Chemical Approaches to Study the Effect of Histone Post-translational Modifications

Esha Upadhyay

A thesis

submitted in partial fulfilment of the

requirements for the degree of

Master of Science

University of Washington

2019

Committee

Champak Chatterjee

Robert E. Synovec

Program Authorized to Offer Degree

Chemistry

University of Washington

Abstract

Chemical approaches to Study the Effect of Histone Post-translational Modifications

Esha Upadhyay

Chair of the Supervisory Committee:

Champak Chatterjee

Department of Chemistry

Genetic material is stored in a massive nucleoprotein complex called Chromatin. The basic unit of Chromatin is the Nucleosome Core Particle or NCP, made up of 147 bp DNA wound around a core of histone protein octamer comprising two of each histone H2A, H2B, H3 and H4. The positively charged tails of the histones extend outwards into the nucleosol and are subjected to various post-translational modifications (PTMs) by chromatin modifying enzymes. Histone PTMs are important regulators of a variety of critical cellular processes such as DNA damage repair, transcriptional regulation. Multiple human cancers, autoimmune and neurodegenerative diseases can occur from the misregulation and mistranslation of histone PTMs.

Ubiquitylation and SUMOylation are two such examples of histone PTMs. Studies on the ubiquitylation of histone H2A in a nucleosomal context by the breast cancer predisposition genes BRCA1/BARD1 heterodimer, showed that missense mutations in the RING domain of BARD1 disrupts the ubiquitin ligase activity of the heterodimer as well as interferes with its ability to bind to nucleosomes. This in turn leads to loss of transcriptional control of genes that

encode estrogen- metabolizing enzymes. These studies were facilitated by access to quantities of wild-type mononucleosomes produced in the Chatterjee lab at UW that was sufficient for carrying out biochemical assays and NMR experiments.

The effect of the small ubiquitin-like modifier (SUMO) or SUMOylation is less well understood and requires access to uniformly and site-specifically modified NCPs. A successful protein-semisynthetic strategy developed by the Chatterjee lab utilized a chemical ligation auxiliary, 2-(aminooxy)ethanethiol in order to install SUMO at Lys 12 in histone H4 (suH4). This process generated sufficient quantities of suH4 that enabled investigations into its effect on chromatin and cross-talk with other histone PTMs such as acetylation and methylation, mediated by transcriptionally repressive complexes. The large-scale reconstitution of mononucleosomes and optimization of the semisynthesis of suH4 is described here.

Table of Contents

Chapter 1: Ubiquitylation of nucleosomal H2A by the BRCA1/BARD1 heterodimeric complex

1.1 Introduction

Genetic material is stored in a massive and dynamic nucleoprotein complex called Chromatin. The fundamental building block of chromatin is the nucleosome core particle (NCP) that comprises of 147 bp of double-stranded DNA wrapped around a core of histone proteins.(H2A, H2B, H3 and H4).[1] Extending into the nucleosol are the functional- group rich histone tails which are the sites for a variety of chemically distinct and reversible post-translational modifications (PTMs). Some of these include arginine and lysine side-chain methylation; serine, threonine, or tyrosine phosphorylation, and modification of lysine side-chains by smaller acetyl to longer crotonyl and butyryl groups and even entire proteins such as ubiquitin (Ub) and SUMO (small ubiquitin-like modifier). Studies into the regulation of chromatin structure and function by reversible chromatin modifications have generated enormous interest in the past decade. As recently as August 2018, chemical and biochemical studies identified a new histone mark, lysine benzoylation, regulated by SIRT2, an NAD^+ dependent protein deacetylase. [2-4]

Histone PTMs are essential for transcriptional regulation as well as for maintaining a euchromatic or heterochromatic state. Specific marks have been shown to be associated with active or repressed transcription. For example, histone H3/H4 acetylation, H3K4me2/3, H3K36me2/3 and H2BK120ub are shown to be associated with active transcription but H3K9me3, H3K27me3, H2AK119ub, and H4K12 SUMOylation are considered to be repressive marks.[5,6]

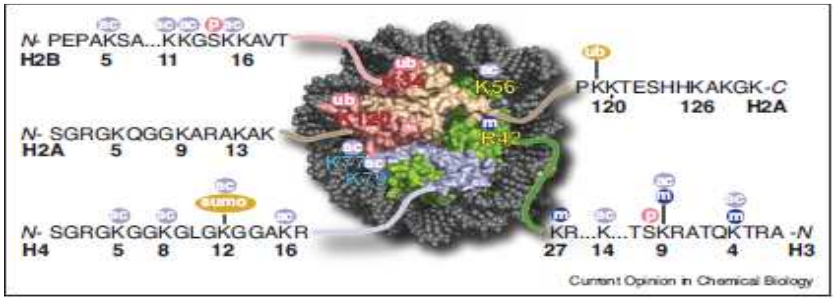


Figure 1. Histone tail marks. Scheme showing the diversity of marks that have been accessed by chemical protein synthesis and semisynthesis. Chemical groups are indicated as ac, acetyl; m, methyl; p, phosphoryl; sumo, small ubiquitin-like modifier; ub, ubiquitin. The globular core of the NCP is shown with histones coloured as H2A (gold), H2B (red), H3 (green) and H4 (blue) and double-stranded DNA (gray). PDB code 1KX5. [5]

Histone PTMs seldom occur by themselves. Histone tails can have multiple modifications, which combined together constitute a so-called 'histone code' capable of regulating crucial cellular processes such as transcription, DNA damage repair, gene silencing etc. [7]

One such example is the ubiquitylation of Lys127 and Lys129 in the C-terminal tail of histone H2A by the BRCA1/BARD1 heterodimer, which is an important step in DNA damage repair and in the transcriptional regulation of certain estrogen-metabolizing enzymes by BRCA1.

BRCA1 is the breast cancer predisposition gene, found associated with BARD1 and belong to the tumour suppressor gene family. Both proteins contain a RING domain and can exist as homodimers but prefer to exist as a heterodimer with ubiquitylating activity.[8] When part of the heterodimer, the BRCA1 RING domain, interacts with an E2 enzyme, facilitating the transfer of ubiquitin to the substrate.[9] Missense mutations in the critical residues in the BRCA1 RING domain have been linked to high risks of breast and ovarian cancer.[10] However the effect of loss-of-function mutations in the BARD1 RING domain so far remained unclear.

A recent study by Stewart *et al*. at the Klevit lab, Department of Biochemistry, University of Washington (UW) showed that missense mutations of the zinc-binding residues of BARD1 identified in a family severely affected by breast cancer, indicated that the presence of BARD1 in the heterodimer is essential for the ubiquitylation of histone H2A on nucleosomal substrates.[8] Further, preliminary NMR studies conducted in the Klevit lab on the BRCA1/BARD1/Nucleosome complex showed that specific residues in BARD1 might play a role in the recognition and binding of the heterodimeric complex to nucleosomal substrates.

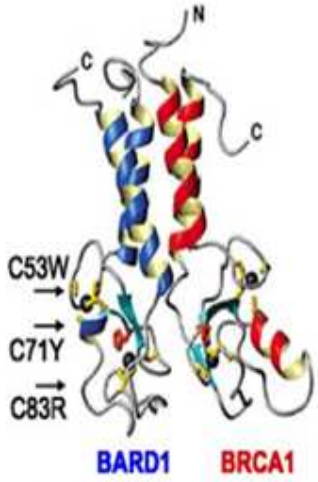


Figure 2. The BRCA1/BARD1 heterodimer (based on pdb entry 1JM7) indicates the positions of the missense mutations in the BARD1 RING domain (shown as blue subunit). [8]

1.2 Experimental procedures

1.21 Histone octamer formation

Individual histones were overexpressed in E. coli BL21 (DE3) cells and purified by size exclusion chromatography on a Superdex 200 column followed by reverse- phase high performance liquid chromatography (RP-HPLC).Histone stock solutions were made upto a concentration of 4 mg/ml by dissolving 1-2 mgs each in appropriate amounts of unfolding buffer containing 7 M Gn.HCl, 20 mM Tris, pH=7.5 at 25°C. Exact concentrations were determined using the 280 nm extinction coefficients: H2A, $\varepsilon = 4470\ M^{-1}cm^{-1}$; H2B,$\varepsilon = 7450\ M^{-1}cm^{-1}$; H3 C110A, $\varepsilon = 4470\ M^{-1}cm^{-1}$; H4, $\varepsilon = 5960\ M^{-1}cm^{-1}$. Appropriate volumes of each stock was combined in equimolar amounts and dialysed against 2 lt of refolding buffer containing 2 M NaCl, 10 mM Tris,1 mM EDTA, pH= 7.5 at 4°C two times for 3 h and then once overnight to generate octamers followed by purification by size-exclusion chromatography. Purified octamers generated were at a concentration of 73.3 µM and final volume of 500 µl. Final octamer concentration was determined with the 280 nm extinction coefficients: $\varepsilon = 44{,}700\ M^{-1}cm^{-1}$

