Synthesis, biochemical characterization and genetic encoding of acetyl-lysine mimicking amino acids for bromodomains and deacetylases

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Sören Benjamin Kirchgäßner
aus Ludwigsburg

Tübingen
2021

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	04.02.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Dirk Schwarzer
2. Berichterstatter:	Prof. Dr. Thorsten Stafforst
3. Berichterstatter:	Prof. Dr. Michael Lammers

Declaration of Authorship

I hereby declare that I alone wrote the doctoral work submitted here under the title "Synthesis, biochemical characterization and genetic encoding of acetyl-lysine mimicking amino acids for bromodomains and deacetylases", that I only used the sources and materials cited in the work. I have explicitly marked all material, which has been quoted either literally or by content from the used sources. Parts of this work were done in collaboration with other scientists whose names are stated in the dissertation.
I declare that I adhered to the guidelines set forth by the University of Tübingen to guarantee proper academic scholarship (Senate Resolution 25.05.2000). I declare that these statements are true and that I am concealing nothing. I understand that any false statements can be punished with a jail term of up to three years or a financial penalty.

Date: Tübingen, den

Signed:

Abstract

Nε-acetylation of lysine residues is one of the most abundant post-translational modifications (PTMs) of proteins in all kingdoms of life. Lysine acetylation is installed by histone acetyltransferases (HATs), removed by histone deacetylases (HDACs) and serves as binding site for bromodomains (BRDs). The orchestrated interplay of these proteins is involved in regulation of various cellular processes including chromatin remodeling and transcription. Dysregulation of HDACs and BRDs can result in the development of diseases such as cancer. This renders HDACs and BRDs as promising therapeutic targets. Insight into the substrates and binding sites of HDACs and BRDs is a prerequisite for understanding the disease mechanisms and the development of highly specific HDAC and BRD inhibitors. A first generation of synthetic inhibitors and small-molecule ligands have already been developed for HDACs and BRDs, which could potentially be re-designed into chemical probes for investigating HDAC and BRD substrate selectivities.

In this study, a set of triazole-containing amino acids was synthesized as potential acetyl-lysine mimics for BRDs on the basis of BRD inhibitor JQ1. The amino acids were incorporated into histone H3 and H4 derived peptide probes and tested for BRD binding in pull-down experiments. These experiments revealed aminoheptanedioic acid-ε-3-methyl-1,2,4-triazole (ApmTri) as most potent acetyl-lysine mimic for BRDs of the BET family. Recombinant BET BRD proteins BRD3(2), BRD4(2) and BRD4(1) bound ApmTri-containing peptide substrates with similar affinity as the native acetyl-lysine substrate as observed in pull-down experiments and indicated by K_D values. Obtained crystal structures of BRD3(2) and BRD4(1) with bound ApmTri substrates illustrated that the triazole moiety efficiently mimicked the Nε-amide in acetyl-lysine. ApmTri was further shown to be stable towards HDACs, enabling the subsequent genetic encoding and incorporation of this non-canonical amino acid into proteins of HEK cells. Co-expression of BRD3(2) and ApmTri-containing fusion proteins allowed co-precipitation of both interaction partners from native HEK lysates.

In order to generate new tools for HDAC research, the already established HDAC-trapping amino acid AsuHd was genetically encoded as well. Recombinant AsuHd-containing proteins were able to co-immunoprecipitate endogenous HDACs from HEK lysates. The ability of AsuHd-containing protein probes was further extended to whole nucleosome core particles (NCPs). In this case, sortase-mediated ligation was used for linking synthetic AsuHd-containing histone H3 tails to the NCP structure. AsuHd designer NCPs recruited HDAC complexes more specifically and efficient than AsuHd-containing peptide probes. Mass spectrometry-based interactome profiling of AsuHd-NCPs uncovered efficient enrichment of HDACs 1, 2, 6 and 38 new potential HDAC binding proteins.

Zusammenfassung

Die Nε-Acetylierung von Lysin-Seitenketten ist eine der häufigsten posttranslationalen Modifikationen von Proteinen. Lysin-Acetylierungen werden von Histon-Acetyltransferasen (HATs) eingeführt, von Histon-Deacetylasen (HDACs) abgespalten und dienen Bromodomänen (BRDs) als Bindemotiv. Das Zusammenspiel dieser Proteine reguliert eine Vielzahl zellulärer Prozesse, wie zum Beispiel Chromatin-Remodellierung und Transkription. Eine Deregulierung von HDACs und BRDs steht in Verbindung mit der Entwicklung von Krebs und einer Vielzahl weiterer Erkrankungen, was HDACs und BRDs zu vielversprechenden therapeutischen Angriffspunkten macht. Eine Voraussetzung für die Entwicklung hochspezifischer Inhibitoren für diese Proteine ist die Erforschung ihrer Substrate und Bindungsstellen. Eine erste Generation von synthetischen Liganden für HDACs und BRDs wurde bereits entwickelt und könnte die Entwicklung chemischer Sonden für die Erforschung der Substratselektivität dieser Enzyme ermöglichen.

In dieser Arbeit wurden Triazol-basierte Aminosäuren auf der Basis des BRD-Inhibitors JQ1 synthetisiert, welche potentiell als Acetyllysin-Mimikry fungieren könnten. Diese Aminosäuren wurden in von Histon H3 und H4 abgeleitete Peptidsequenzen eingefügt und in Pulldown-Experimenten auf deren Bindung von BRDs hin untersucht. Anhand dieser Versuche konnte Aminoheptansäure-ε-3-methyl-1,2,4-triazol (ApmTri) als wirkungsvollstes Acetyllysin-Mimikry für BRDs der BET-Familie identifiziert werden. Pulldown-Assays und Dissoziationskonstanten (K_Ds) zeigten auf, dass die BET-BRDs BRD3(2), BRD4(2) und BRD4(1) ApmTri-Peptidsubstrate mit vergleichbarer Affinität wie ihre nativen Acetyllysin-Substrate binden können. Es wurden Kristallstrukturen von BRD3(2) und BRD4(1) mit gebundenen ApmTri-Substraten gelöst, welche zeigten, dass die Triazolstruktur das Nε-Amid von Acetyllysin erfolgreich imitierte. Im Folgenden wurde die Stabilität der nichtkanonischen Aminosäure ApmTri gegenüber HDACs beobachtet, was die genetische Kodierung und den Einbau von ApmTri in Proteine in HEK-Zellen ermöglichte. Die Co-Expression von BRD3(2) und Fusionsproteinen mit inkorporiertem ApmTri erlaubte die gemeinsame Präzipitation beider Interaktionspartner aus nativem HEK-Zelllysat.

Die bereits etablierte HDAC-bindende Aminosäure AsuHd wurde ebenfalls genetisch codiert und somit ein neues Werkzeug für die Erforschung von HDACs erhalten. Mithilfe von rekombinanten, AsuHd enthaltenden Proteinen konnten endogene HDACs aus HEK-Zelllysat co-immunopräzipitiert werden. Das Potenzial von AsuHd-basierten Proteinsonden wurde ebenfalls für die Generierung von vollständigen Nukleosomen-Grundpartikeln (NCPs) genutzt, indem mithilfe von Sortase-vermittelter Ligation synthetische AsuHd-H3-Peptide an die NCP-Struktur ligiert wurden. Diese Designer-NCPs waren in der Lage, HDAC-Komplexe mit höherer Spezifität und Effektivität zu rekrutieren als die entsprechenden Peptid-basierten AsuHd-Sonden. Eine massenspektrometrische Analyse des Interaktoms von AsuHd-NCPs zeigte die erfolgreiche Anreicherung von HDAC 1, 2, 6 und weiterer 38 neuer, potentieller HDAC-Interaktionspartner.

