. Peptides were prepared by using Wang-Peg resin (Watanabe Chemical Industries) for peptides corresponding to the C-terminal of Ubiquitin or Cl-Trt(2-Cl) resin (Watanabe Chemical Industries) for peptides corresponding to the N-terminal of Ubiquitin. The isolated

yields of each peptide were estimated by using the molecular weights of TFA salt at N-terminus amino groups, Arg, Lys, and His positions. Peptide synthesis. All peptides were synthesized using Intavis ResPep SL (Intavis). Amino acids protected by 9-fluorenylmethoxycarbonyl (Fmoc) group were coupled with O-(1H-Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HBTU) as activator and 4-methylmorpholine (NMM) as the base. For the coupling of Fmoc-His(Trt)-OH, Fmoc-Cys(Trt)-OH, DIC (6 equiv.), and HOAt (6 equiv.) were employed to avoid racemization during coupling reaction.

Preparation of hydrazine resin. To prepare the C-terminal hydrazide peptide, Cl-Trt(2-Cl) resin(10 μmol scale) was used. The preparation procedure was followed as described in the previous report (J. Zheng, S. Tang, Y. Qi, Z. Wang, L. Liu, Nat. Prot.2014, 8, 2483-2495). Briefly, the resin was swelled in 50% DCM/DMF for 30 min. After removing the solvent, 5% hydrazine in DMF (400μL) was added to the resin, and the mixture was agitated for 1 h at r.t. Then the solvent was drained and washed by DMF. This operation was conducted again, and the resin was washed by DMF, DCM, and DMF. Next, 5% MeOH/DMF (400μL) was added and stirred for 10 min. After removing the solvent, the resin was washed by DMF, DCM, and DMF. Immediately, to the resin were added Fmoc-amino acids (4 equiv.), HBTU (3.8 equiv.), and DIEA (8 equiv.). The mixture was stirred for 60 min at r.t and then washed by DMF, DCM, and DMF three times respectively. The yield of hydrazidation was calculated by UV absorbance at 301 nm (extinction coefficient: 7800 M-1cm-1) derived from Fmoc group after deprotection with 20% piperidine in DMF. Peptide cleavage from resin. To each resin was added TFA cocktails (TFA 90%, thioanisole 5%, ethanedithiol 3%, and anisole 2%). The mixture was rotated at room temperature for 2 h under argon atmosphere. Then, filtered the mixture to remove resin, and cooled ether 10 ml was added. The tube was vortexed well and centrifuged 5,000 g at room temperature for 1 min. Ether was decanted and these operations were repeated three times. The crude precipitation was dried by speed-vac

Thioesterification after cleavage from resin. Peptide precipitation was dissolved in 6 M Gn·HCl and 0.2 M NaH2PO4 at pH 3.0 (peptide concentration: 2–3 mM). The solution was cooled to -15 °C and 1 M NaNO2 aq was added (10 equiv. against peptide). The mixture was stirred at -15 °C for 15 min, and then 1 M MESNa aq (50 equiv. against each peptide) was added to the reaction mixture. The pH was adjusted to 6.5–7.0 with 6 N NaOH aq and the solution was stirred at room temperature for 30 min. The peptide solution was diluted by a mixture of water/acetonitrile containing 0.1% TFA and purified by HPLC.

Desulfurization. To ubiquitin, peptides were added denaturing buffer (6 M Gn.HCl, 0.2 M NaH2PO4 at pH 7.0), TCEP solution (500 mM dissolved in denaturing buffer), and glutathione

solution (1 M dissolved in denaturing buffer). To the mixture were VA-044 solution (200 mM dissolved in denaturing buffer). Final concentration of the mixture was peptide (0.8 mM), TCEP (300 mM), glutathione (150 mM) and VA-044 (15 mM). The reaction mixture was stirred at 37 °C under argon atmosphere for 24 h. The peptide solution was diluted by a mixture of water/acetonitrile containing 0.1% TFA and purified by HPLC.

Native chemical ligation. The MESNa Thioester peptide was dissolved in an aqueous buffer containing 6M guanidine hydrochloride (Gn·HCl), 200 mM NaH_2PO_4, 100mM MPAA and TCEP pH 6.4 to 6.5 adjusting with 6M NaOH at a final concentration of 2mM. The Cys-peptide was added, and the ligation started. Analytic HPLC and MALDI were used to follow the reaction. When the reaction was completed, the mixture was treated with equal volumes of 5% TCEP and then purified by HPLC. The purified ligation product was confirmed by MALDI.

Alloc deprotection (on the resin). DCM solution containing $Pd(PPh_3)_4$(0.4eq) and $PhSIH_3$(10eq) was added to the resins in the column and shaded with aluminum foil. The reaction was accomplished in 30min, and the deprotection of alloc was confirmed by the Kaiser test.

Di-Gly coupling. Fmoc-Gly-OH (4equiv.), HBTU (3.8equiv.), DIPEA (4equiv.) were mixed in DMF and added to the resin for 30min at 40 °C. The coupling of Fmoc-Gly-OH was checked by Kaiser test. After coupling the first Glycine, the Fmoc was deprotected by adding 20% piperidine for 15minx3. Then the next mixture of Fmoc-Gly-OH (4equiv.), HBTU (3.8equiv.), DIPEA (4equiv.) in DMF was added to resin for 30min at 40 °C. The coupling of Fmoc-Gly-OH was checked by the Kaiser test, followed by the removal of the Fmoc group by adding 20% piperidine.

MS/MS analysis. For shotgun MS analysis, an Easy nLC 1200 system (Thermo Fisher Scientific) was connected online to an Orbitrap Fusion LUMOS mass spectrometer (Thermo Fisher Scientific) with a nanoelectrospray ion source. The Orbitrap Fusion LUMOS instrument was operated in a data-dependent MS/MS mode by using the Xcalibur software (Thermo Fisher Scientific). Data were processed by using Qual browser software (Thermo Fisher Scientific).

Synthesis of Fmocisotope-labeled valine. The isotope-labeled valine was purchased, and the fmoc group was attached following the protocol. The L-amino acid was dissolved in water containing sodium carbonate decahydrate. The solution was cooled in an ice bath, and dioxane was added slowly over 1min with stirring. Towards the end of the addition of a fine white powder began to precipitate. A solution of Fmoc-Cl in dioxane was added dropwise with rapid stirring over 10min. Almost all the solid dissolved toward the end of the addition, but a new solid then began to separate. The mixture was stirred at ice bath temperature for 2hr then at

room temperature for 5.5hr. The solvent was evaporated in vacuo, and the solid remaining dissolved in water. The pH of the solution was about 8.5. It was extracted with ether. The aqueous phase was adjusted to the pH2 with HCL. A white solid precipitated. The mixture was extracted with ethyl acetate, and the combined extracted were washed with saturated brine, then dried over MgSO4, filtered, and the solvent evaporated, leaving a white solid. The solid was stirred with hexane for 60min, the hexane decanted, and the solid dried in vacuo. The solid was dissolved in boiling ethyl acetate, and hexane was added slowly while warming gently. Crystallization began during this time. The mixture was allowed to cool slowly to rt then left at 5degree overnight. The solid was isolated by filtration, washed with hexanes and vacuum dried at rt.

References

1. Darwin, K. H. Prokaryotic ubiquitin-like protein (Pup), proteasomes and pathogenesis. *Nat. Rev. Microbiol.* 7, 485–491 (2009).

2. Geis-Asteggiante, L., Lee, A. E. & Fenselau, C. *Analysis of the topology of ubiquitin chains*. *Methods in Enzymology* vol. 626 (Elsevier Inc., 2019).