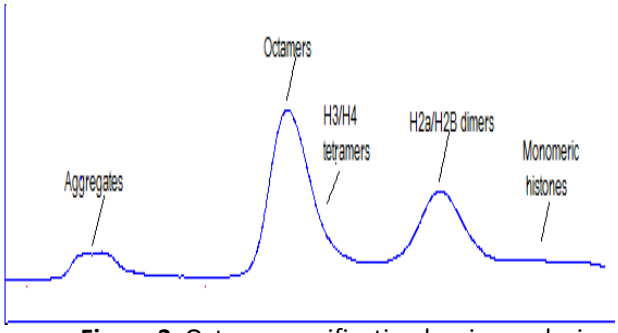


Figure 3. Octamer purification by size-exclusion chromatography

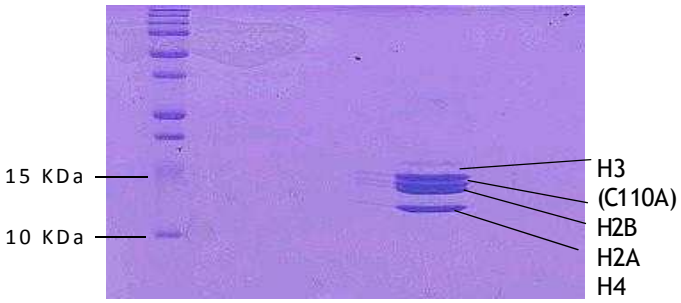


Figure 4. Histone octamers visualized on 15% SDS-PAGE gel

1.22 Generation of 147 bp 601 DNA

Equal volumes of single stranded-DNA were separately dissolved in buffer containing 10 mM Tris, 50 mM NaCl, 1mM EDTA, pH=8, at 25°C, to a final concentration of 100 µM and

incubated at 50°C followed by cooling for 60 secs. The DNA stocks were then combined and heated to a temperature of 70°C followed by slow bench-top cooling to a final concentration of 42.4 µM and final volume of 400 µl. Successful annealing was verified by visualizing the annealed DNA on a 2% agarose gel followed by staining with EtBr and imaged using GE Typhoon Imager. Final concentration of DNA was calculated with the 260 nm extinction coefficient ε = 2.7845 $M^{-1}cm^{-1}$

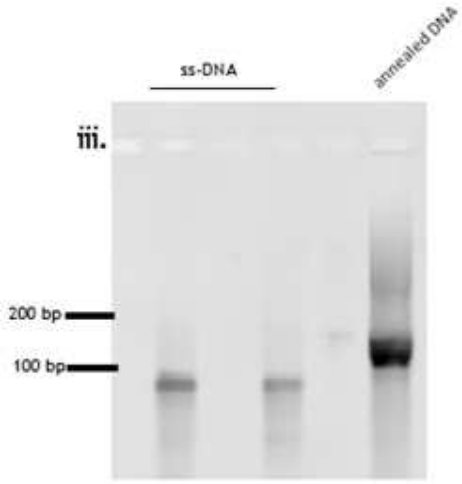


Figure 5. Annealed 147 bp DNA visualized on 2% agarose gel

1.23 Mononucleosome (MNs) reconstitution

Different concentrations of octamers and 147 bp DNA (1_147_601) [11, 12] were combined, to determine the best ratio for MN formation with minimal amount of free DNA. Octamers and DNA were then combined in the previously determined ratio in high-salt refolding buffer followed by slow dialysis into low-salt buffer to generate MNs at a concentration of 68 µM and final volume 80 µl, which were visualized on a 5% TBE gel run in 0.5x TBE buffer, followed by staining with ethidium bromide. Final concentration of MNs was calculated with the 260 nm extinction coefficient ε = 2.7845 $M^{-1}cm^{-1}$

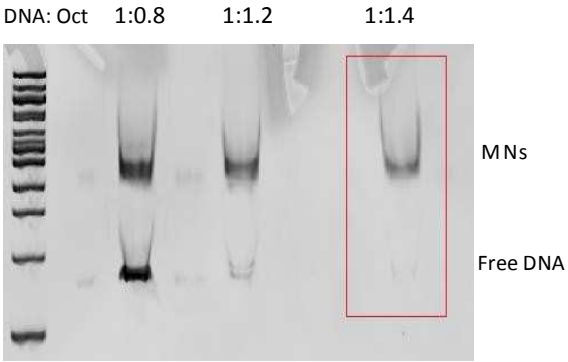


Figure 6. Tests to determine optimum DNA: Octamer ratio for MN formation

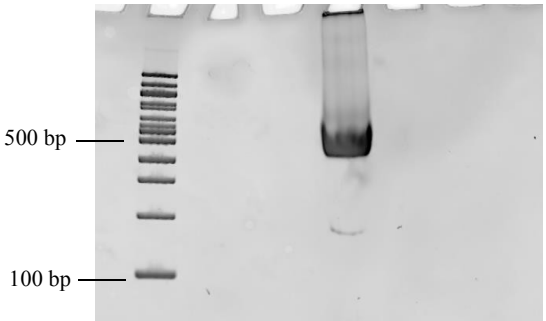


Figure 7. Large-scale MN formation, DNA: Oct:: 1:1.4 (conc.)

The reconstituted MNs were then used to assess the ubiquitylating and binding ability of the BARD1 variants towards nucleosomal substrates.

1.3 Results and discussion

The results from experiments conducted by Stewart *et al*. indicated that BARD1 played a major role in the ubiquitylating activity of the BRCA1/BARD1 heterodimer. The heterodimeric complex incorporating any of the 3 missense mutations identified in the highly conserved zinc coordinating residues of the BARD1 ring domain: p.C53W, p.C71Y, p.C83R, failed to ubiquitylate H2A in the nucleosomal context. However these BARD1 variants retained their ability to ubiquitylate free H2A as well as other substrates such as ERα and autoubiquitylate BRCA1. [8]

In figure 8, ubiquitylated H2A appears as higher molecular weight bands corresponding to mono and di- ubiquitylated H2A. The band for unmodified H2A disappears. The positive control used for these experiments is BRCA1 p.C61G, a known BRCA1 mutant that interferes with the ubiquitylating activity of the heterodimer towards all substrates. Also shown here is the BARD1 p.C78S variant, containing a mutated cysteine, not present in the zinc binding domain, which shows similar activity as its wild-type counterpart.

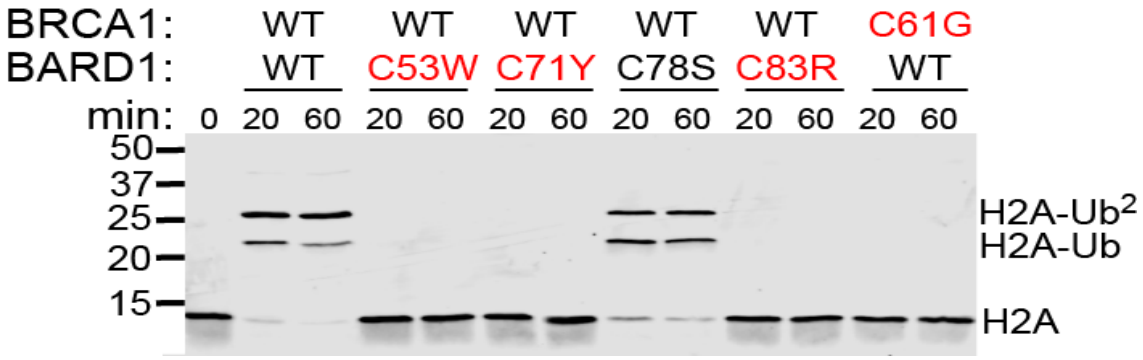


Figure 8. Wild-type BRCA1 in complex with BARD1 p.C53W, BARD1 p.C71Y, or BARD1 p.C83R is deficient at in vitro ubiquitylation of nucleosomal histone H2A, as is BRCA1 p.C61G in complex with wild-type BARD1. In contrast, wild-type BRCA1 in complex with BARD1 p.C78S has normal ubiquitylation activity. [8]