Contents

List of Abbreviations

9-BBN ..... 9-Borabi-cyclo[3.3.1]nonane
Aad ..... 2-Aminoadipic acid
Ac ..... Acetyl
ACN ..... Acetonitrile
Ahx ..... 6-Aminohexanoic acid
Amp ..... Ampicillin
Apm ..... 2-Aminoheptanedioic acid
ApmTri ..... Apm(3-methyl-1,2,4-triazole)
APS ..... Ammonium persulfate
Asu ..... 2-Aminosuberic acid
AsuHd ..... α-Aminosuberic ζ-hydroxamic acid
BCA ..... Bicinchoninic acid
BET ..... Bromodomain and extra-terminal domain
BFP ..... Blue fluorescent protein
Boc ..... *tert*-Butyloxycarbonyl
BocK ..... *N*ε-Boc-L-Lysine
BRD ..... Bromodomain
BSA ..... Bovine serum albumin
BV ..... Bed volume
CFP ..... Cyan fluorescent protein
CHAPS ..... 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CoIP ..... Co-immunoprecipitation
COMU ..... (1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate
CoP ..... Co-precipitation
CREB ..... cAMP response element-binding protein
CREBBP ..... CREB-binding protein
CTCL ..... Cutaneous T-cell lymphoma
Cy5 ..... Cyanine 5
DCM ..... Dichlormethane
DIC ..... Diisopropylcarbodiimid
DIPEA ..... *N, N*-Diisopropylethylamine
DMAP 4-Dimethylaminopyridine
DMF ..... *N, N*-Dimethylformamide
DMSO ..... Dimethylsulfoxide
DNA ..... Deoxyribonucleic acid
DTT ..... Dithiothreitol
EDTA ..... Ethylenediaminetetraacetic acid
Em. ..... Emission

EPL ..... Expressed protein ligation
eq ..... Equivalents
Ex. ..... Excitation
FA ..... Formic acid
FDA ..... U.S. Food and Drug Administration
Fmoc ..... Fluorenylmethoxycarbonyl
Fmoc-OSu .. *N*-(9-Fluorenylmethoxycarbonyloxy)succinimid
FRET ..... Fluorescence resonance energy transfer
GFP ..... Green fluorescent protein
GMFI ..... Geometric mean fluorescence intensity
HAc ..... Acetic acid
HAT ..... Histone acetyltransferase
HATU ..... 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HBTU ..... 3-[Bis(dimethylamino)methyliumyl]-3*H*-benzotriazol-1-oxide hexafluorophosphate
HCCA ..... α-Cyano-4-hydroxycinnamic acid
HDAC ..... Histone deacetylase
HDACi ..... Histone deacetylase inhibitor
HEPES ..... 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMT ..... Histone methyltransferase
HPLC ..... High-performance liquid chromatography
iBET ..... Bromodomain and extra-terminal domain inhibitor
IPTG ..... Isopropyl β-D-1-thiogalactopyranoside
ivDde ..... 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)isovaleryl
K_D ..... Dissociation constant
K_i ..... Inhibition constant
Kan ..... Kanamycine
KDM ..... Lysine demethylase
LB ..... Lysogeny broth
LC-MS ..... Liquid chromatography mass spectrometry
LFQ ..... Label-free quantification
m/z ..... Mass-to-charge ratio
MALDI-TOF .. Matrix-assisted laser desorption/ionization-time-of-flight
MS ..... Mass spectrometry
MST ..... Microscale thermophoresis
MTCTK ..... Methylthiocarbonyl-thialysine
Mtt 4-Methyltrityl
ncAA ..... Non-canonical amino acid
NCL ..... Native chemical ligation
NCP ..... Nucleosome core particle
NES ..... Nuclear export signal

NHEJ ..... DNA non-homologous end joining
NLS ..... Nuclear localization signal
NMM ..... *N*-Methylmorpholine
NMP *N*-Methyl-2-pyrrolidone
NMR ..... Nuclear magnetic resonance
nt ..... non-treated
o/n ..... Over night
OD ..... Optical density
PCR ..... Polymerase chain reaction
PD ..... Pull-Down
PHD ..... Plant homeodomain
ppm ..... parts per million
PTM ..... Post-translational modification
PVDF ..... Polyvinylidene difluoride
PylRS ..... Pyrrolysyl-tRNA synthetase
PyOxim ..... [Ethyl cyano(hydroxyimino)acetato-O^2]tri-1-pyrrolidinylphosphonium hexafluorophosphate
RFP ..... Red fluorescent protein
RNA ..... Ribonucleic acid
RP-HPLC ..... Reversed-phase HPLC
rt ..... Room temperature
s.d. ..... Standard deviation
SDS ..... Sodium dodecyl sulfate
SDS-PAGE ... Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SML ..... Sortase-mediated ligation
SPPS ..... Solid-phase peptide synthesis
SrtA ..... Sortase A
$t_{ret.}$ ..... Retention time
tBu ..... *tert*-Butyl
TEMED ..... Tetramethylethylenediamine
TEV ..... Tobacco Etch Virus
TFA ..... Trifluoracetic acid
THF ..... Tetrahydrofuran
Thr ..... Thrombin cleavage site
TIPS ..... Triisopropylsilane
Tris ..... Tris(hydroxymethyl)aminomethane
Trt ..... Trityl
Tub ..... Tubulin
YFP ..... Yellow fluorescent protein
Znf ..... Zinc finger domain

1 Introduction

1.1 Chromatin structure as storage of genetic information in eukaryotes

The haploid human genome comprises of 23 chromosomes with over 3 billion base pairs (bps) of DNA including an estimated number of 19,000-20,000 protein-coding genes [1]. With a length of approximately 3.4 Å for each base pair, the DNA strand of a human diploid cell can be assumed to be 2.07 meters in length [2]. Considering this and the size of an eukaryotic cell nucleus, the genome requires storage in a compact and condensed form, realized in a DNA-protein complex called chromatin.

1.1.1 Chromatin structure and composition

The genomes of eukaryotic cells are packaged into the chromatin structure, a highly organized and regulated complex consisting of genomic DNA, proteins and RNA. In addition to its primary function in DNA packaging, chromatin also plays an important role in the regulation of DNA accessibility and consequently in cellular processes like cell division, DNA replication, DNA repair and gene expression.
The genomic DNA is organized in different distinct substructures (see figure 1). The formation of the first condensation level requires with the histone protein family. These essential protein component of chromatin consist of the four core histones H2A, H2B, H3 and H4 and the linker histone H1. The amino acid sequences of the core histones are among the most evolutionary conserved of all eukaryotic proteins [3] and histone-like proteins have also been identified in Archaea [4, 5]. In Eubacteria, only a homolog for the linker histone H1 was found so far [6].
A DNA strand of approximately 146 base pairs is wrapped in a left-handed super-helical turn around a histone octamer, formed by two H2A-H2B dimers and a H3-H4 tetramer [7, 8], constituting the nucleosomal core particle (NCP). Nucleosomes are connected by short segments of linker DNA varying in length between species and range from 5 to 60 base pairs [9]. Multiple linked nucleosomes are forming a "beads-on-a-string" nucleosomal array with a diameter of 11 nm. Binding of the linker histone H1 to nucleosome and linker DNA leads to coiling of the "string" into a helical structure called 30 nm chromatin fibre [10]. The addition of further non-histone enzymes and scaffold proteins like topoisomerase II and ScII compacts the chromatin into further higher-order structures up to the highly condensed mitotic chromosomes [11, 12].
The packaging of eukaryotic genomic DNA into chromatin results in reduced accessibility of DNA and therefore, chromatin remodeling is required to allow crucial cellular processes like gene expression. Based on the condensation level, two different chromatin states can be distinguished: A lightly condensed chromatin form on the "beads-

on-a-string" level, containing transcriptionally active genes, is defined as euchromatin, whereas the more tightly packed 30 nm chromatin fibre is categorized as heterochromatin, containing mostly transcriptionally silent genes [13]. Coordinated and regulated conversion of chromatin between the inactive and active state is accomplished by various factors. Especially ATP-dependent chromatin remodeling complexes and histone-modifying enzymes playing key roles in these processes [14].