3. Goldberg, A. L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–899 (2003).

4. Huang, X. & Dixit, V. M. Drugging the undruggables: exploring the ubiquitin system for drug development. *Cell Res.* **26**, 484–498 (2016).

5. Voorhees, P. M. & Orlowski, R. Z. THE PROTEASOME AND PROTEASOME INHIBITORS IN CANCER THERAPY. *Annu. Rev. Pharmacol. Toxicol.* **46**, 189–213 (2006).

6. Ozawa, T. Analysis of a multiqueue model for an ISDN access interface. *Perform. Eval.* **15**, 65–75 (1992).

7. Rose, A. & Mayor, T. Exploring the rampant expansion of ubiquitin proteomics. *Methods Mol. Biol.* **1844**, 345–362 (2018).

8. Wilkinson, K. D. Ubiquitination and deubiquitination: Targeting of proteins for degradation by the proteasome. *Semin. Cell Dev. Biol.* **11**, 141–148 (2000).

9. Rajalingam, K. & Dikic, I. SnapShot: Expanding the Ubiquitin Code. *Cell* **164**, 1074-1074.e1 (2016).

10. Okatsu, K. *et al.* Phosphorylated ubiquitin chain is the genuine Parkin receptor. *J. Cell Biol.* **209**, 111–128 (2015).

11. Chau, V. *et al.* A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science (80-.).* **243**, 1576 LP – 1583 (1989).

12. Husnjak, K. & Dikic, I. Ubiquitin-Binding Proteins: Decoders of Ubiquitin-Mediated Cellular Functions. *Annu. Rev. Biochem.* **81**, 291–322 (2012).

13. Valkevich, E. M., Sanchez, N. A., Ge, Y. & Strieter, E. R. Middle-Down mass spectrometry enables ch aracterization of branched ubiquitin chains. *Biochemistry* **53**, 4979–4989 (2014).

14. Bezstarosti, K., Wal, L. Van Der & Demmers, J. A. A. Detection of Protein Ubiquitination Sites by Peptide Enrichment and Mass Spectrometry. *J. Vis. Exp.* **157**, 1–9 (2020).

15. Bignell, G. R. *et al.* Identification of the familial cylindromatosis tumour-suppressor gene. *Nat. Genet.* **25**, 160–165 (2000).

16. Sylvestersen, K. B., Young, C. & Nielsen, M. L. Advances in characterizing ubiquitylation sites by mass spectrometry. *Curr. Opin. Chem. Biol.* **17**, 49–58 (2013).

17. Xiao, W. *et al.* Ac-LysargiNase Complements Trypsin for the Identi fi cation of Ubiquitinated Sites **91 (**24), 15890-15898 (2019).

18. Swatek, K. N. *et al.* Insights into ubiquitin chain architecture using Ub-clipping. *Nature* **572**, 533–537 (2019).

19. Deol, K. K., Eyles, S. J. & Strieter, E. R. Quantitative Middle-Down MS Analysis of Parkin-Mediated Ubiquitin Chain Assembly., American Society for Mass Spectrometry **31(**5), 1132-1139 (2020).

Chapter 3

Chemical synthesis of H2AK119 ubiquitin and H3.1K27me3 to understand the effect on nucleosome

Abstract

Histone proteins undergo various post-translational modifications, one of the important and widely studied modifications is ubiquitination of H2A and H2B. H2AK119 ubiquitination leads to gene repression and coordinates with other modifications such as histone methylation. In this study, we are interested in the reconstitution of a complex nucleosome consisting of two modifications, i.e., H2AK119ubiqutination and H3.1K27me3. Here, we are using chemically synthesized H2AK119ubiquitination and H3.1K27me3. Both of these marks act in gene repression. Although the role of these marks is known separately, the cross-talk between these modifications is still a controversy. Therefore, we wish to study the effect of these marks together on the nucleosome. Further, we wish to use this complex to conduct gene regulation study.

Introduction

The nucleosome is composed of an octamer made of four histones wrapped around 147bp DNA. Histones are a proteinaceous substance of ~12kDa and a highly conserved core domain. The N- and C-terminal of histones protrude outside of the nucleosome, providing the binding site for various proteins. These proteins modify these terminals by reading, writing, or erasing different marks[1].

DNA constantly undergoes damage due to the endogenous and exogenous damaging agents, which leads to mutation. To overcome the damage response in cells, several damage response agents are recruited to the site of damage. Histones also play an important role in the recruitment of these damage responsive agents to the damage site and other effector proteins such as readers, writers, and erasers[2]. Histones proteins undergo various post-translational modifications such as methylation, ubiquitination, acetylation to regulate the cellular activities[3]. The modifications regulate processes such as DNA transcription, repair, and overcome damage-induced stress in cells. One of the important post-translational modifications that play an important role is ubiquitination[4]. Histone Ubiquitination, most importantly, H2A and H2B, have been widely studied to play an important role in nucleosome dynamics[5]. These modifications also interact with other post-translational modifications such as H2AK119ubiqutination cross-talk with H3K27tri-methylation and H2BK cross talks with H3K4 and H3K79 tri-methylation[6,7]. These modifications are carried out by a group of a protein complex known as the Polycomb group (PcG). H2A ubiquitination contributes only 5-15% of the total ubiquitination observed in the cell and has been linked to X inactivation[8]. In 2004, in HeLa cells, the enzyme responsible for carrying out H2Aubiquitination was identified as PRC1, which consists of three subunits Ring1, Ring2, and Bmi1[9]. Knockdown and several studies confirmed the importance of the Ring2 domain to carry out the ubiquitination of proteins[9]. Recently, a new E3 ligase carrying out the ubiquitination of H2A was identified as 2A-HUB, which also belongs to the Ring family of protein[10]. Studies have also suggested the interaction between Ring2 and PRC1 and 2A-HUB and histone deacetylase complex N-CoR/HDAC1/2[10]. The ubiquitination of different sites on H2A is carried out by different E3 ligases such as RNF168, RING1B(RNF2), BRCA1/BARD1 carries out modification at H2A K13, K15, K119, K127/129 respectively[11]. Each E3 ligase leads to a different response during DNA damage such as BRCA1/BARD1 promote HR, RNF168 promotes NHEJ, PRC1 promotes transcriptional silencing[11]. In the DNA damage response ubiquitination of H2A by

BRCA1/BARD1 is believed to promote HR, and ubiquitination by RNF168 seems to promote NHEJ[12]. EZH2 auxiliary domain of PRC2 complex is responsible for methylation on the H3K27 position. H3K27 mark is associated with gene repression, such as the inactive X chromosome (Xi)[13].

Polycomb Group Proteins (PcG) play an essential role in the maintenance of cellular homeostasis, transcriptional state, and regulation in different human diseases[14,15]. The major gene repression signals in the heterochromatin region are mostly controlled by these PcG[15]. The PcG was first identified in *Drosophila melanogaster* and played an important role in the maintenance of Spatio-temporal repression during the development[16]. In mammals, they play a role in maintaining repression on CpG island[15]. The group can be characterized into two complexes PRC1 and PRC2 on the basis of auxiliary domains. PRC1 consists of the RING1A/B domain, thus having the E3 ligase activity and controls H2AK119 ubiquitin ation[17]. On the other hand, PRC2 possesses the methyltransferases domain EZH1 and EZH2 and interacts with the scaffold protein SUZ12 and EED and controls the methylation of H3.1K27[18]. PRC1 and PRC2 can be further classified into different classes on the basis of the presence of different subunits and thus leading to different function[19]. PRC2 can be categorized into PRC2.1, has PHF1, MTF2, and PHF19 and recognizes the unmethylated CpG island[18]. PRC2.2 has JARD1(binds to H2K119ubiquitin) and AEBP2[17]. Similarly, PRC1 can be categorized into six complexes. Among these PRC1.2, PRC1.4 consists of CBX complex needed to bind H3.1K27[20]. The recruitment action of PRC1 and PRC2 to the target position is not clear until now. PRC1 signals for recruitment of PRC2 complex or vice versa is still a question of controversy[15,17,21]. Studies suggest that H2AK119 ubiquitin acts as the glue that keeps the PRC1 and PRC2 complex together and keeps the gene repression state[19]. Loss of H2AK119 ubiquitin leads to loss of PRC2 (PRC2.1 and PRC2.2) complex affecting the H3.1K27me3 status and loss of PRC1 complex and affecting the repression state of the gene[22]. This cross-talk has an effect on the development of certain human cancer such as Uveal melanoma, mesothelioma[15]. Studies have suggested that loss of H2K119ub and H3K27me3 led to an increase in HOX gene depression, inhibition of cell proliferation[23].