Preliminary NMR studies undertaken in the Klevit lab at UW Biochemistry, identified the BARD1 residue W91 to be important for the recognition and binding of nucleosomal substrates. The BARD1 p.W91A variant when compared to one of the previously identified missense mutations C53W, showed a similar loss-of-function towards nucleosomal H2A. (Figure 9)

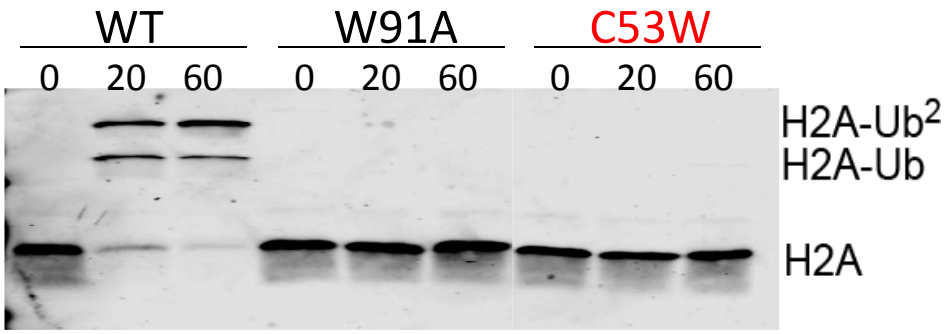


Figure 9. BARD1 p.W91A identified in NMR studies, is important for nucleosome binding.
(Unpublished data from Mikaela Stewart, Klevit Lab, UW Biochemistry)

1.4 References

1. Luger, K., Mäder, A., Richmond, R., Sargent, D. and Richmond, T. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), pp.251-260.
2. Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell*, 128(4), pp.693-705.
3. Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S., Falck, J., Peng, J., Gu, W. and Zhao, Y. (2007). Lysine Propionylation and Butyrylation Are Novel Post-translational Modifications in Histones. *Molecular & Cellular Proteomics*, 6(5), pp.812-819.
4. Huang, H., Zhang, D., Wang, Y., Perez-Neut, M., Han, Z., Zheng, Y., Hao, Q. and Zhao, Y. (2018). Lysine benzoylation is a histone mark regulated by SIRT2. *Nature Communications*, 9(1).
5. Tessarz, P. and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews Molecular Cell Biology*, 15(11), pp.703-708.
6. Zhang, T., Cooper, S. and Brockdorff, N. (2015). The interplay of histone modifications - writers that read. *EMBO reports*, 16(11), pp.1467-1481.
7. Strahl, B. and Allis, C. (2000). The language of covalent histone modifications. *Nature*, 403(6765), pp.41-45.
8. Stewart, M., Zelin, E., Dhall, A., Walsh, T., Upadhyay, E., Corn, J., Chatterjee, C., King, M. and Klevit, R. (2018). BARD1 is necessary for ubiquitylation of nucleosomal histone H2A and for transcriptional regulation of estrogen metabolism genes. *Proceedings of the National Academy of Sciences*, 115(6), pp.1316-1321.
9. Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., Ogata, H. and Ohta, T. (2001). The RING Heterodimer BRCA1-BARD1 Is a Ubiquitin Ligase Inactivated by a Breast Cancer-derived Mutation. *Journal of Biological Chemistry*, 276(18), pp.14537-14540.
10. Densham, R., Garvin, A., Stone, H., Strachan, J., Baldock, R., Daza-Martin, M., Fletcher, A., Blair-Reid, S., Beesley, J., Johal, B., Pearl, L., Neely, R., Keep, N., Watts, F. and Morris, J. (2016). Human BRCA1–BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection. *Nature Structural & Molecular Biology*, 23(7), pp.647-655.
11. Lowary, P. and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, 276(1), pp.19-42.
12. Dyer, P., Edayathumangalam, R., White, C., Bao, Y., Chakravarthy, S., Muthurajan, U. and Luger, K. (2003). Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. *Chromatin and Chromatin Remodeling Enzymes, Part A*, pp.23-44.

Chapter 2: Semisynthesis of SUMOylated histone H4

2.1 Introduction

Several decades since the discovery of histone ubiquitylation, a report by Eisenman and Shiio, described the modification of the histone H4 tail and to a much lesser extent the H2A, H2B and H3 tail, by two isoforms of the small ubiquitin-related modifier protein, SUMO-1 and SUMO-3. In order to demonstrate transcriptional repression due to SUMOylation on H4, the authors fused UBC9 (a SUMO-conjugating enzyme) to the DNA-binding domain of GAL4 and examined the subsequent effect on a GAL4 reporter gene. The GAL4-UBC9 strongly repressed transcription, which was consistent with the hypothesis that SUMOylation of chromatin mediates transcriptional repression. [1,18] In 2011, a study by Galisson *et al*. identified lysine 12 in histone H4 as the preferred site for SUMOylation by SUMO-3 in human cells.[2] In 2014, a study by Dhall *et al.* demonstrated that the steric bulk of SUMO prevented chromatin compaction, making SUMOylation of H4 somewhat of a paradox. It is a transcriptionally repressive mark that does not simultaneously favour the compaction of chromatin. [3]

However, in spite of these recent discoveries, a lot remains unknown regarding the precise role of SUMO in chromatin as well as its crosstalk with other transcriptionally activating marks like histone acetylation and methylation. [4] The challenges faced when studying SUMOylation and its effects are similar to those of Ubiquitylation. Additional complications with SUMOylation arise due to the presence of multiple SUMO isoforms, multiple sites of modifications on histones and the absence till date of a suH4 specific antibody. [5]

Therefore the successful study of histone post-translational modifications and the unravelling of their crosstalk, requires uniformly and site-specifically modified Nucleosome Core particles (NCPs). This requirement has driven the development of several chemical biology tools that provide access to sufficient quantities of homogeneously and site-specifically modified histones. [6]

One such useful technique is protein semisynthesis, which has provided access to well-defined NCP substrates, enabling the study of their cross-talk and effect on chromatin. [7] An important advantage of protein semisynthesis is that it makes it possible to install several differently modified residues in close proximity to each other. Native chemical ligation (NCL) is a commonly used semisynthetic strategy that is able to overcome the limited yields from solid-phase peptide synthesis of proteins containing over 50 amino acid residues.[8] The NCL strategy allows the formation of an amide bond between two asymmetrical polypeptide fragments, one

containing a C-terminal thioester and the other incorporating an N-terminal Cysteine. The shorter fragment is usually the more synthetically accessible fragment and the one containing the desired histone PTMs. In the first step, both fragments are joined via a thioester linkage followed by a spontaneous N-S acyl shift to generate the native amide backbone between neutral to alkaline pH. [9] An extension of the NCL strategy is Expressed Protein Ligation or EPL, where a recombinantly expressed protein undergoes splicing in bacteria to generate the C-terminal thioester which is then ligated to the N-terminal cysteine containing peptide fragment. [10]

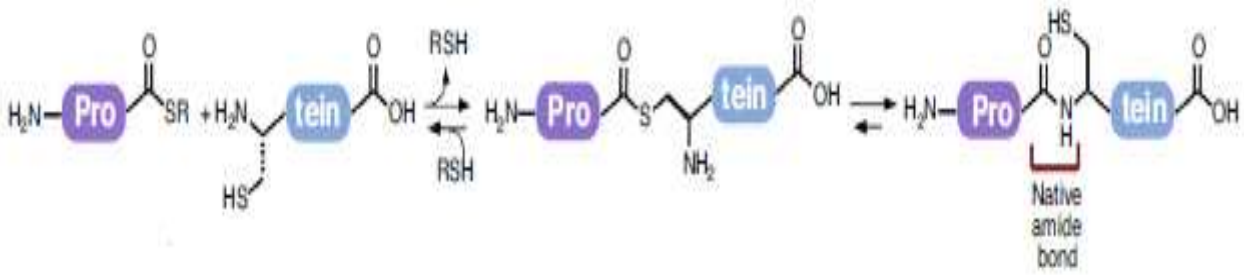


Figure 10. Native chemical ligation and strategies for histone semisynthesis. Native chemical ligation leads to the formation of a native amide bond between two polypeptide fragments. [5]

A strategy for the semisynthesis of suH4 was devised by the Chatterjee lab, Department of Chemistry at the University of Washington.[11] The method was based on the successful incorporation of a ligation auxiliary (2-aminooxy ethanethiol) developed by Kent and co-workers,[12] to synthesize SUMOylated peptides. The SUMOylated peptide, containing the first 14 residues of histone H4, was then ligated to the rest of the truncated histone H4. The complete scheme (adapted from Dhall *et al*.) is shown below. [13]

Detailed experimental procedures described below have been modified from Dhall *et al*. [1]