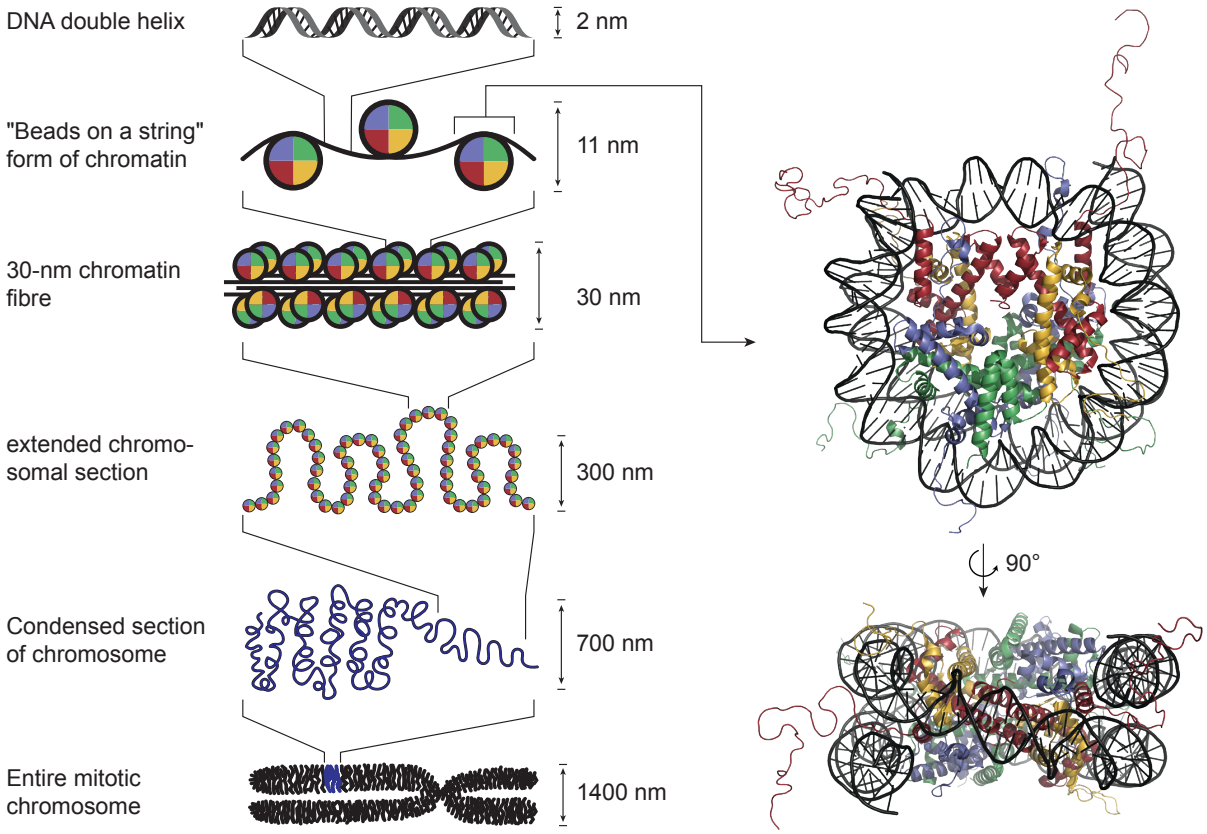


Figure 1: The organization of DNA in chromatin structure. (Left side) Overview of the different chromatin condensation levels and their approximate size (redrawn from [15]). (Right side) Crystal structure of the nucleosome core particle from *Xenopus laevis* at a resolution of 1.9 Å (PDB code: 1KX5 [16]). For side view, nucleosome was rotated by 90° around the x-axis. DNA is colored in gray, and core histones in purple (H2A), green (H2B), red (H3) and yellow (H4).

1.1.2 Histone tails and post-translational modifications

Besides the globular fold, forming the nucleosomal octamer, each of the core histones contains a highly flexible and disordered N-terminal tail, protruding from the nucleosome. Furthermore, the histones H2A and H2B also contain a C-terminal tail. The sequences of the tails are highly evolutionary conserved [17] and exhibit a high number of positively charged residues.
The amino acids of the histone tails are targeted by histone modifying enzymes introducing covalent post-translational modifications (PTMs). A large number of PTMs have

been identified on histone tails including lysine and arginine methylation, lysine acetylation, phosphorylation of serine/threonine/tyrosine residues, lysine ubiquitylation, lysine SUMOylation and ADP ribosylation [18, 19]. In terms of complexity methylation stands out because lysine residues can carry one, two or three methylation marks. Arginine residues have been identified in mono-methylation as well as symmetric and asymmetric dimethylated form [20, 21]. Additional PTMs have been reported in recent years, including various lysine Nε-acylations such as formylation, propionylation, butyrylation, crotonylation, succinylation, malonylation, 2-hydroxyisobutyrylation, glutarylation, β-hydroxybutyrylation and benzoylation [22–25]. Currently, at least 17 kinds of PTMs at over 60 different histone residues have been described [26]. A selection of PTMs and their positions on histone tails is shown in figure 2.