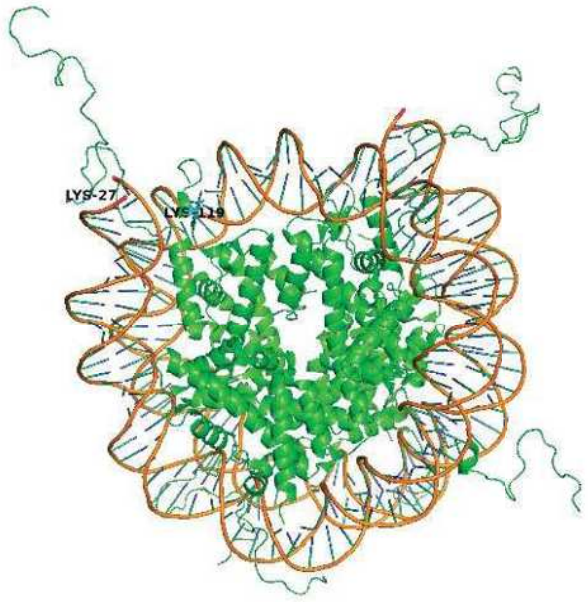


Figure 1. The structure of the nucleosome is shown with the target H2AK119 and H3.1K27 position shown as sticks, PDB 1KX5.

H2AK119 ubiquitination being the most abundant, but it is less understood modification. The discovery of E3 ligase made the H2AK119 ubiquitination research easy. In one of the studies by Nakagawa *et al.,* they interpreted the role of H2A and H2B ubiquitination after reconstitution[24]. They found that there is a negative cross-talk between H3K4 methylation by MLL3 and H2A ubiquitination[24]. Thus, signifying the role of H2AK119 ubiquitination in transcriptional repression and its cross-talk with other modifications[24]. Studies have suggested that H2AK119 ubiquitin has slight inhibition on the PRC2 activity[25]. However, the exact reason for such inhibition is not known. One of the reasons could be the steric hindrance of H2AK119 ubiquitin on the PRC2 binding site. However, the exact mechanism is not yet clear[26,27]. The H2Aub is found at the repressed gene promoter region and heterochromatin[24]. Studies have suggested the role of H2AK119 ubiquitin in transcription initiation on its interaction with H3K4me2/3[24]. Therefore, in this study, we aim to understand how H2AK119 cross-talk with PRC2 and understand its role in the PcG mediated gene silencing. We will reconstitute the nucleosome to mimic the dynamic of histone ubiquitination *in vivo* and investigate the kinetic control of H2AK119 ubiquitin turnover to gene regulation[6]. Further, use the synthesized nucleosome to create a euchromatin and heterochromatin structure. Studies have suggested that the PRC2 activity is ~10fold for heterochromatin than euchromatin[28]. We can also study the effect of H1 linker histone on the heterochromatin and euchromatin can be studied.

The synthesis of monoubiquitinated histone can be achieved using the semisynthesis method or chemical synthesis methods. Earlier disulfide chemistry was used to achieve the ubiquitination on histone H2AK119 and H2BK120. It provides the advantage of no requirement of special chemical synthesis and allows ubiquitination at the desired lysine by mutating the Lys to Cys[29]. Fierz and colleagues used the EPL reaction to achieve the target ubiquitinated histone. EPL reaction involves the reaction between UbG76C, C-terminal thioester, and chemically synthesized histone fragments. The drawback of the EPL method was that after the ligation in the desulfurization step, the C-terminal glycine at 76 positions of ubiquitin is mutated to alanine[30]. Later Liu grouped synthesized H2Aub derivatives using the TFA Labile auxiliary[31]. In this study, we are adopting the methodology of the Liu group for the chemical synthesis of H2AK119ub.

Results and discussion

Human histone H2A, which consists of 129 amino acids, does not contain Cys residues, which are both indispensable for NCL reaction and suitable for use as peptide cleavage sites. However, desulfurization reactions, in which Cys residues are transformed into Ala residues through radical reactions, have been developed to enable the design of peptide fragments at Ala residues. We envisaged dividing the H2A sequence into three peptide fragments at A47 and A86 (Scheme). 2-mercaptoethanesulfonate was used as a C-terminal thioester precursor for the N-terminal fragment (1), the middle fragment (2) remained in the hydrazine form. The N-terminal Cys residues of fragments 2 and 3 were protected by the Alloc group to prevent a self-ligation reaction. Each fragment was synthesized through Fmoc SPPS, purified by HPLC, and identified by MALDI-TOF MS figure 2. The isolated yield of **1,2,** and **3** were 12%, 10%, and 14%, respectively. To the plasmid of recombinant ubiquitin first, the mutation was done at the G76C position, which enables the thiol group incorporation at the C-terminal. After the successful incorporation of the Cys, the recombinant ubiquitin was expressed in Rosetta competent cells and purified using AKTA. Later the recombinant ubiquitin was converted to its hydrazine (4) form using the Macmillan method. The reaction takes 78hrs for maximum conversion yield of 74.6% and an isolated yield of 40% in figure 3. After purification of the hydrazine form, it is converted to 2-mercaptoethanesulfonate thioester form by using the standard protocol. The isolated yield of UbMESNa thioester (4*) was 18.11% in figure 4.

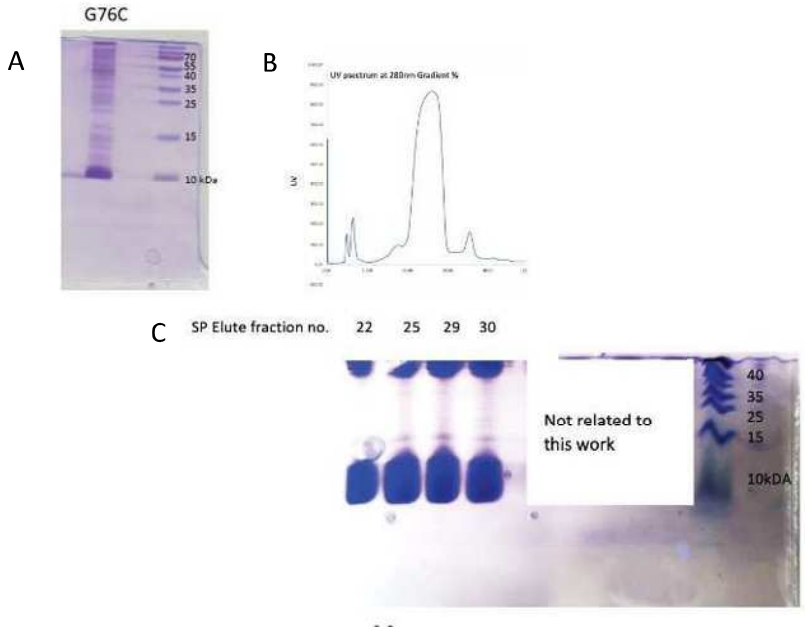


Figure 1. Recombinant ubiquitin-protein expression check (A) and purification (B) using cation exchange column and purity check of the eluted fractions (C)