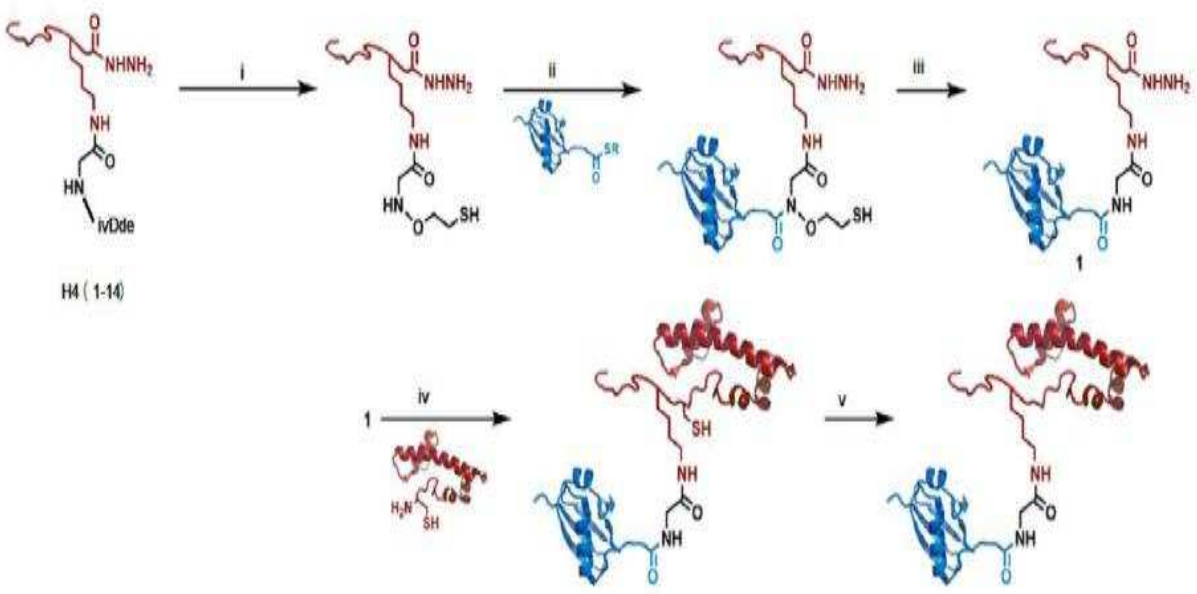


Figure 11. Semisynthesis of suH4

i. Site-specific coupling of the ligation auxiliary to H4 (1-14) Lys12 followed by acidolytic release of the fully unprotected peptidyl hydrazide from the solid-phase.

ii. Expressed protein ligation of H4 (1-14) aux -$CONHNH_2$ with SUMO-3(2-91)C47S-MES thioester to generate SUMOylated peptide hydrazide,H4(1-14)Su(C47S)(aux) -$CONHNH_2$.

iii. Zn-mediated auxiliary removal of the ligation auxiliary to yield SUMOylated H4 (1-14)-C $ONHNH_2$ followed by diazotisation to generate a highly reactive thioester.

iv. Native chemical ligation of the H4 (1-14) Su (C47S) - thioester with the H4 (15-102) A15C truncant protein to yield full-length SUMOylated H4 (A15C).

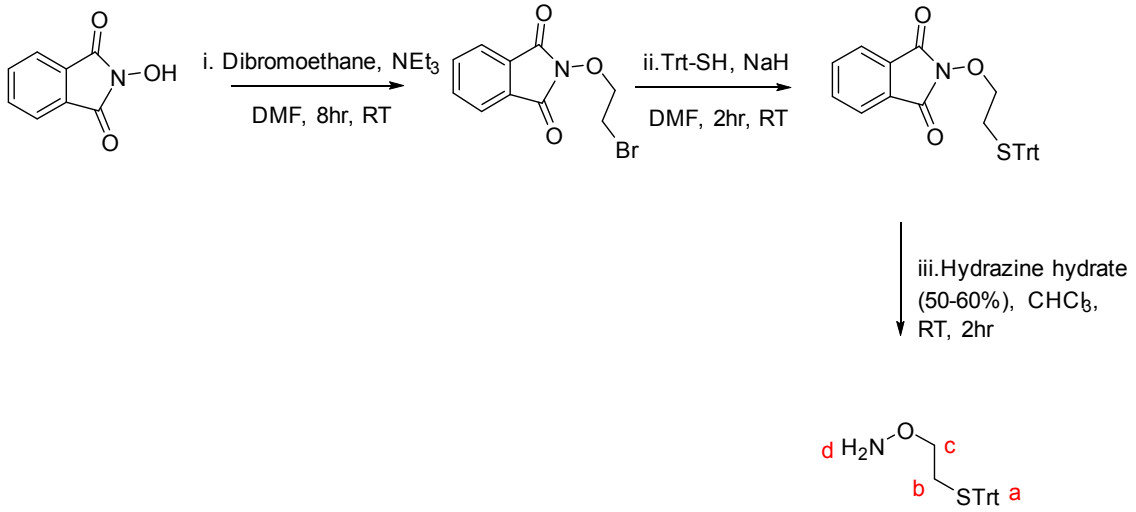
v. Desulfurization of Cys15 in H4 to generate full-length wild-type suH4.

2.2 Experimental procedures

2.21 Synthesis of the ligation auxiliary

The ligation auxiliary O-(2-(tritylthio)ethyl)hydroxylamine was prepared with 57% yield over 3 steps from N-hydroxyphthalimide as described in detail in Weller *et al*.[14] The steps are represented briefly in the scheme below.

a)



b)

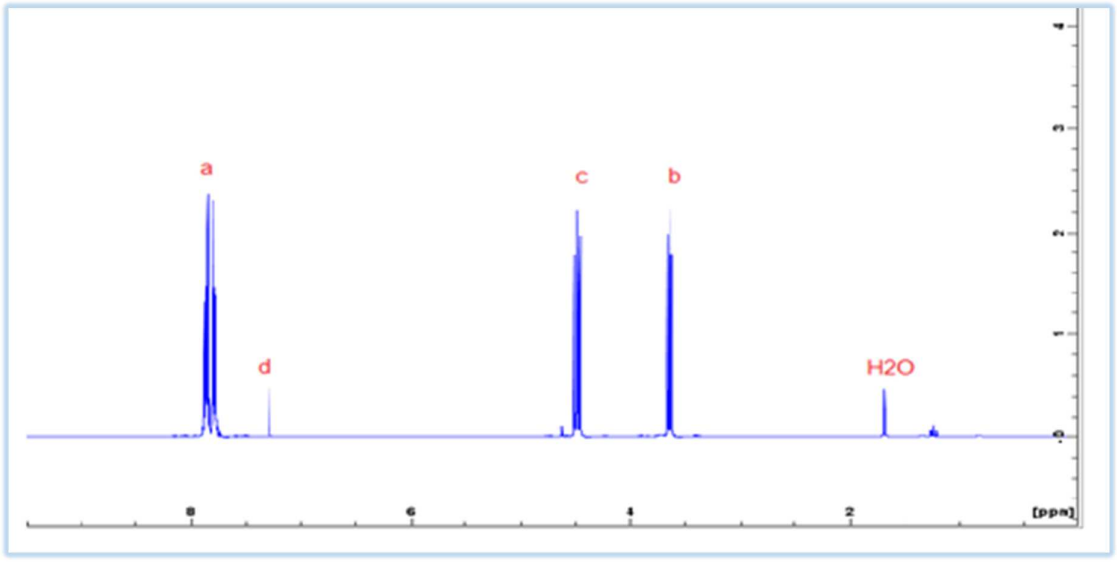


Figure 12. a) Reaction scheme for the chemical synthesis of O-(2-(tritylthio) ethyl) hydroxylamine
b) ^{1}HNMR of O-(2-(tritylthio) ethyl) hydroxylamine