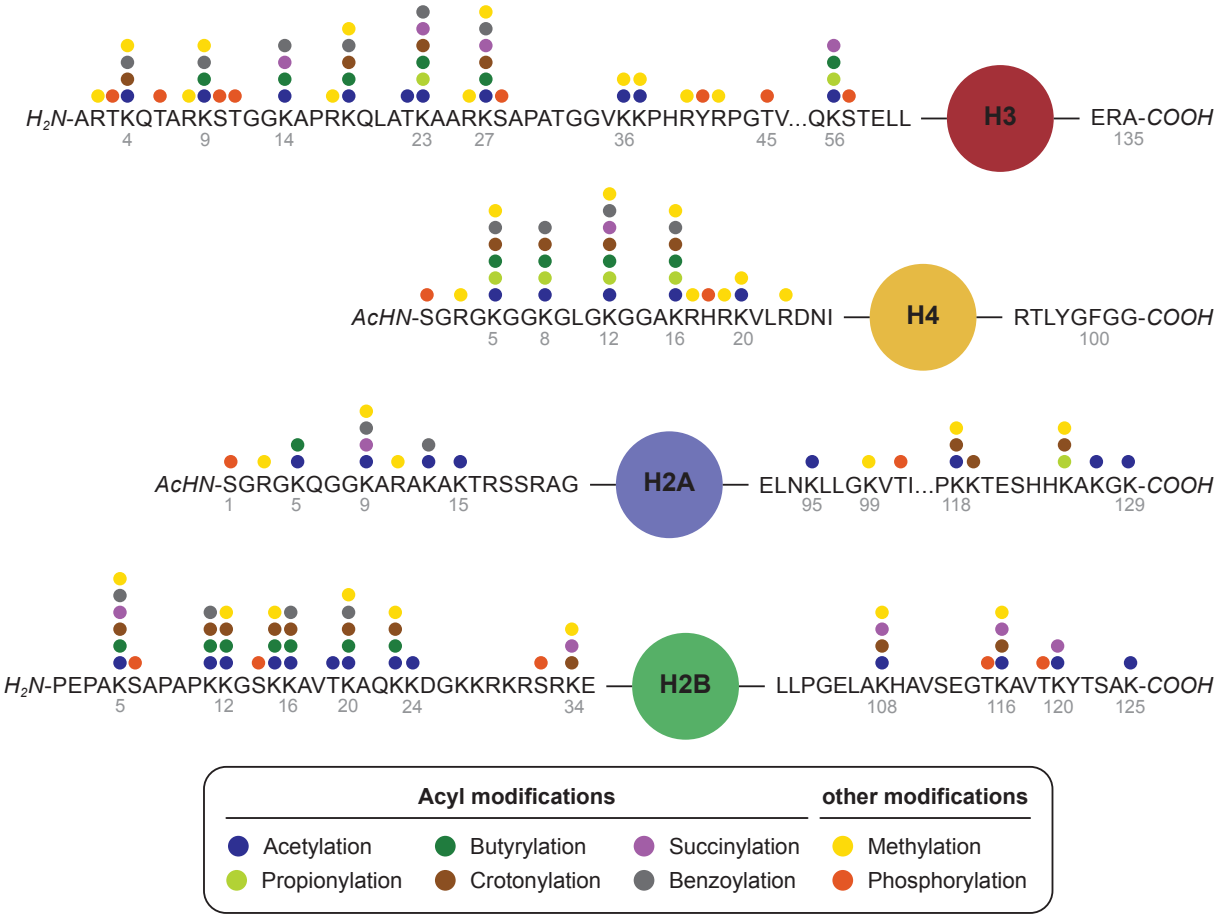


Figure 2: Post-translational modifications on core histone tails. Identified sites of selected acyl modifications, methylation and phosphorylation sites on tails of core histones H3, H4, H2A and H2B are indicated. The marked PTMs are described in [25, 27, 28].

The majority of covalent histone modifications are located on the N-terminal histone tails but were also found on the C-terminal tails of H2A, H2B and on the globular domains of core histones [29, 30].

The effects of histone modifications on chromatin remodeling and cellular processes like gene regulation have been investigated and two main mechanisms were postulated to

explain the crosstalk between histone PTMs and chromatin activity: The first theory bases on a direct impact of PTMs on the overall chromatin structure (euchromatin and heterochromatin) and the second theory focuses on the recruitment of effector molecules to histone PTMs, which modulate chromatin activity upon binding [31].
The first theory considers the charge-alteration effect of PTMs like acetylation and phosphorylation at the modification sites, influencing the dynamic electrostatic interactions between histone tails and the DNA backbone or the nucleosomal surface [32, 33]. Especially the interaction of the histone H4 tail with the nucleosomal surface is well studied. In unmodified state the highly basic H4 tail interacts with an eight acidic residue comprising cluster formed by H2A and H2B. The interaction with this so-called "acidic patch" leads to chromatin folding into 30 nm fibres and higher order structures [34–36]. The high number of known acetylation sites on the H4 tail (H4K5, H4K8, H4K12, H4K16 and H4K20) indicates that the positive charge of the tail can be effectively neutralized by hyperacetylation, causing a disruption of the H4 tail–acidic patch interaction [31]. Furthermore, this results in decondensation of chromatin and increased DNA accessibility, required for DNA replication and also leading to increased transcriptional activity [37–39].
Acetylation of H4 tails is not only influencing chromatin structure by charge alteration, but also provides recruitment sites for chromatin-associated factors containing specific binding domains. This effect is the basis of the second model, commonly referred to as "histone code theory". First proposed in 2000, the histone code theory proposes that individual or combinations of histone PTMs recruit so-called 'reader' proteins, involved in regulation of gene transcription [40, 41]. Reader proteins can effect cellular processes directly, or the binding to one modification site can trigger further modifications at adjacent residues. For example, such kind of histone crosstalk is described for phosphorylated serine 10 on the histone H3 tail (H3S10), stimulating acetylation of lysine 14 on histone H3 (H3K14) by the histone acetylase (HAT) GCN5 [42]. The modification of nearby residues can establish new binding sites for additional reader proteins, but can also prevent recognition of the site by other chromatin factors [43, 44]. The number and potential combinations of histone PTMs render the histone code hypothesis highly complex and requires deciphering of the crosstalks between PTMs and a large number of chromatin factors.
The histone code theory is an integral part of epigenetics, describing heritable changes in gene expression that do not result from alteration of the DNA sequence [45]. Besides protein and DNA modifications, non-coding RNAs, including small interfering RNAs (siRNAs), microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs) and long non-coding RNAs (lncRNAs) are reported elements of the epigenetic code, regulating the chromatin remodeling and DNA access [46].

1.2 Histone modifying enzymes

In addition to the already mentioned reader proteins for histone modifications, the enzymes depositing and removing PTMs, 'writers' and 'erasers', play important roles in establishing and interpretating the histone code.
In case of histone methylation, two different major types of histone methyltransferases (HMTs) and two main classes of histone demethylases (HDMs) are responsible for transferring and removing methyl groups at lysine and arginine residues. Mono-, di- and trimethylation of lysines is catalyzed by lysine-specific HMTs containing SET (Su(var)3-9, Enhancer-of-zeste and Trithorax) domains as catalytic unit [47]. Non-SET HMTs (Dot1-like) have been identified as well [48]. Symmetrical or asymmetrical methylation of arginine residues is catalyzed by the family of protein arginine N-methyltransferases (PRMTs). Despite differences in the catalytic mechanisms, all HMTs utilize S-adenosyl methionine as cofactor and methyl group donor [49]. Histone demethylases are divided in two main classes, defined by their mechanisms. The first family of lysine-specific demethylases (KDMs) are flavin adenine dinucleotide (FAD)-dependent amine oxidases, that oxidize the Nε-amine to the corresponding imine, followed by hydration and elimination of formaldehyde. The second family are Jumonji C (JmjC)-domain-containing histone demethylases (JHDMs), catalyzing direct oxidation of the methyl group via the cofactor α-ketoglutarate and oxygen [50–52].
Histone phosphorylation occurs on serine, threonine and tyrosine residues. This PTM is found on all four core histones and introduced by several protein kinases using adenosine triphosphate (ATP). Notable examples are the phosphorylation of H3S10 and H3S28 by Aurora B [53, 54] and H3T3 phosphorylation by the haspin protein kinase [55]. Phosphorylation levels are further controlled by various phosphatases removing the phosphoryl groups, including protein phosphatase 1, which catalyzes dephosphorylation of H3T3ph, H3S10ph and H3S28ph [56].
The set of histone modifying enzymes also comprises proteins catalyzing lysine ubiquitination [57], SUMOylation [58] or ADP-ribosylation [59]. However, only few sites of these PTMs have been identified so far. In terms of abundance, lysine acetylation stands out as the most common PTM of histones. Lysine acetylation is installed by histone acetyltransferases (HATs) and removed by the histone deacetylase (HDAC) family.