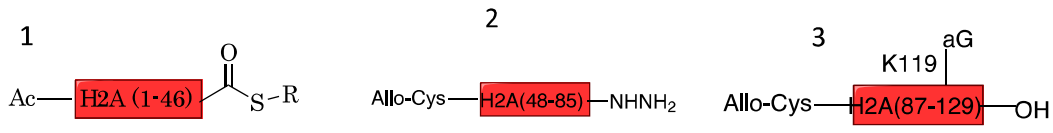
1 47

SGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLTAEI

86

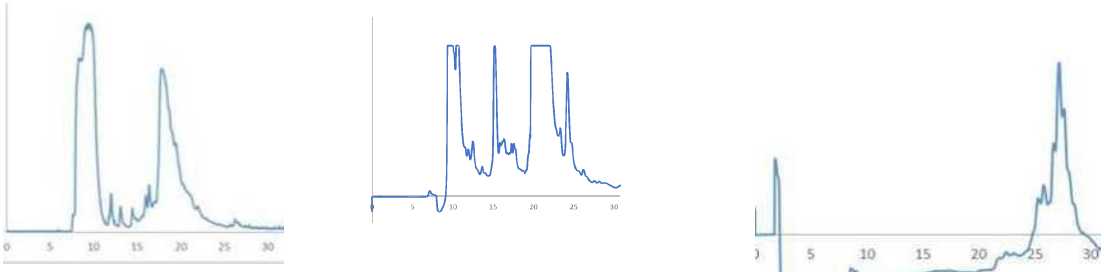
LELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK

SGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERVGAGCPVYLAAVLEYLTAEILE
LAGNAARDNKKTRIIPRHLQLCIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK



Crude HPLC of each fragment of H2A

A fragment1 B fragment 2 C fragment 3



Purity check of each fragment respectively

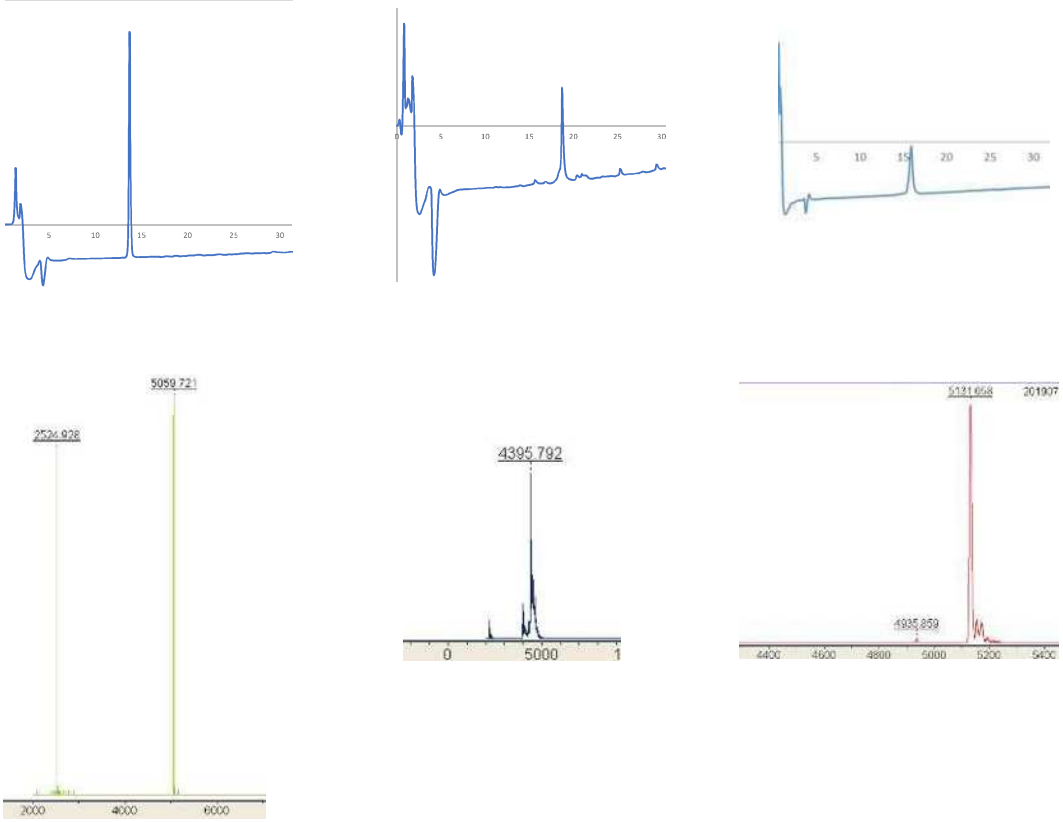


Figure 2. Synthesis of peptide fragments 1 (a), 2(b), and 3(c). HPLC charts of crude and purified peptide fragments were shown in blue. Peaks indicated by black arrows correspond to the desired peptides. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1%TFA. Gradient: (a) 20-40%, 30min. (b) 40-60%, 30min. (c)

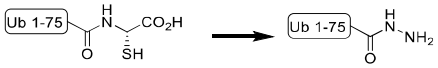
25-45%, 30min. MALDI-TOF MS of each peptide were shown: peptide 1, observed mass 5059.72 Da(calcd. 5058.25), peptide 2, observed 4395.79 Da(calcd. 4393.14), peptide 3, observed 5131.63 Da(calcd. 5129.56)

Ubiquitin peptide sequence

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLH LVLRLRGC

4

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLH LVLRLRGNHNH2



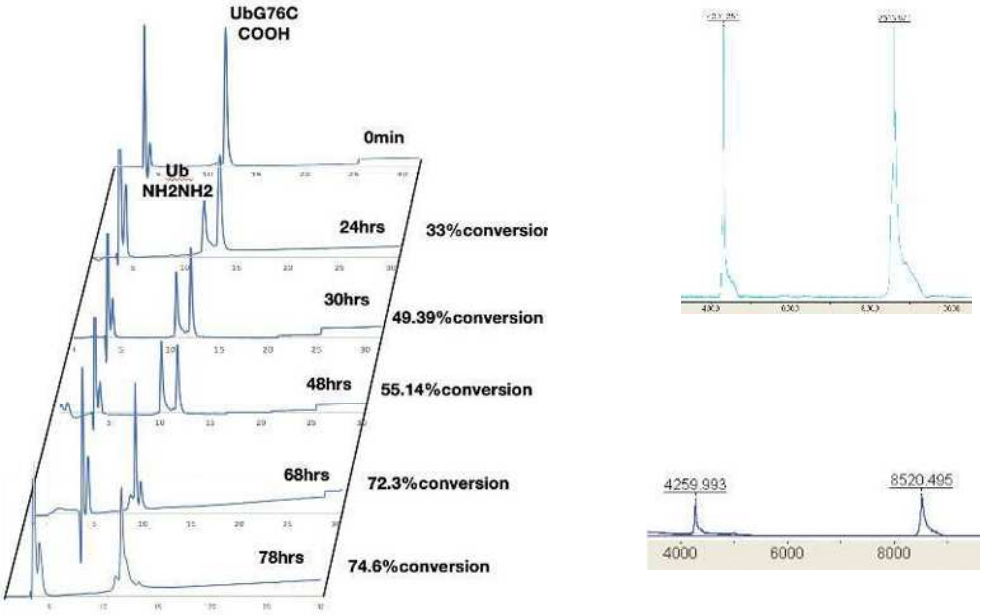


Figure 3. HPLC charts for the conversion of UbG76C to UbNHNH2 using the Macmillan method. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1% TFA gradient 20-60%, 30min. MALDI-TOF MS of each spectrum was shown. UbG76C, $[Peptide+H]^+$ observed mass 8613.87 (calcd 8611.83), UbNHNH2, $[Peptide+H]^+$ observed mass 8520.49 (calcd 8521.72)