2.22 Synthesis of H4 (1-14) aux-$CONHNH_2$

Synthesis of BocHN-H4 (1-14)-2-chlorotrityl hydrazine resin

The peptide BocHN-SGRGKGGKGLGKGG-$CONHNH_2$ corresponding to the first 14 N-terminal residues of the human histone H4 protein was synthesized by solid-phase peptide synthesis on a 0.25 mmol scale using standard 9-fluorenylmethoxycarbonyl (Fmoc)-based Nα -deprotection chemistry. Briefly, 2-chlorotrityl hydrazine resin was prepared by reacting 2-chlorotrityl chloride resin (1.33 mmol/g) in a 10% (v/v) solution of hydrazine in DMF at 30 °C for 30 min. The reaction was repeated once with fresh hydrazine solution. The resin was then treated with 10% (v/v) methanol in DMF for 10 min to cap any unreacted sites on the resin. The first amino acid, Gly, was coupled in 4-fold molar excess. The coupling reaction containing Fmoc-Gly-OH (1.0 mmol), O-(6-Chlorobenzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HCTU, 0.95 mmol), and DIEA (2.0 mmol) proceeded for 60 min at 30 °C. From glycinyl 2-chlorotrityl hydrazine resin each remaining amino acid was coupled in 5-fold molar excess based on initial resin loading. Deprotection of the Fmoc- group was achieved by treating resin with 20% (v/v) piperidine in DMF for 3 min at 75 °C. Coupling reactions were undertaken for 5 min at 75 °C with a mixture of Fmoc-amino acid (1.31 mmol), HBTU (1.28 mmol) and DIEA (2.75 mmol) in DMF. For Fmoc-Arg(Pbf)-OH, an additional coupling reaction was performed for 25 min at 75 °C. The Lys at position 12 was orthogonally protected with the 4, 4-dimethyl-2, 6-dioxocyclohex-1-ylidene (ivDde) protecting group. Finally, the peptide was protected at the α-NH2 position with Boc group by reaction with di-tert-butyl dicarbonate (2.0 mmol) and DIEA (4.0 mmol) in DMF for 2 hours.

Attachment of the ligation auxiliary

Deprotection of the ivDde group was achieved by reacting resin bound peptide with a solution of 5% (v/v) hydrazine in DMF for 5 min. The deprotection was repeated twice, for a total of three times. The peptidyl resin was then coupled to bromoacetic acid (8-fold molar excess) with N, N′-Diisopropylcarbodiimide (DIC, 8-fold molar excess) in DMF for 45 min at room temperature. The coupling was repeated once. Subsequently, dry peptidyl resin was placed in a solution containing 9 equivalents of auxiliary (0.5 M in DMSO) and shaken for 24 hours at room temperature.

Initial attempts at attaching the auxiliary to the lysine 12 of the H4 peptide failed. ESI-MS data showed the presence of a truncated 1306 Da species alongside the correct product

corresponding to 1363 Da. Further analysis by Tandem MS, indicated the possibility of a chloride capping mechanism shown below:

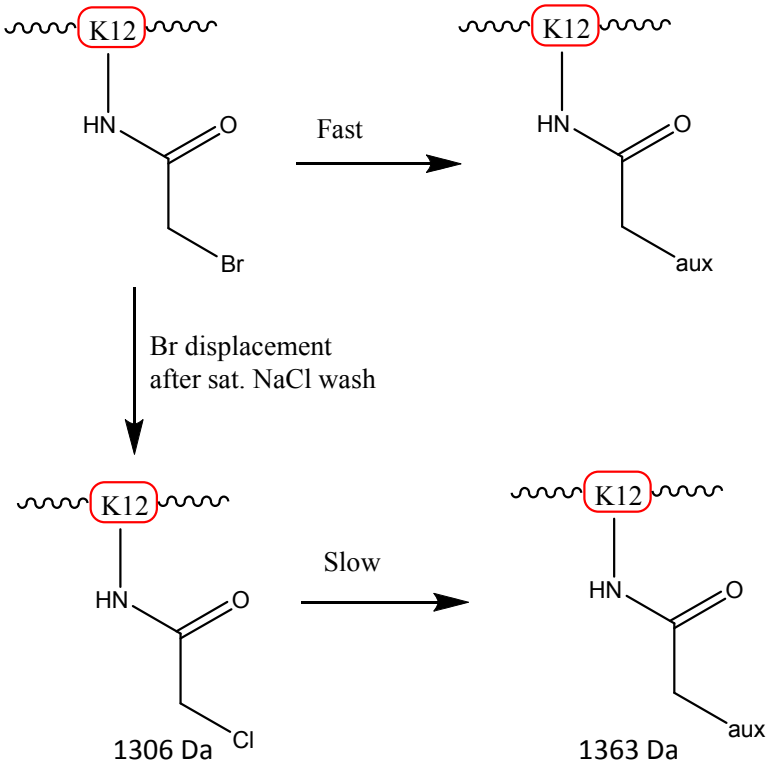


Figure 13. Schematic showing a competing chloride capping mechanism

In the first step, chloride ions displace bromide in the bromoacetic acid coupled peptide forming the 1306 Da species. The displacement of chloride ions by the auxiliary is a much slower process that leads to the accumulation of the 1306 Da intermediate species detected by ES1-MS. To prevent this competing mechanism, a saturated NaCl wash performed in the final step of the ligation auxiliary synthesis was omitted as it was thought to be the source of the chloride ions in the reaction. This precaution ensured the successful displacement of bromine by the ligation auxiliary. The H4(1-14)-aux peptide was cleaved and deprotected by reaction of resin at 20 µL/mg with Reagent K (TFA: thioanisole: H2O: phenol: 1,2-ethanedithiol 82.5:5:5:5:2.5 v/v) for 1.5 hours at room temperature, then precipitated and washed 2 times with ice-cold diethyl ether. The dry peptide was dissolved in RP-HPLC buffer A and purified by C18 preparative RP-HPLC with a gradient of 0-50% B over 60 min. This yielded 9% of the peptide-auxiliary conjugate H4 (1-14) aux-C (O) NHNH2 based on initial resin loading. ESI-MS of H4 (1-14) aux-$CONHNH_2$, calculated m/z [M+H] + 1,363.6 Da, observed 1,363.8 Da.

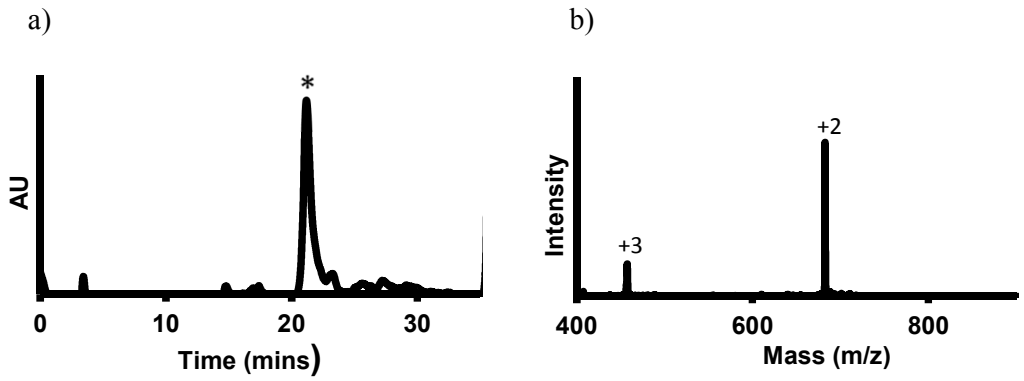


Figure 14. a) C18 analytical RP-HPLC chromatogram of H4 (1-14)-aux-$CONHNH_2$, gradient of 0-73% B, 30 min. b) ESI-MS of H4 (1-14)-aux-$CONHNH_2$. Asterix indicates the product

2.23 Overexpression and purification of SUMO-3(2-91) C47S-MES

E. coli BL21(DE3) cells with the plasmid pTXB1-SUMO-3(2-91)C47S were grown in 4 L Luria-Bertani medium supplemented with 100 µg/mL of Ampicillin at 37 ˚C with shaking at 250 rpm until OD600 ~0.6-0.8. Cells were cooled to 16 ˚C and induced with 300 µL 1 M IPTG. Cells were harvested by centrifugation (20 min at 6000 xg at 4 ˚C) after 18 h. Cells were resuspended in lysis buffer containing 50 mM sodium phosphate, 300 mM sodium chloride, 5 mM imidazole, pH 8, and sonicated on ice. Lysed cells were spun down (30 min at 20,000xg at 4 ˚C) and supernatant vacuum filtered and applied to 5 mL of Ni-NTA resin pre-equilibrated with lysis buffer. Supernatant was bound for 1 h at 4 ˚C. The flow- through from the column was then colleceted and Ni-NTA resin washed with 20 column volumes of lysis buffer, followed by 10 column volumes of 50 mM sodium phosphate, 300 mM sodium chloride, 20 mM imidazole, pH 8 and then with 5 x 1 column volume of 50 mM sodium phosphate, 300 mM sodium chloride, 50 mM imidazole, pH 8. Protein was eluted with 5 x 1 column volume of 50 mM sodium phosphate, 300 mM sodium chloride, 250 mM imidazole, pH 8. Fractions were evaluated by 15% SDS-PAGE gel, pooled and dialyzed twice against thiolysis buffer containing 100 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, 1 mM MESNa, pH 7.2 at 4 ˚C. The contents of the dialysis bag were then added to a volume of 1 M MESNa, 100 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, pH 7.2 to a final MESNa concentration of 200 mM. The solution was then placed in 45 mL falcon tubes, and nutated at 30 ˚C. The thiolysis reaction was allowed to proceed for over 6 hours, then lyophilized to dryness. The protein fraction was re-solubilized in 6 M Guanidine. HCl and purified by C18 preparative RP-HPLC employing a gradient of 0-73% B over 60 min. Typical

yield was 8-10 mg/L of cell culture. ESI-MS for SUMO-3(2-91) C47S-MES, calculated m/z [M+H] + 10,444.7 Da, observed 10,445.8 ± 3.6 Da.