1.2.1 Histone acetyltransferases

Histone acetyltransferases are enzymes catalyzing Nε-acetylation of lysine residues on histone proteins using acetyl coenzyme A (acetyl-CoA) as co-substrate and acetyl group donor. HAT domains were found as integral part of many transcription factors. Based on structural motifs and sequence homology, HATs can be grouped into three main families [60]. The GNAT (GCN5-related N-acetyltransferase) family is named after the well characterized GCN5 enzyme in yeast. In addition to the HAT catalytic domain,

GNAT family members possesses a PTM reader domain for acetylated lysines, referred to as bromodomain [61]. Members of the MYST protein family (named after the first members MOZ, Ybf2, Sas2, and Tip60) contain an additional zinc finger domain and a chromodomain binding of methylated lysine residues. The third family - p300/CBP - is named after the histone acetyltransferase paralogs p300 and the transcription factor CREBBP (CBP), two well-studied enzymes involved in transcriptional regulation [62]. Due to high functional homology of p300 and CBP, both proteins are synonymously referred to as p300/CBP [63]. Members of the p300/CBP family contain several additional domains including multiple zinc fingers, bromodomains and contact sites for further transcription factors [64].
Some HATs also recognize other non-histone substrates. One of the most intensively studied proteins is the tumor suppressor protein p53 with several identified acetylation sites (K370, K372, K373, K381, K382). Acetylation of those sites is catalized by the transcriptional coactivators p300/CBP and p300/CBP-associated factor (PCAF) [65]. Acetylation enhances p53 transcriptional activity by directly affecting the protein conformation causing increased binding to certain response elements in target genes [66]. Lysine acetylation was also found on proteins not involved in transcription regulation. For example, lysine K40 acetylation of α-tubulin by the acetyltransferase αTAT1 is associated with stabilization of long-lived microtubules and linked to cell migration, autophagy and cell adhesion [67, 68].

1.2.2 Histone deacetylases

HDAC families

Whereas HATs are responsible for histone acetylation, histone deacetylases catalyze the opposite reaction by removing Nε-acetylations of lysine residues. So far, 18 human HDACs were identified, which can be grouped into two major families, depending on their distinct enzymatic mechanism. The first family of histone deacetylases consists of zinc-dependent "classical" HDACs (in the following referred to as HDACs), whereas members of the second family referred to as sirtuins, share a nicotinamide adenine dinucleotide (NAD^+)-dependent reaction mechanism. Based on sequence similarity to yeast prototype HDACs, the enzymes are further subdivided into class I (HDAC1, 2, 3, and 8), class IIa (HDAC4, 5, 7 and 9), class IIb (HDAC6 and 10) and class IV (HDAC11). The sirtuin proteins are categorized as class III HDACs (SIRT1-7) [69–71].
Class I HDACs show high homology to the deacetylase Rpd3 of *S. cerevisiae* and are ubiquitously expressed in mammals. Compared to the other HDAC classes, they share an N-terminal highly conserved deacetylase domain and possesses a nuclear localization signal (NLS), localizing Class I HDACs almost exclusively to the nucleus. An additional nuclear export signal (NES) of HDAC3 enables cytoplasmic localization of this enzyme as well [72, 73]. Cytosolic localization was also reported for HDAC8 in smooth muscle cells [74].

A sequence similarity of 86 % for HDAC1 and 2 indicates redundant and overlapping functions. Both enzymes are part of the same multiprotein complexes (mainly Co-REST, Sin3 and NuRD), essential for HDAC recruitment to chromatin and modulation of deacetylase activity (HDAC complexes are further described below). HDAC3 provides the catalytic domain for a multiprotein complex called N-CoR/SMRT (nuclear receptor co-repressor/silencing mediator for retinoid and thyroid hormone receptors) and is further able to associate with members of class II HDACs (HDAC4, 5 and 7) [75, 76]. HDAC8 is well studied in terms of catalytic mechanism and protein structure. Several histone and non-histone substrates (ERRα, inv(16), CREBBP) of HDACs were identified. However, there is only little information about potential multiprotein complexes of HDAC8 [77, 78].

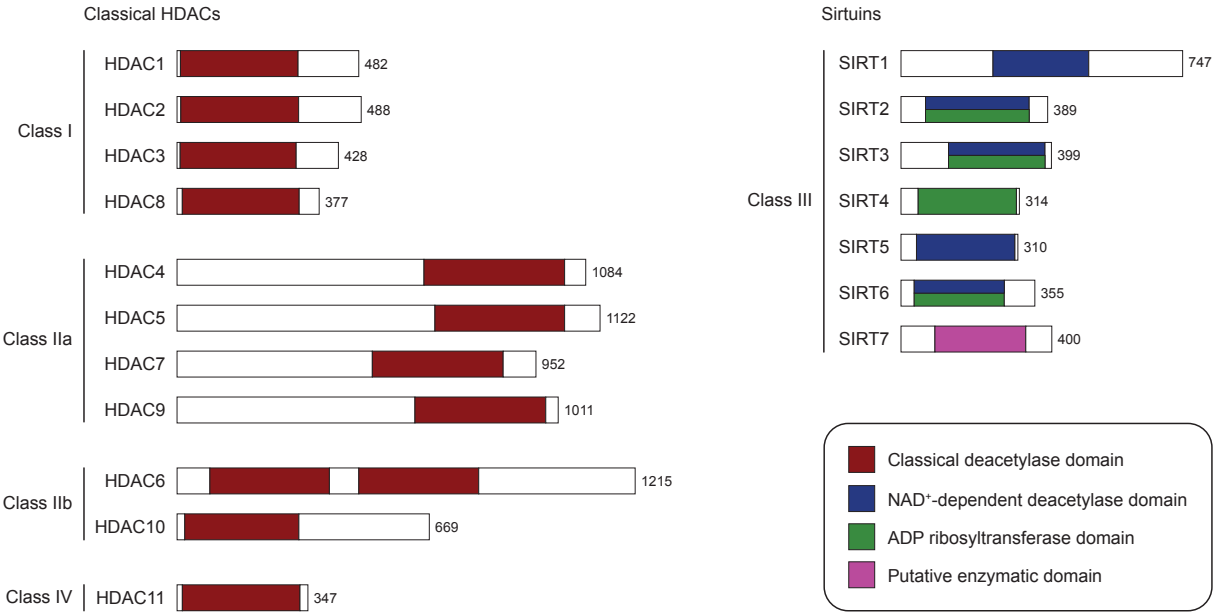


Figure 3: Classification and domain organization of classical HDACs and Sirtuins. Bar charts showing the domain organization and length of human HDACs. The longest isoform of each HDAC is illustrated. Classical deacetylase domains are colored in red, NAD$^+$-dependent deacetylase domains in blue, ADP ribosyltransferase domains in green and putative enzymatic domain in pink. Illustration is adapted from [69].