4*

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRG
MESNa

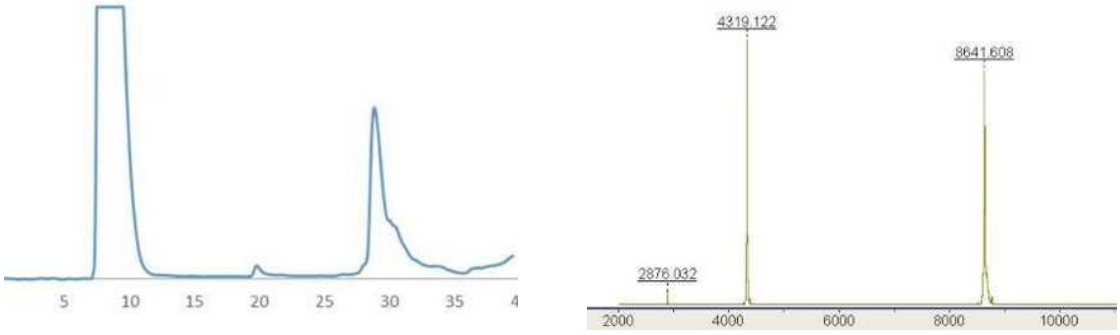
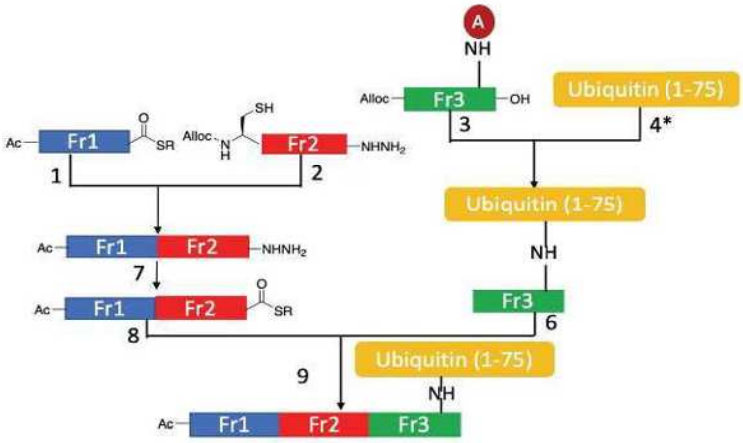


Figure 4. HPLC charts for the conversion of UbNHNH2 to UbMESNa to achieve the peptide 4*. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1% TFA gradient 20-60%, 30min UbMESNa [Peptide+H] $^{+}$, observed mass 8641.60 (calcd 8639.81).

Fragments **4*** and **3** were first assembled to generate the ubiquitin attached fragment **5** in the presence of 4-mercaptophenylacetic acid (MPAA), which accelerates the NCL, the reaction was monitored by HPLC, and the target mass was checked by MALDI-TOF MS. After the completion of the reaction, one-pot alloc deprotections were carried out using the Pd/TPPTS complex. The alloc deprotection finishes in 30min. The proceeding of reaction is monitored by HPLC, and the mass is checked by MALDI-TOF MS, followed by HPLC purification to obtain fragment **6** figure 5. A second ligation was conducted between peptides 1 and 2 to generate 85-aa peptide **7** in the presence of 4-mercaptophenylacetic acid (MPAA). The progress of the reaction is monitored by HPLC, with an isolated yield of 15.6% in figure 6. The peptide 7 is then converted to its MESNa thioester **8** for and purified with an isolated yield of 13.2%. The final reaction is conducted between peptide **8** and peptide **6** in the presence of 4-mercaptophenylacetic acid, the progress of the reaction was monitored by HPLC, and the target mass was confirmed by MALDI-TOF MS to achieve the target product **9** with the isolated yield of 20% figure7. After purification, the auxiliary group is removed under acidic condition using TFA:TIPS: H2O as the TFA cocktail, for 2hrs and later precipitated by **ether.** The overall reaction scheme is shown below in scheme1.



Scheme 1. The synthetic strategy used for the synthesis of ubiquitinated histone H2A.

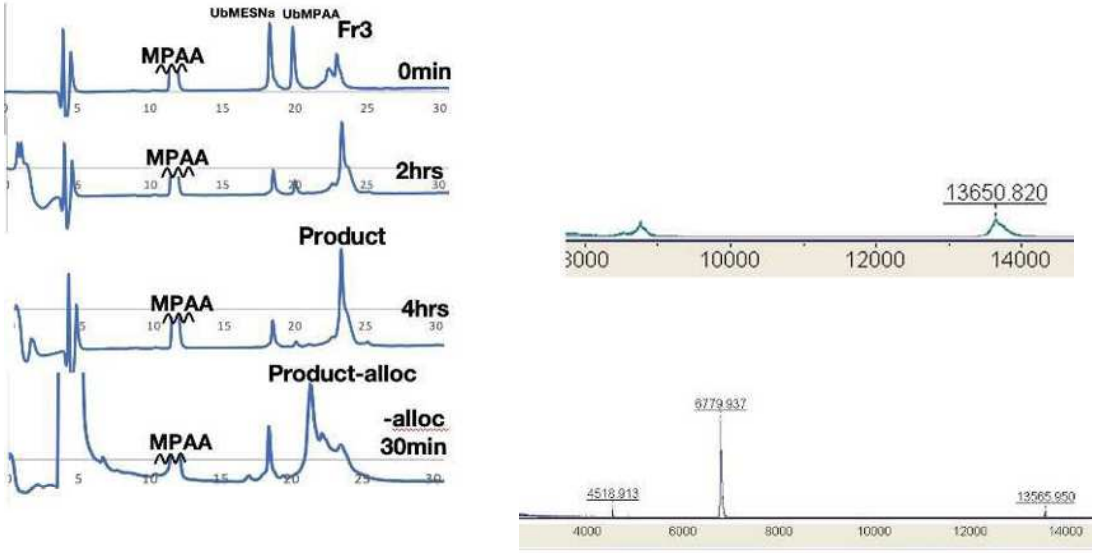


Figure 5. HPLC charts for the 1st NCL of peptide 4* and peptide 3 of H2A to achieve peptide 6. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1% TFA gradient 25-48%, 30min. MALDI-TOF MS of each spectrum was shown. UbFr3(+alloc)[Peptide+H]$^+$, observed mass 13650.82 (calcd 13647.4), UbFr3(-alloc)[Peptide+H]$^+$, observed mass 13565.95 (calcd 13564.37)