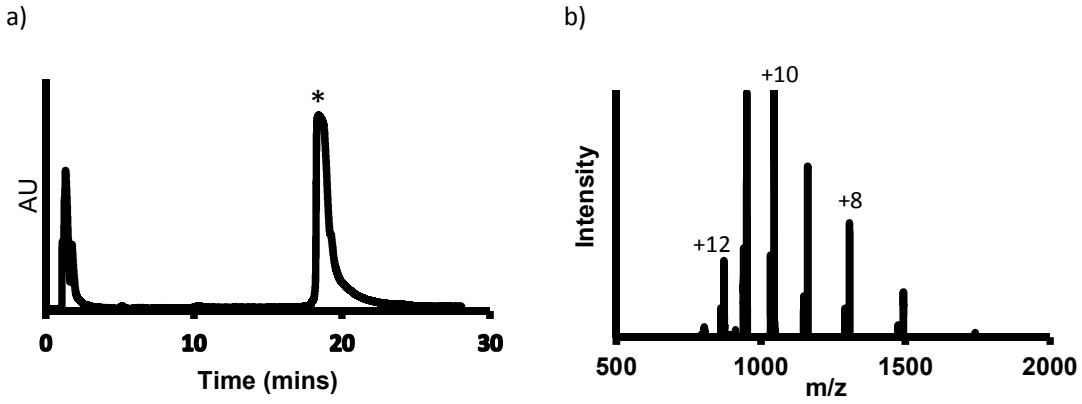


Figure 15. a) C18 analytical RP-HPLC chromatogram of SUMO-3(2-91) C47S-MESNA thioester, gradient of 0-73% B, 30 min. b) ESI-MS of SUMO-3(2-91) C47S-MESNA thioester

2.24 Expressed protein ligation of H4 (1-14) (aux)-$CONHNH_2$ and SUMO-3(2-91) C47S-MES

Purified H4(1-14)(aux) (14.8 mg) and SUMO-3(2-91)C47S-MES (18.6 mg, 6 eq, 0.25 mM) were dissolved in ligation buffer consisting of 6 M Gn·HCl, 100 mM Na_2HPO_4, and 10 mM TCEP, pH 7.3 at RT. Ligation proceeded with gentle shaking at 25 °C for 24 h. Ligation product was purified by C18 preparative RPHPLC employing a gradient of 25-50% B over 60 min to give 6.4 mgs (43%). ESI-MS of H4(1-14)K12su(C47S)(aux)-$CONHNH_2$, calculated m/z [M+H] + 11,666.1 Da, observed 11,668 ± 1.34 Da.

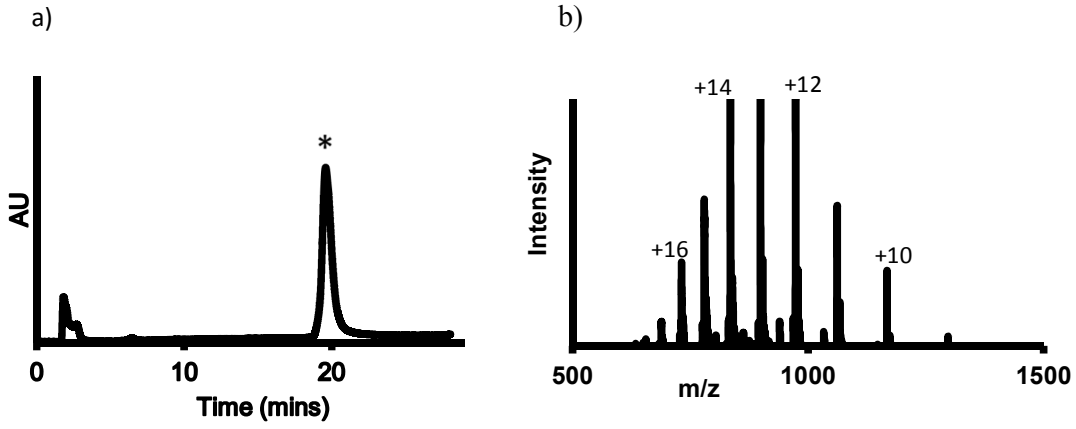


Figure 16. a) C18 analytical RP-HPLC chromatogram of H4 (1- 14) Su (C47S) (aux)-$CONHNH_2$, gradient of 0-73% B, 30 min. b) ESI-MS of H4 (1- 14) Su (C47S) (aux)-$CONHNH_2$

2.25 Zn-mediated auxiliary removal from H4 (1-14) Su (C47S) (aux)-$CONHNH_2$

Metallic Zn was freshly activated by stirring in a solution of 5% HCl for 5 min followed by washing with water, ethanol, and diethyl ether, and dried over vacuum. Activated Zn powder (1 g) added to degassed 6 M Gn·HCl, pH 3. Degassing accomplished by 3 freeze-thaw cycles under Ar using isopropanol-dry ice mixture. SUMOylated- H4 peptide was dissolved in reaction buffer at a conc. of 0.5 mg/ml and reaction allowed to proceed at 37 ^{0}C with gentle shaking for 24 h, protected from light. The reaction mixture was briefly centrifuged at 13,000 rpm to pellet the Zn and the supernatant containing reduced products was removed. The pelleted Zn was washed twice with 0.5 mL of 6 M Gn·HCl, 50 mM EDTA, pH 3. The combined supernatant and washes were purified by C18 semi-preparative RP-HPLC with a gradient of 30-80% B over 45 min to give 4.7 mg of reduced product in 73% final yield. ESI-MS of H4 (1-14) Su (C47S)-$CONHNH_2$, calculated m/z [M+H] + 11,589.9 Da, observed 11,591.4 ± 0.8 Da.

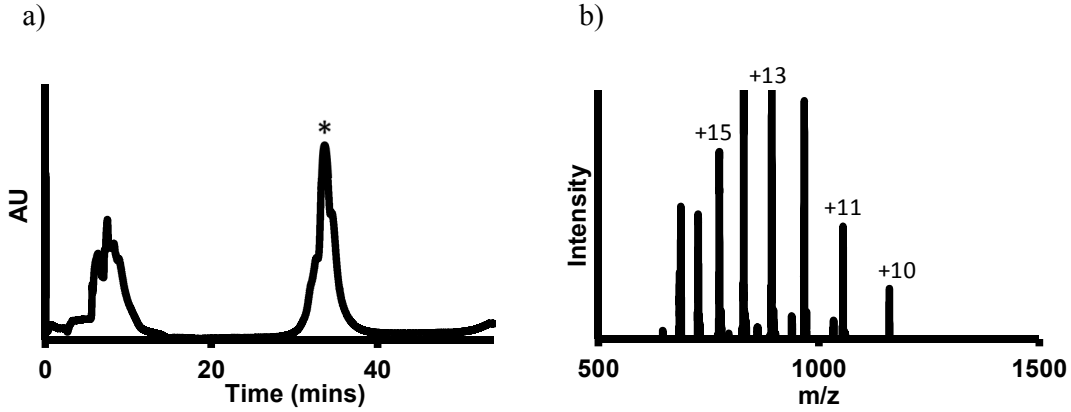


Figure 17. a) C18 semi-preparative RP-HPLC chromatogram of H4 (1- 14) Su (C47S)-$CONHNH_2$, gradient of 30-80% B, 30 min. b) ESI-MS of H4 (1- 14) Su (C47S)-$CONHNH_2$