Class II HDACs show sequence similarity with deacetylase Hda1 of *S. cerevisiae*. Class IIa members (HDAC4, 5, 7, 9) consist of a conserved C-terminal deacetylase domain and a unique adapter domain at the N-terminus, providing a binding platform for the DNA-binding transcription factor MEF2 (myocyte enhancer factor 2) and multiple phosphorylation sites recruit 14-3-3 proteins. An additional NES in class II HDACs enables shuttling between nucleus and cytoplasm [79], although it was shown that HDAC6 is predominately localized in the cytoplasm [80, 81]. Class IIa HDACs are forming large multiprotein complexes with the N-CoR/SMRT complex of HDAC3 and show very low autonomous enzymatic activity, likely caused by the substitution of a conserved tyrosine residue serving as transition state stabilizer in class I HDACs. A gain-of-function mutation H976Y restores the deacetylase activity of HDAC4, questioning the relevance of deacetylase activity of class IIa HDACs [82]. Thus, specific substrates remain unknown and it is pre-

sumed, that class IIa HDACs act as adaptors of repressor complexes, explaining their crucial roles in development and differentiation processes [83].
The class IIb HDACs (HDAC6 and 10) are evolutionary related, displaying a sequence homology of 55 %. They possess N-terminal positioned deacetylase domains and an extended C-terminus with the particularity that HDAC6 contains two deacetylase domains arranged in tandem [70]. Due to its predominant cytoplasmic localization, reported HDAC6 substrates are mainly non-histone proteins such as α-tubulin [80]. The role of HDAC10 is only poorly understood. Like HDAC6, HDAC10 localizes to the nucleus and cytoplasm. The cytoplasmic localization is caused by four copies of putative leucine-rich NES sequences in the C-terminal region [84]. HDAC10 is found to interact with several other HDACs and their complexes, indicating a potential main function as recruitment platform for other proteins than as deacetylase [77, 85]. Furthermore, HDAC10 is a robust polyamine deacetylase (PDAC) with strong preference towards N^8-acetylspermidine [86].
Class III HDACs are homologous of the silent information regulator 2 (Sir2) from *S. cerevisiae* and commonly referred to as sirtuins (silent mating type information regulation 2 homolog). Seven sirtuins (SIRT1-7) have been reported in humans, sharing the NAD^+-dependent mechanism. In addition, some sirtuins may exhibit further enzymatic activities such as ADP-ribosyltransferase or deacylase activity [70]. Sirtuins are widely expressed but vary in cellular localization. Most of the sirtuins are found in the nucleus (SIRT1, SIRT2, SIRT3, SIRT6 and SIRT7). SIRT3, SIRT4 and SIRT5 localize to mitochondria and SIRT1 and SIRT2 can also be found in the cytoplasm [87, 88]. SIRT1 is among the most extensively studied sirtuins and numerous SIRT1 substrates were identified, including transcription factors like the tumor suppressor protein p53, NF-κB (nuclear factor kappa B) and PPARγ (peroxisome proliferator-activated receptor gamma) [89].
The only member of **Class IV** is HDAC11, sharing sequence similarities with yeast Hos3 and is most closely related to human HDAC3 and HDAC8. HDAC11 is localized in the nucleus but can be co-immunoprecipitated with cytoplasmic HDAC6 [90]. HDAC11 is associated with diverse regulatory functions, for example the negative regulation of interleukin-10 (IL-10) or deacetylation of the replication licensing factor CDT1 [91]. A remarkable feature of HDAC11 is the preferred deacylation of fatty acids from lysine residues, reported to be >10,000 times more efficient than deacetylase activity [92].

Catalytic mechanism of zinc-dependent HDACs

The HDACs of class I, II and IV share a common zinc-dependent enzymatic mechanism. The proposed mechanism bases on solved X-ray structures of the histone deacetylase-like protein (HDLP) from the bacterium *Aquifex aeolicus* and human HDAC8 [69, 93]. The amino acid residues, involved in the proposed mechanism (see figure 4), are conserved in all HDACs except members of class IIa, comprising a tyrosine to histidine mutation. In general, the catalytic mechanism is similar to metalloproteases. In HDACs the zinc(II) ion activates an adjacent water molecule, promoting the nucleophilic attack at the

metal-coordinated carbonyl group of the acetyl-lysine substrate. A neutral histidine residue serves as acceptor of the free proton. In the next step, the proton is transferred to the amide nitrogen atom facilitating the cleavage of the amide bond. The oxyanion tetrahedral intermediate is presumably stabilized by the zinc ion and the tyrosine residue absent in class IIa HDACs. After release of the lysine residue and the formed acetate, the active site is capable of binding new water molecules and the deacetylation cycle is repeated [69, 94].

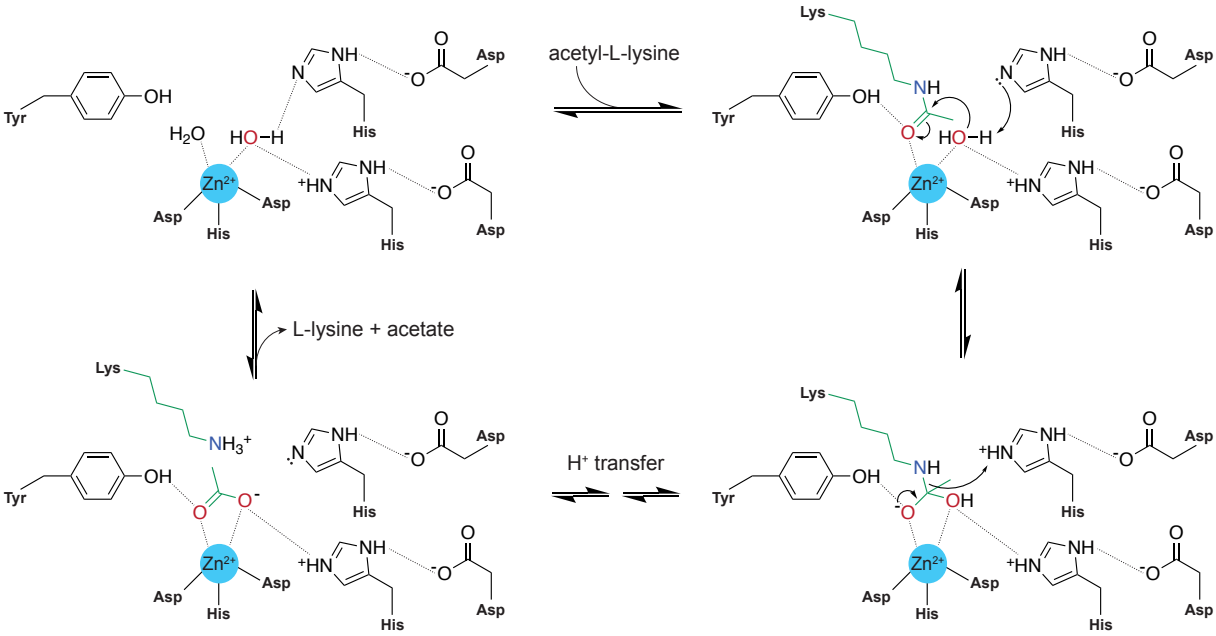


Figure 4: Proposed general mechanism of class I, II and IV HDACs. A water molecule in the active site is activated by the zinc ion (blue), promoting the nucleophilic attack at the metal-coordinated carbonyl group of the acetyl-L-lysine substrate (green). The resulting oxyanion (relevant oxygen atoms are colored in red) of the tetrahedral intermediate is stabilized by metal coordination and the proton required for decomposition is provided by the system. For simplification, only involved amino acid side chains are shown. Figure is adapted from [95].