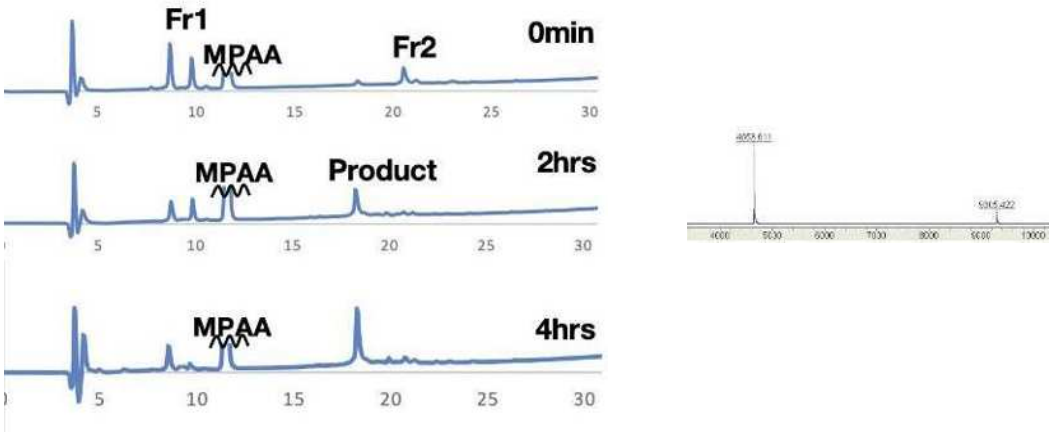


Figure 6. HPLC charts for the 2nd NCL of peptide 1 and peptide 2 of H2A to achieve peptide 7. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1% TFA gradient 20-80%, 30min. MALDI-TOF MS of each spectrum was shown. Fr1Fr2[Peptide+H]$^{+}$, observed mass 9305.42 (calcd 9307.73)

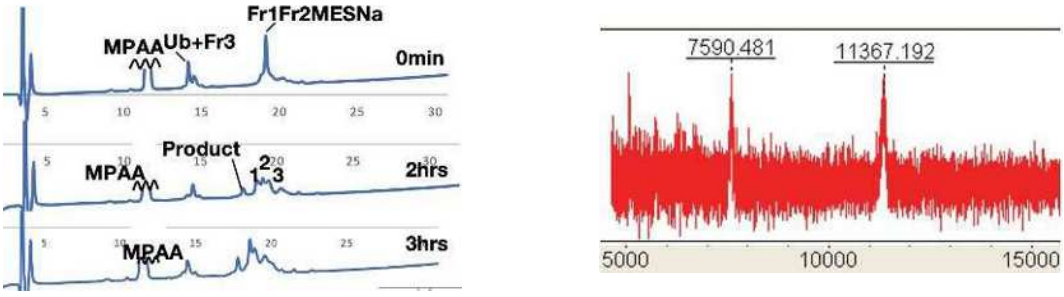


Figure 7. HPLC charts for the 2nd NCL of peptide 4* and peptide 3 of H2A to achieve peptide 6. HPLC peaks were monitored at 220nm in the linear gradient with water/acetonitrile containing 0.1% TFA gradient 20-80% 30min. MALDI-TOF MS of each spectrum was shown. Fr1Fr2Fr3Ub observed mass 11367.19 (calcd 11405.56).

After the synthesis of H2AK119ubiquitin structural analysis was carried out to understand the effect of ubiquitin adheres to H2A. To compare the data H2A and full-length Ubiquitin was taken as the control (figure 8)

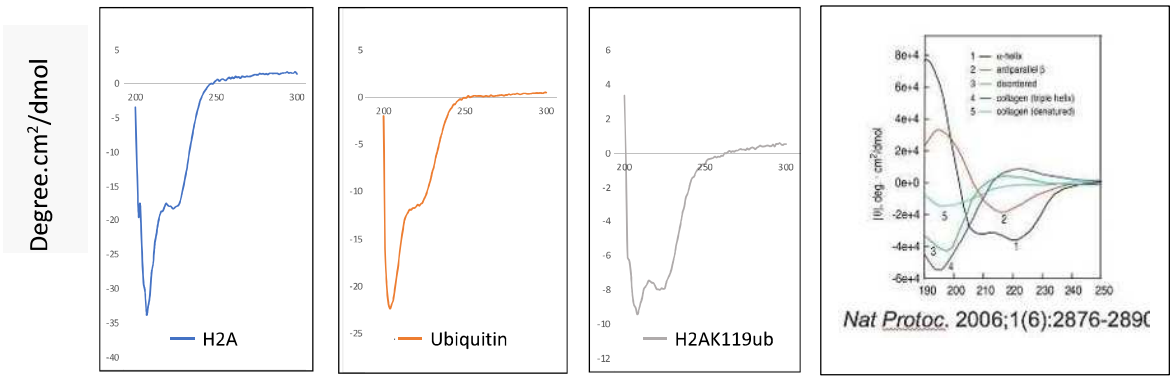


Figure 8. CD spectrum of H2A, Ubiquitin, and H2AK119ubiquitin.

The secondary structure analysis shows that the ubiquitin has a sharp dip at 208nm and 222nm, which corresponds to alpha-helix and 219nm dip for beta-sheets, similarly, H2A shows a strong peak at 208nm and 222nm which corresponds to alpha-helix and 219nm which corresponds to beta-sheets. As we want to understand the change in the structure when ubiquitin is attached to H2A, we checked the structure of H2AK119ubiquitin and founds that the H2AK119ubiquitin has stronger peak dip at 208nm and 222nm which corresponds to alpha-helix and 219nm dip for beta sheets.

Conclusion

H2AK119ubiquitin is one of the most studied histone ubiquitin molecules. H2AK119ubiquitin is a gene repression mark and plays an important role in regulating polycomb silencing. Similar to H2AK119ubiquitin, another histone mark H3.1K27me3 is also a gene repression mark. These marks are well studied separately and identified as gene repression marks. In this study, we aim to reconstitute nucleosome consisting of two gene repression mark. We would like to understand the changes in nucleosome stability because of the presence of two gene repression marks by conducting thermal stability assay.

To conduct this analysis, we successfully synthesized H2AK119ubiquitin using the native chemical ligation method used by the Lei Liu group. The recombinant ubiquitin was expressed and purified with the desired G76C mutation. Further, the glycine auxiliary was synthesized and introduced manually at the K119 position of H2A ubiquitin. After the successful synthesis of the H2A fragments and obtaining ubiquitin in the mesna thioester form, native chemical ligation was conducted. The method of ligation between the fragments was optimized by changing the temperature and pH of the reaction. After incorporation of the

desired modification to the C-terminal of H2A, the protecting groups were deprotected, and the full-length H2AK119 ubiquitin was obtained.

After obtaining a large amount of the product, the CD analysis was conducted. In this analysis, unmodified H2A and ubiquitin were used as control molecules. We observed that H2AK119ubiquitin has a merged structure of H2A and ubiquitin as expected. We also found that the H2AK119 ubiquitin has some turns, which could be because of the linker region between the H2A and ubiquitin. Recombinant H2AK119ubiquitin should be used to confirm the observed structural ambiguity observed for the synthesized H2AK119ubiquitin.

We will further use the synthesized H2AK119ubiquitin to reconstitute the nucleosome and conduct the thermal stability assay.

Material and methods

Ubiquitin was expressed in Rosetta 2(DE3) E. coli cells in LB media supplemented with appropriate kanamycin at 37ºC to OD600 ~0.6-0.8 and transferred to 20 ºC for overnight after induction with 0.3mM IPTG. Cultures were harvested, flash-frozen, resuspended in lysis buffer (50 mM Ammonium acetate pH 4.51, 200mM NaCl, 1mM EDTA, 5mM Mercaptoethanol, 1mM PMSF), lysed by sonication, and clarified at 60,00xg for 30 min at 4ºC. Clarified lysate was then dialysed in SP buffer A (50mM Ammonium acetate pH4.5, 0.1mM EDTA, 50mM NaCl, 5mM Mercaptoethanol). After dialysis, the protein was filtered and purified using the cationic exchange column. The purified fractions were dialyzed in Ub storage buffer (50mM Tris HCl pH 7.5 and 50mM NaCl). The purity of the fractions was checked by SDS PAGE.
Ub1-75 hydrazide (Ub75-NHNH2) was generated using the Macmillan method. In brief, the native protein (like WT Ub and Ub Cys mutants) of which amino acid of the C terminal is Gly-Cys could undergo N-S acyl transfer, and hydrazine could be a suitable nucleophile, leading a reliable C Terminal hydrazinolysis of proteins. 20 mg/mL Ub Cysn-G76C, 5 mg/mL TCEP, 50 mg/mL NHNH2•HCl, 100mg/ml MesNa were mixed in 20 mM S4 Tris, pH 6.5, and the reaction was rotated at 60 rpm, 50 °C for 5 days. The final products were purified by RP-HPLC.
MALDI-TOF mass spectra were recorded with micro flex (BRUKER), using Protein Calibration Standard 2 as an external standard. Reversed-phase HPLC was performed on a 5C18-Protein-R for analysis and purification, with a PU-2080 plus Intelligent HPLC Pump (JASCO) and MD-2018 plus Photodiode Array Detector (JASCO) at 195 to 650 nm. All solvents and reagents were commercially available and used without further purification.
Peptides preparation. Peptides were prepared by using Wang-Peg resin (Watanabe Chemical Industries) for peptides corresponding to the C-terminal of Ubiquitin or Cl-Trt(2-Cl) resin