2.26 Overexpression and purification of H4 (15-102) A15C

E. coli BL21(DE3) cells containing pET15b-His6-[TEV]-H4(15-102)A15C were grown in 3 L Luria Bertani medium supplemented with 100 µg/mL of Ampicillin at 37 ^{0}C with shaking at 250 rpm until OD600 ~0.6. Overexpression was induced by the addition of 0.3 mM IPTG and cells were grown for an additional 1.5 h at 37 ^{0}C. The cells were harvested by centrifugation at 7,000xg for 15 min. Cells were resuspended in wash buffer (20 mM tris, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 7.5, 1% triton X-100) and lysed by sonication on ice. Inclusion bodies were pelleted by centrifugation at 20,000xg for 20 min and washed twice

with wash buffer. Inclusion bodies were then dissolved in extraction buffer (EB) containing 6 M Gn·HCl, 20 mM tris, 1 mM 2-mercaptoethanol, pH 7.5 and applied to Ni-NTA resin. Column binding proceeded for 3 h at 4 °C, after which the resin was washed with 10 CV extraction buffer containing 25 mM imidazole. The protein was eluted with 3 x 1 CV extraction buffer containing 400 mM imidazole, then dialyzed into water containing 1 mM DTT. After dialysis, 10x cleavage buffer was added for final concentrations of 50 mM tris, 1 mM EDTA, 10 mM DTT, 10 mM L-cysteine, pH 6.9. Purified TEV protease was added to ~20% of the final volume, and the cleavage reaction proceeded for 1 h at 37 °C. The reaction was then dialyzed back into extraction buffer, incubated overnight at 4 °C with Ni-NTA resin to remove the his-tagged TEV protease and cleaved H4 N-terminal tail, and the column flow-through containing H4(15-102)A15C purified by C4 semi-preparative RP-HPLC employing a gradient of 40- 70% B over 60 min. Typical yields were 4-5 mg/L of cell culture. ESI-MS of H4 (15-102) A15C, calculated m/z [M+H] + 10,071.8 Da, observed 10,072.7 ± 0.6 Da.

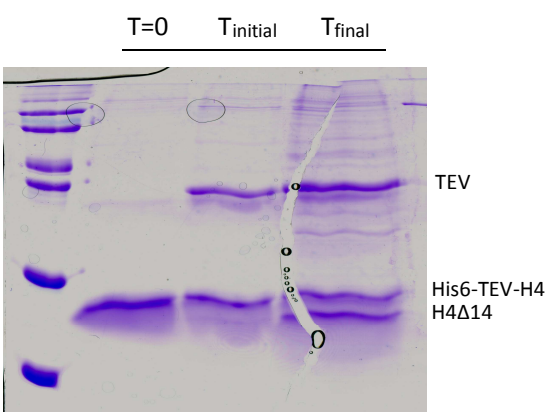


Figure 18. His6-H4Δ14 cleavage by TEV protease

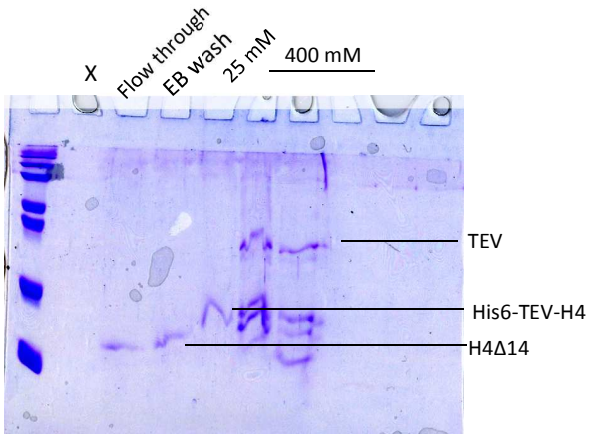


Figure 19. Ni column fractions showing cleaved H4Δ14 Separated from His6-H4Δ14 after 1 h incubation with Ni-NTA resin

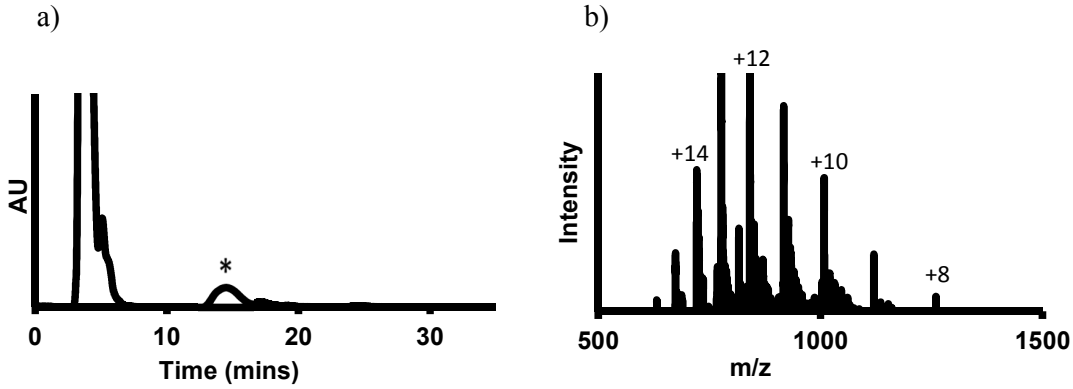


Figure 20. a) C4 semi-preparative RP-HPLC chromatogram of H4 (15- 102) A15C, gradient of 40-70% B, 30 min. b) ESI-MS of H4 (15- 102) A15C

2.27 Expressed protein ligation of H4 (1-14) Su (C47S)-$CONHNH_2$ and H4 (15-102) A15C

Ligation was accomplished by first converting the C-terminal hydrazide of the SUMOylated-H4 peptide to an acyl azide with $NaNO_2$ via the diazotization reaction, as described in Zheng *et al*.[21] Subsequent addition of MPAA served to both quench the remaining $NaNO_2$ and generate a highly reactive C-terminal thioester for the ligation reaction. Purified SUMOylated-H4 peptide was dissolved at 0.5 mM conc. in 200 mM Na_2HPO_4, 6 M Gn·HCl, pH 3, and kept at -20 °C for 20 min. 500 mM solution of $NaNO_2$ in water added to the above for a final concentration of 10 mM. Reaction was briefly mixed, and kept at -20 °C for 15 min. This was followed by the addition of a solution of H4 (15-102) A15C (2.4 eq) dissolved at 1 mM in 200 mM Na_2HPO_4, 6 M Gn·HCl, 200 mM MPAA, pH 6.5, to the above. The reaction mixture was allowed to warm up to room temperature, and pH adjusted with 3 M NaOH to 6.8-7.0. The ligation reaction allowed to proceed at 25 °C for 24 h. The ligation product purified by C18 semi-preparative RP-HPLC employing a gradient of 30-70% B over 45 min. to give 2.6 mg. ESI-MS of H4 (A15C) Su (C47S), calculated m/z [M+H] + 21,628.7 Da, observed 21,634.0 ± 9.1 Da.

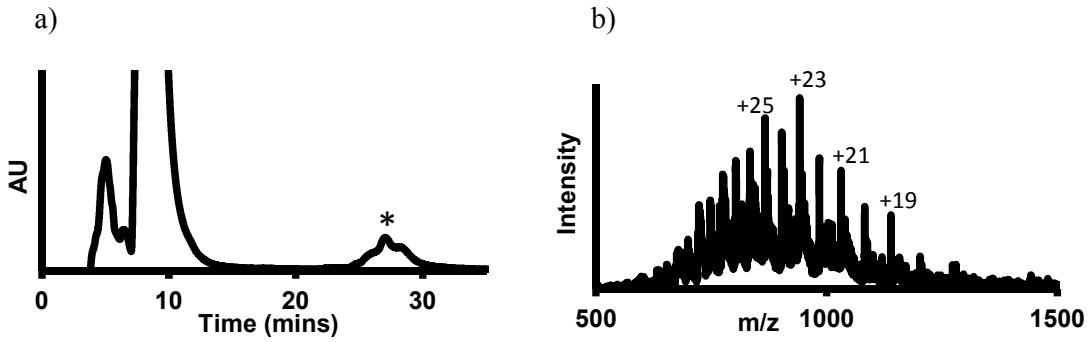


Figure 21. a) C18 semi-preparative RP-HPLC chromatogram of H4 (A15C) Su (C47S), gradient of 30-70% B, 30 min. b) ESI-MS of H4 (A15C) Su (C47S)

2.28 Desulfurization of H4 (A15C) Su (C47S)

Purified suH4 from the previous step was dissolved in degassed buffer containing Na_2HPO_4, 6 M Gn·HCl, 500 mM TCEP, 100 mM MESNa, pH 7.5 to a final concentration of 90 μM. 2-methyl-2-propanethiol (280 mM) and radical initiator 2,2'-Azobis[2-(2-imidazolin-2-yl)propane dihydrochloride (VA-044) (10 mM) was added to the above solution and the reaction incubated at 37 °C for 24 h. The product was purified by C18 analytical RP-HPLC

gradient of 0-73% B over 30 min. ESI-MS of desulfurized suH4, calculated m/z [M+H] + 21,596.7 Da, observed 21,606.9 ± 7 Da.