Class I HDAC multiprotein complexes

Regulation of chromatin remodeling and activity requires an orchestrated interplay of several types of proteins including PTM readers, DNA binding domains or proteins carrying additional regulatory domains that interact with chromatin in order to execute their functions. Chromatin recruitment, substrate recognition and catalytic activity of many HDACs is achieved by formation of multi-subunit protein complexes (figure 5). HDAC1 and HDAC2 are part of three identified main complexes, CoREST, NuRD (nucleosome remodeling deacetylase) and the transcriptional co-repressor complex Sin3, all of which are essential for HDAC recruitment to chromatin by transcription factors [96]. HDAC3 forms a stable complex with members of the N-CoR/SMRT complex, resulting in activation of deacetylase action [97]. In general, HDAC1-3 show only weak deacetylase activity as isolated enzymes. HDAC8 is the only exception of class I enzymes that is not imbedded into multiprotein complexes, showing full activity in absence of any binding partners [98].

The CoREST complex includes HDAC1 and 2, the co-repressor RE1 silencing transcription factors 1-3 (CoREST1-3), the PHD finger protein 21A (PHF21A) and a lysine-specific histone demethylase 1A (LSD1) also known as KDM1A. CoREST can also form larger complexes by interacting with the C-terminal-binding protein 1 (CtBP1) complex, consisting of the histone-lysine N-methyltransferases (EHMT1 and 2), the zinc phosphodiesterase ELAC protein 2 (ELAC2), the Ras-responsive element-binding protein 1 (RREB1) and an additional zinc finger protein 516 (ZNF516). CoREST can further associate with the ATP-dependent chromatin-remodeling complex SWI/SNF (SWItch/Sucrose Non-Fermentable) and ZNF217, a Krüppel-like zinc finger protein [96, 99, 100]. The crosstalks between different enzymatic activities of a chromatin-remodeling complex are demonstrated by a described histone H3 deacetylation event, leading to demethylation of H3K4me2 and repression of REST-responsive gene expression [101].
CoREST1, LSD1 and PHF21A are also described as components of the BRAF–histone deacetylase complex (BHC complex), assembled by the SMARCE1-related protein HMG20B (also known as BRAF35) containing a high-mobility group (HMG) domain for DNA binding in a sequence-non-specific manner. Recent studies have suggested, that HMG20B is not essential for the formation of CoREST-LSD1-HDAC protein complexes, thus the BHC complex is probably only a specific, HMG20B-containing CoREST complex variant [102–104].

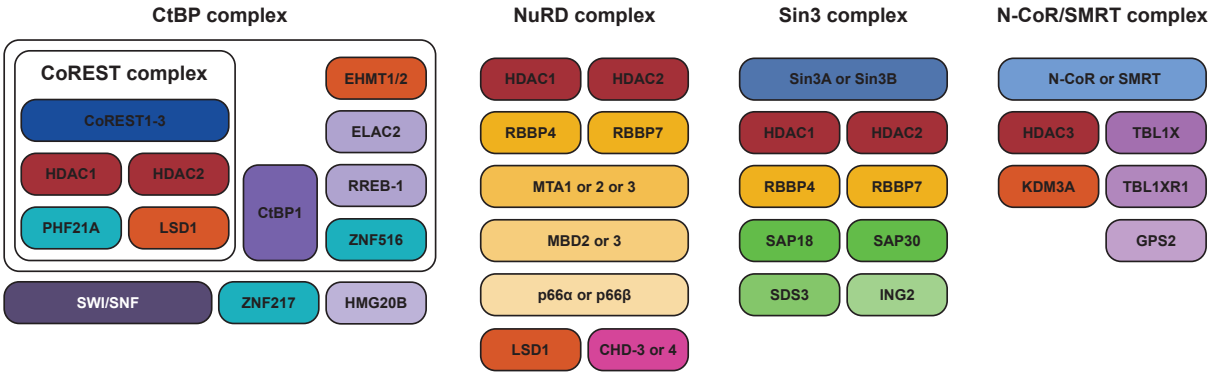


Figure 5: Schematic representation of the class I HDAC multiprotein complexes CoREST/CtBP, NuRD, Sin3 and N-CoR/SMRT. Main complex components are colored in blue, methyltransferases/demethylases in orange and zinc finger containing proteins in turquoise. The illustrated complex components are described in [96, 105, 106].

The NuRD and Sin3 complexes share a common core of HDAC1, HDAC2 and the histone-binding proteins RBBP4 and 7. In case of NuRD, this core complex is supplemented with metastasis-associated proteins MTA1, 2 or 3, methyl-CpG-binding domain proteins MBD2 or 3, transcriptional repressor p66α or p66β, chromodomain-helicase-DNA-binding proteins CHD-3 or 4 and the lysine-specific histone demethylase 1A (LSD1). In the Sin3 complex, the core proteins are complemented with either paired amphipathic helix protein Sin3A or Sin3B, histone deacetylase complex subunits SAP18 and SAP30, the Sin3 histone deacetylase co-repressor complex component (SDS3) and the inhibitor of growth protein 2 (ING2).

The HDAC3 containing N-CoR/SMRT complex consists of either N-CoR or SMRT, the F-box-like/WD repeat-containing proteins TBL1X and TBL1XR1, the G protein pathway suppressor GPS2 and the lysine-specific demethylase 3A (KDM3A) also known as JMJD2A. The mitotic deacetylase complex (MiDac) was the last multiprotein complex discovered that contains HDAC1 and HDAC2, the co-repressor protein MIDEAS (mitotic deacetylase-associated SANT domain), DNTTIP1 (deoxynucleotidyltransferase terminal-interacting protein 1) and/or the closely related TRERF1 (transcriptional-regulating factor 1) and ZNF541 (zinc finger protein 541). MIDEAS, TRERF1 and ZNF541 have similar ELM2–SANT domains for potential interactions with HDAC1 and 2 [102, 107].
The multiple components of class I HDAC complexes already indicate diverse roles in cellular processes, such as chromatin remodeling, regulation of histone PTMs, transcription regulation and DNA methylation [96, 99, 105]. However, the formation of multiprotein complexes complicates biochemical investigation of HDAC functions and activity.

1.2.3 HDAC inhibitors

In addition to physiological functions, HDACs play crucial roles in many diseases, including neurodegenerative diseases and cancer, such as cutaneous T-cell lymphoma (CTCL) and acute myeloid leukemia (AML) [108–110]. Consequently, HDACs are considered as promising drug targets. To date, a wide range of natural and synthetic HDAC inhibitors (HDACi) were developed targeting activity of all types of HDACs (pan-inhibitors) and specific HDAC types (HDAC isoform-selective inhibitors). Depending on molecular structure and functional groups, HDAC inhibitors can be divided into four main groups: hydroxamates, benzamides, aliphatic acids and cyclic tetrapeptides (figure 6). Another group of HDACis is formed by specific class III (sirtuin) inhibitors, including the pan-inhibitor nicotinamide, sirtinol as specific SIRT1 inhibitor and cambinol as specific SIRT2 inhibitor [111].
The first identified molecules with inhibitory activity towards histone deacetylases were short-chain fatty acids (SCFAs) such as sodium butyrate, valproic acid and sodium phenylbutyrate. However, with inhibitor constants (K_is) in the millimolar range, inhibitory effects are weak [112, 113]. Nevertheless, the utility of these compounds in cancer treatment was tested in clinical trials [114]. To date, valproic acid is an approved clinical drug used for treatment of patients suffering from migraines or epilepsy, but the contribution of HDACi activity to the therapeutic effects is not entirely clear [115].
HDAC inhibitors based on a hydroxamate moiety exert inhibitory effects by coordinating the zinc(II) ion in the active site of HDACs by the carbonyl and the hydroxy group. In the zinc-bound state, hydroxamic acid mimics the activated water molecule, thus simulating the proposed transition state of the enzymatic deacetylation reaction (see 1.2.2) [116]. This type of inhibitory mechanism renders hydroxamate molecules as pan-inhibitors, affecting members of class I, II and IV HDACs. The first identified hydroxamate inhibitor was trichostatin A (TSA), a natural antifungal antibiotic produced by *Streptomyces hygroscopicus* [117]. The structural related Vorinostat (suberanilohydroxamic

acid, SAHA) was the first FDA-approved HDACi utilized for treatment of refractory CTCL [118]. TSA and SAHA share three structural features: the hydroxamic acid for zinc(II) ion chelation in the active site, an aliphatic spacer reaching into the hydrophobic pocket and an aromatic cap blocking the channel to the active site [113].