(Watanabe Chemical Industries) for peptides corresponding to the N-terminal of Ubiquitin. The isolated yields of each peptide were estimated by using the molecular weights of TFA salt at N-terminus amino groups, Arg, Lys, and His positions. Peptide synthesis. All peptides were synthesized using Intavis ResPep SL (Intavis). Amino acids protected by 9-fluorenylmethoxycarbonyl (Fmoc) group were coupled with O-(1H-Benzotriazol-1-yl)-N, N, N', N'-tetramethyluronium hexafluorophosphate (HBTU) as activator and 4-methylmorpholine (NMM) as the base. For the coupling of Fmoc-His(Trt)-OH, Fmoc-Cys(Trt)-OH, DIC (6 equiv.), and HOAt (6 equiv.) were employed to avoid racemization during coupling reaction.

Preparation of hydrazine resin. To prepare the C-terminal hydrazide peptide, Cl-Trt(2-Cl) resin (10μmol scale) was used. The preparation procedure was followed as described in the previous report (J. Zheng, S. Tang, Y. Qi, Z. Wang, L. Liu, Nat. Prot.2014, 8, 2483-2495). Briefly, the resin was swelled in 50% DCM/DMF for 30 min. After removing the solvent, 5% hydrazine in DMF (400μL) was added to the resin, and the mixture was agitated for 1 h at r.t. Then the solvent was drained and washed by DMF. This operation was conducted again, and the resin was washed by DMF, DCM, and DMF. Next, 5% MeOH/DMF (400μL) was added and stirred for 10 min. After removing the solvent, the resin was washed by DMF, DCM, and DMF. Immediately, to the resin were added Fmoc-amino acids (4 equiv.), HBTU (3.8 equiv.), and DIEA (8 equiv.). The mixture was stirred for 60 min at r.t and then washed by DMF, DCM, and DMF three times, respectively. The yield of hydrazidation was calculated by UV absorbance at 301 nm (extinction coefficient: 7800 M-1cm-1) derived from the Fmoc group after deprotection with 20% piperidine in DMF. Peptide cleavage from resin. To each resin was added TFA cocktails (TFA 90%, thioanisole 5%, ethanedithiol 3%, and anisole 2%). The mixture was rotated at room temperature for 2 h under argon atmosphere. Then, filtered the mixture to remove resin, and cooled ether 10 ml was added. The tube was vortexed well and centrifuged 5,000 g at room temperature for 1 min. Ether was decanted, and these operations were repeated three times. The crude precipitation was dried by speed-vac

Thioesterification after cleavage from resin. Peptide precipitation was dissolved in 6 M Gn·HCl and 0.2 M NaH2PO4 at pH 3.0 (peptide concentration: 2–3 mM). The solution was cooled to -15 °C and 1 M NaNO2 aq was added (10 equiv. against peptide). The mixture was stirred at -15 °C for 15 min, and then 1 M MESNa aq (50 equiv. against each peptide) was added to the reaction mixture. The pH was adjusted to 6.5–7.0 with 6 N NaOH aq, and the solution was stirred at room temperature for 30 min. The peptide solution was diluted by a mixture of water/acetonitrile containing 0.1% TFA and purified by HPLC.

Native chemical ligation

Peptide fragment ligations were carried out as previously reported. Briefly, the MESNa Thioester peptide was dissolved in an aqueous buffer containing 6M guanidine hydrochloride (Gn·HCl), 200 mM NaH_2PO_4, 100mM MPAA and TCEP pH 6.4 adjusting with 6M NaOH) at a final concentration of 2mM. The Cys-peptide was added, and the ligation started. Analytic HPLC and MALDI were used to follow the reaction. When the reaction was completed, the mixture was treated with equal volumes of 5% TCEP and then purified by HPLC. The purified ligation product was confirmed by MALDI.

Alloc deprotection. The alloc deprotection can be carried out by using 3eq of Pd/TPPTS complex at room temperature for 30min.

CD Measurement. H2A and H2AK119 ubiquitin were dissolved in PBS buffer, and ubiquitin was dissolved in the storage buffer to a final concentration of 0.005mM. CD measurement was performed and analyzed with the circular dichroism spectrometer J-82O (JASCO).

References:

1. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
2. Williamson, E. A., Wray, J. W., Bansal, P. & Hromas, R. Overview for the histone codes for DNA repair. *Prog. Mol. Biol. Transl. Sci.* **110**, 207–227 (2012).
3. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
4. Meas, R. & Mao, P. Histone ubiquitylation and its roles in transcription and DNA damage response. *DNA Repair (Amst).* **36**, 36–42 (2015).
5. Meas, R. & Mao, P. Histone ubiquitylation and its roles in transcription and DNA damage response. *DNA Repair (Amst).* **36**, 36–42 (2015).
6. Whitcomb, S. J. *et al.* Histone monoubiquitylation position determines specificity and direction of enzymatic cross-talk with histone methyltransferases Dot1L and PRC2. *J. Biol. Chem.* **287**, 23718–23725 (2012).
7. Shukla, A., Chaurasia, P. & Bhaumik, S. R. Histone methylation and ubiquitination with their cross-talk and roles in gene expression and stability. *Cell. Mol. Life Sci.* **66**, 1419–1433 (2009).
8. Vissers, J. H. A., Nicassio, F., van Lohuizen, M., Di Fiore, P. P. & Citterio, E. The many faces of ubiquitinated histone H2A: Insights from the DUBs. *Cell Div.* **3**, 1–14 (2008).

9. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
10. Zhou, W. *et al.* Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. *Mol. Cell* **29**, 69–80 (2008).
11. Uckelmann, M. & Sixma, T. K. Histone ubiquitination in the DNA damage response. *DNA Repair (Amst).* **56**, 92–101 (2017).
12. Simon, J. A. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49**, 808–824 (2013).
13. Rougeulle, C. *et al.* Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol. Cell. Biol.* **24**, 5475–5484 (2004).
14. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science (80-.).* **293**, 1074 LP – 1080 (2001).
15. Tamburri, S. *et al.* Histone H2AK119 Mono-Ubiquitination is Essential for Polycomb-Mediated Transcriptional Repression. *bioRxiv* 690461 (2019)
16. Pengelly, A. R., Kalb, R., Finkl, K. & Müller, J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev.* **29**, 1487–1492 (2015).
17. Kasinath, V., Beck, C., Sauer, P., Poepsel, S. & Kosmatka, J. JARID2 and AEBP2 regulate PRC2 activity in the presence of H2A ubiquitination or other histone modifications. 1–20 (2020).
18. van Mierlo, G., Veenstra, G. J. C., Vermeulen, M. & Marks, H. The Complexity of PRC2 Subcomplexes. *Trends Cell Biol.* **29**, 660–671 (2019).
19. Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445–1459 (2014).
20. Cao, R. *et al.* Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science (80-.).* **298**, 1039 LP – 1043 (2002).
21. Cooper, S. *et al.* Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat. Commun.* **7**, 13661 (2016).
22. Cao, R., Tsukada, Y. I. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and hox gene silencing. *Mol. Cell* **20**, 845–854 (2005).
23. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
24. Nakagawa, T. *et al.* Deubiquitylation of histone H2A activates transcriptional initiation via. *Genes Dev.* 37–49 (2008)

25. M. Christopher, A. M. L. S. Recruiting Polycomb to chromatin. *Physiol. Behav.* **176**, 100–106 (2016).
26. Richly, H. *et al.* Transcriptional activation of polycomb-repressed genes by ZRF1. *Nature* **468**, 1124–1128 (2010).
27. Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
28. Yuan, W. *et al.* Dense chromatin activates polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science (80-.).* **337**, 971–975 (2012).
29. Chatterjee, C., McGinty, R. K., Fierz, B. & Muir, T. W. Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nat. Chem. Biol.* **6**, 267–269 (2010).
30. Obradovic, J. Stability of nucleosomes containing homogenously ubiquitylated H2A and H2B prepared using semisynthesis Beat. *Bone* **23**, 1–7 (2005).
31. Ai, H. *et al.* Examination of the Deubiquitylation Site Selectivity of USP51 by Using Chemically Synthesized Ubiquitylated Histones. *ChemBioChem* (2018).

Chapter 4

Conclusions and Future perspectives

The author, throughout this thesis, demonstrated the chemical synthesis of ubiquitin derivatives. These ubiquitin derivatives have an important role in understanding the ubiquitin codon. This chapter briefly summarizes the work and describes future perspectives on this study.

In chapter1, the author shortly reviewed previous and current progress of ubiquitination from biological and chemical viewpoints. The author mentioned the various methods earlier used to conduct quantification studies. Here, the different strategy used for the chemical synthesis of protein is also mentioned. This chapter introduced that decoding the ubiquitin code can be achieved using the chemical methods, which is an attractive and competitive research field.

In chapter 2, the author achieved the total chemical synthesis of ubiquitin chains, the bedrock of this thesis. The chains synthesized for the study imitates the homogenous and heterogenous ubiquitin chains. Further, the Middle down the mass spectrum of them was generated, which confirmed the specificity of the chains. The MS1 and MS2 data shows the differentiation pattern of the linear and branched chains. Using the data, the chemically synthesized chains such as UbK48-GG, UbK63-GG can be differentiated. These reference models mimic the isolated ubiquitin chains and retain their chain topology. Further, the author wants to conduct the quantification study of the ubiquitin chains isolated chains from the cell. For doing so, the ubiquitin chains after isolated would be treated with the Lb Pro enzyme, which leaves the signature "GG-motif" on the unique position. The advantage of this enzyme is that the chains retain their topology. After isolation, the MS data would be generated, using the same reaction conditions as that of the reference model, shown in the figure below:

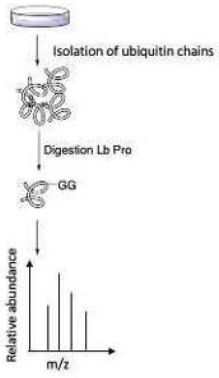


Figure 1. Schematic representation of future experiments to be done for conducting the

quantification study.

The reference models were validated by comparing the MS1 spectrum of previous research. The comparison showed that the chemically synthesized compounds could be used for conducting further analysis.

In chapter3, the author described the chemical synthesis of H2AK119 ubiquitin, which plays an important role in genome regulation. The author wants to use this synthesized compound to conduct a crosstalk study with H3.1K27me3 and understand the genome compaction study.

A large number of studies about the chemical synthesis of ubiquitin and ubiquitinated histones are paving the way for the understanding of epigenetics at the molecular level. Artificial preparation of modified histones has led to complicated and elaborated biochemical assays in full-length histone, mono-nucleosome, and poly-nucleosome context. Muir and co-workers are pioneering researchers in the field of chromatin chemical biology. They have designed ubiquitinated histones and demonstrated in vitro elegant assays. PTM research in the nucleosome level is treasure-trove because higher-order structures such as acidic patch function as a scaffold of various factors, whereas conventional peptide-based approach is difficult to consider the effect of these 3D structures. While semisynthetic way and site-selective labeling of PTMs are prevalent in this area, fully synthetic ways also can address the issues by which multiple molecules, including PTMs, photoaffinity tags, and fluorescent dyes, are introduced. The author is also focusing on the structural effect on the nucleosome due to the presence of a gene repression mark. Therefore, the author conducts a CD spectrum analysis. Further, nucleosome formation and thermal stability assay would be carried out.

This is the first study of synthesizing two histone gene repression marks and reconstituting nucleosome. After the nucleosome reconstitution and analyzing the effect of thermal stability assay heterochromatin formation will be carried out. The heterochromatin chains would help us to study the role of H2AK119 ubiquitin in polycomb targeting gene silencing and Hox gene repression[1]. The heterochromatin chain formation will be important to study the effect of compact formed nucleosome because of two gene repression marks[2]. Studies have suggested the effect of distancing between the nucleosome and the signal carried out by the post translational modification[3].

Studies have suggested the role of H2AK119 ubiquitin in double-strand DNA repair (DDR). Thus, our model having ubiquitination and methylation can be used to study the complex process of DDR[4].

Further, H3.1K27me3 has a binding site for the Polycomb protein (Pc), which is a part

of the PRC1 complex. PRC1 complex has an E3 ligase for ubiquitination[5]. Thus, our complex having the H2AK119 ubiquitin and H3.1K27me3 will have an important model for studying the Polycomb complex and signaling between PRC1 and PRC2[6,1].

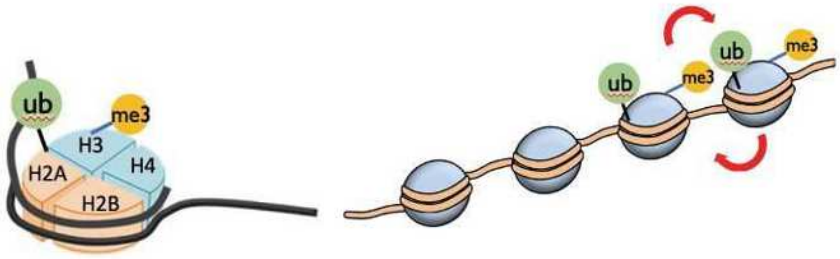


Figure 2. Schematic representation of the future experiments needs to be conducted after the reconstitution of nucleosome and crosstalk between neighboring marks.

1. Shilatifard, A. Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. *Annu. Rev. Biochem.* **75**, 243–269 (2006).
2. Debelouchina, G. T., Gerecht, K. & Muir, T. W. Ubiquitin utilizes an acidic surface patch to alter chromatin structure. *Nat. Chem. Biol.* **13**, 105–110 (2017).
3. Yuan, W. *et al.* Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation. *Science (80-.).* **337**, 971 LP – 975 (2012).
4. Meas, R. & Mao, P. Histone ubiquitylation and its roles in transcription and DNA damage response. *DNA Repair (Amst).* **36**, 36–42 (2015).
5. Smith, E. & Shilatifard, A. The Chromatin Signaling Pathway: Diverse Mechanisms of Recruitment of Histone-Modifying Enzymes and Varied Biological Outcomes. *Mol. Cell* **40**, 689–701 (2010).
6. Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* **21**, 569–571 (2014).