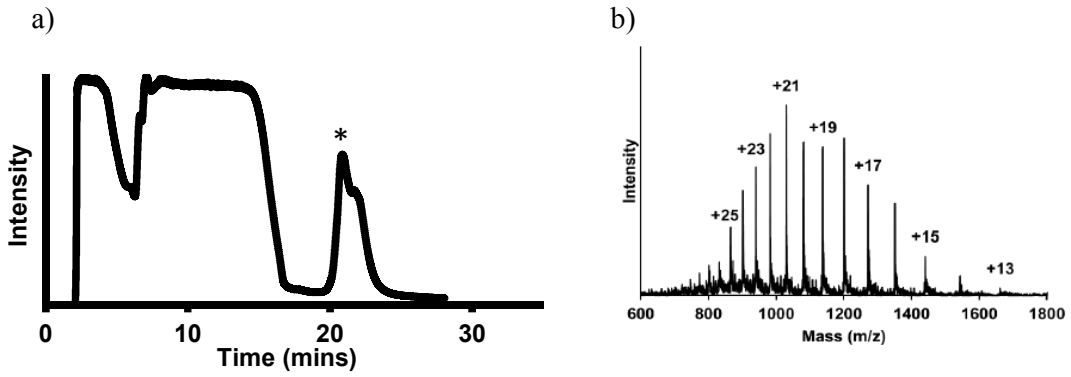


Figure 22. a) C18 analytical RP-HPLC chromatogram of H4 (A15C) Su (C47S), gradient of 0-73% B, 30 min. b) ESI-MS of desulfurized H4 (A15C) Su (C47S) (Data from Caroline Weller)

2.3 Conclusion

Based on this strategy, previous efforts at synthesising suH4 has successfully produced ˜2 mg of protein. The yield is somewhat affected by losses during multiple purifications by HPLC, low efficiency of the ligation steps and competing hydrolysis mechanisms in the latter stages of the process.

The availability of uniformly and site-specifically modified suH4 has enabled the study of its effect on transcription in the Chatterjee lab at UW. Several of the chromatin-modifying enzymes with their protein interaction partners are either modified by SUMO or contain SUMO-interacting motifs (SIM). A SIM comprises of a sequence of less than 10 amino acids and has a core of about 3-4 hydrophobic residues like Val or Ile, surrounded by acidic residues like Glu or Asp.[15] Experiments conducted in the Chatterjee lab have focused on a particularly critical transcriptional regulating complex, LSD1-CoREST-HDAC or LCH complex that contains a SIM.[20]

As part of the LCH complex, LSD1 or lysine specific histone demethylase carries out the demethylation of H3K4me1/2 and is responsible for the repression of a subset of neuronal genes in non-neuronal cells upon association with the RE1-silencing transcription factor (REST). [16] In 2009, it was also shown that the LCH complex represses REST-independent genes upon interaction with SUMO2/3 through its SIM in CoREST. [17]

Since both, suH4 and LSD1 are found to be associated with transcriptional repression, a study by Dhall *et al*. demonstrated potential crosstalk between suH4 and H3K4me2 and discovered a 2-fold increase in the rate of demethylation by LSD1 in a SIM dependent manner.[13] Further studies on the crosstalk between suH4 and other histone marks were conducted to include the deacetylation on of the H3K14Ac mark by the CoREST-HDAC1 complex.[19,20] Similar to the previous model with LSD1-CoREST, an increase in the rate of deacetylation by the histone deacetylase enzyme or HDAC was observed in the presence of suH4 and an intact SIM motif in CoREST. (Unpublished data)

Following the successful completion of all the individual steps required for the suH4 synthesis, future efforts in the lab will focus on scaling up the synthesis to several milligrams for crystallography efforts in collaboration with Ning Zheng's labs at UW Pharmacology.

2.4 References

1. Shiio, Y. and Eisenman, R. (2003). Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences*, 100(23), pp.13225-13230.
2. Galisson, F., Mahrouche, L., Courcelles, M., Bonneil, E., Meloche, S., Chelbi-Alix, M. and Thibault, P. (2010). A Novel Proteomics Approach to Identify SUMOylated Proteins and Their Modification Sites in Human Cells. *Molecular & Cellular Proteomics*, 10(2), pp.M110.004796.
3. Dhall, A., Wei, S., Fierz, B., Woodcock, C., Lee, T. and Chatterjee, C. (2014). Sumoylated Human Histone H4 Prevents Chromatin Compaction by Inhibiting Long range Internucleosomal Interactions. *Journal of Biological Chemistry*, 289(49), pp.33827-33837.
4. Fischle, W. (2008). Talk is cheap--cross-talk in establishment, maintenance, and readout of chromatin modifications. *Genes & Development*, 22(24), pp.3375-3382.
5. Leonen, C., Upadhyay, E. and Chatterjee, C. (2018). Studies of biochemical crosstalk in chromatin with semisynthetic histones. *Current Opinion in Chemical Biology*, 45, pp.27-34.
6. Dhall, A. and Chatterjee, C. (2011). Chemical Approaches To Understand the Language of Histone Modifications. *ACS Chemical Biology*, 6(10), pp.987-999.
7. He, S., Bauman, D., Davis, J., Loyola, A., Nishioka, K., Gronlund, J., Reinberg, D., Meng, F., Kelleher, N. and McCafferty, D. (2003). Facile synthesis of site-specifically acetylated and methylated histone proteins: Reagents for evaluation of the histone code hypothesis. *Proceedings of the National Academy of Sciences*, 100(21), pp.12033-12038.
8. Dawson, P., Muir, T., Clark-Lewis, I. and Kent, S. (1994). Synthesis of proteins by native chemical ligation. *Science*, 266(5186), pp.776-779.
9. Dawson, P. and Kent, S. (2000). Synthesis of Native Proteins by Chemical Ligation. *Annual Review of Biochemistry*, 69(1), pp.923-960.
10. Muir, T., Sondhi, D. and Cole, P. (1998). Expressed protein ligation: A general method for protein engineering. *Proceedings of the National Academy of Sciences*, 95(12), pp.6705-6710.
11. Weller, C., Huang, W. and Chatterjee, C. (2014). Facile Synthesis of Native and Protease-Resistant Ubiquitylated Peptides. *ChemBioChem*, 15(9), pp.1263-1267.
12. Canne, L. E.; Bark, S. J.; Kent, S. B. H. Extending the applicability of native chemical ligation. *J. Am.Chem. Soc.* 1996, *118*, 5891–5896.
13. Dhall, A., Weller, C., Chu, A., Shelton, P. and Chatterjee, C. (2017). Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1–CoREST Complex. *ACS Chemical Biology*, 12(9), pp.2275-2280.
14. Weller, C., Huang, W. and Chatterjee, C. (2014). Facile Synthesis of Native and Protease-Resistant Ubiquitylated Peptides. *ChemBioChem*, 15(9), pp.1263-1267.
15. Wang, Z., Zang, C., Rosenfeld, J., Schones, D., Barski, A., Cuddapah, S., Cui, K., Roh, T., Peng, W., Zhang, M. and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*, 40(7), pp.897-903.
16. Forneris, F., Binda, C., Vanoni, M., Battaglioli, E. and Mattevi, A. (2005). Human Histone Demethylase LSD1 Reads the Histone Code. *Journal of Biological Chemistry*, 280(50), pp.41360-41365.
17. Ouyang, J., Shi, Y., Valin, A., Xuan, Y. and Gill, G. (2009). Direct Binding of CoREST1 to SUMO-2/3 Contributes to Gene-Specific Repression by the LSD1/CoREST1/HDAC Complex. *Molecular Cell*, 34(2), pp.145-154.
18. Nathan, D., Sterner, D. and Berger, S. (2003). Histone modifications: Now summoning sumoylation. *Proceedings of the National Academy of Sciences*, 100(23), pp.13118-13120.
19. You, A., Tong, J., Grozinger, C. and Schreiber, S. (2001). CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proceedings of the National Academy of Sciences*, 98(4), pp.1454-1458.
20. Gocke, C. and Yu, H. (2008). ZNF198 Stabilizes the LSD1–CoREST–HDAC1 Complex on Chromatin through Its MYM-Type Zinc Fingers. *PLoS ONE*, 3(9), p.e3255.
21. Zheng, J., Tang, S., Qi, Y., Wang, Z. and Liu, L. (2013). Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nature Protocols*, 8(12), pp.2483-2495.