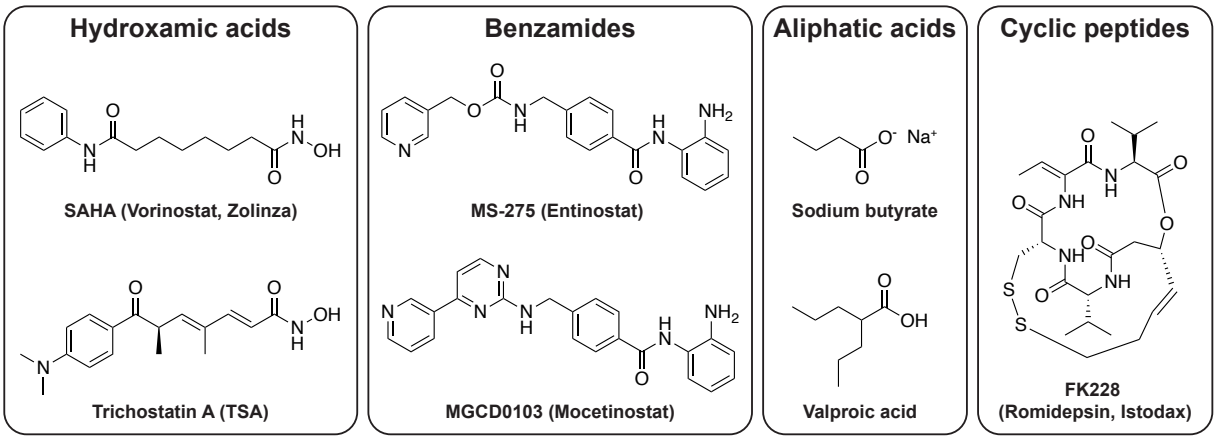


Figure 6: Chemical structures of selected HDAC inhibitors. Inhibitors are grouped by their molecular structure and functional groups.

The synthetic benzamide HDACi MS-275 (Entinostat) was reported in 1999 as a novel inhibitor type. With a K_i in the low micromolar range [119, 120], MS-275 is less potent than hydroxamate inhibitors, but much more selective for class I HDACs. Variants, for example MGCD0103 (Mocetinostat) [121] were developed in the following years and benzamidines were tested in several clinical trials for cancer treatment [122].
Various cyclic tetrapeptide HDACis were found in nature or were developed. They can be grouped by presence (trapoxin B, HC-toxin) or absence of an epoxyketone group (Apicidin, Romidepsin) [113]. Romidepsin (also known as Depsipeptide or FK-228) is produced by *Chromobacterium violaceum* and serves as prodrug that requires intracellular reduction to form a sulfhydryl group, able to bind the zinc(II) ion and the water molecule in the active site of HDACs 1 and 2 [123, 124]. Romidepsin is approved by the FDA for treatment of CTCL and other peripheral T-cell lymphomas (PTCLs) [125].

1.3 Readers of histone modifications

Besides the direct charge alteration effect, PTMs can also serve as recruitment platforms for readers domains ("histone code theory"). Proteins with reader domains often function as binding scaffold for other effector proteins or they carry writer/eraser domains, able to modulate the modification state at adjacent residues. Therefore, reader domains are found as integrated parts of various epigenetic enzymes including HATs, methyltransferases, kinases, helicases and DNA repair proteins. To date, a wide range of reader domains of different histone modifications were identified. In most cases, interactions between readers and potential binding partners were characterized by bio-

chemical and biophysical studies with modified histone peptides. This application limits the obtained information about the biological outcome in terms of chromatin structure and regulation of cellular processes [126]. The investigation of reader domains is further complicated by weak interactions between binders and short peptide probes, which might be caused by lacking the nucleosomal context. Many of the known PTM binders contain multiple reader domains, indicating that multivalent PTM recognition is an essential mechanism for enhancing binding affinities and fine-tune regulation of individual protein-protein contacts [126, 127].
As described previously (see 1.1.2), methylation is found on lysine and arginine residues in three methylation states. Unlike other modifications, the overall charge of the amino acid residues is not altered by methylation. Therefore, recruitment of effector proteins is the main activity for impacting chromatin structure. The recognition of lysine methylation sites (Kme) is achieved by a large and well-characterized group of PTM binders, including Ankyrin repeats, chromodomains, double chromodomains (DCD), chromobarrels, MBT (malignant brain tumor), PHD (plant homeodomain), PWWP (Pro-Trp-Trp-Pro), Tudor, tandem Tudor domain (TTD), WD40 and zf-CW [128]. Most of the methyl-lysine binders consist of an aromatic cage containing two to four aromatic residues, which allow, depending on the pocket size and residue composition, discrimination of different methylation states [129]. Chromo, Tudor, MBT, and PWWP domains are member of the so-called Royal superfamily, consisting of four or five β- strands, forming a characteristic β-barrel structure [130]. Several recognition sites are described for the family members, for example H3K9me2/H3K9me3 (chromodomain of HP1), H3K27me3 (chromodomain of Polycomb), H3K36me3 (Tudor domains of PHF1 and PHF19) and H4K20me (PWWP and chromobarrel) [128, 131–134]. Much less is known about methyl-arginine binders. However, some interactions with histone methylation marks are reported: The Tudor domain TDRD3 binds to H4R3me2a or H3R17me2a, the WD40 repeat of WDR5 is associated with H3R2me2s binding and the DNA methyltransferase Dnmt3a recognizes H4R3me2s but not H4R3me2a via a PHD domain. Thus, methylation of arginines at histone tails can be linked to transcriptional regulation and DNA methylation [135–137].
Recognition of phosporylated serine and tyrosine residues is mediated by binding domains including SH2, BRCT, WW, FHA, WD40, 14-3-3 and LRR. In case of phosphorylated histones, only a few specific bindings were described for phosphorylated serine (Sph) [126]. S139ph of histone H2AX is recognized by MDC1 via its tandem BRCT domain, regulating cellular responses to DNA double-strand breaks, whereas 14-3-3ζ binds to H3S10ph [138, 139].
Different reader domains were also described for ubiquitin-modified histones (ubiquitylation and SUMOylation). The DNA methyltransferase DNMT1 comprises a unique two-mono-ubiquitylation-binding module for the interaction with two mono-ubiquitylations simultaneously. This allows its recruitment to modified H3 tails as part of DNA methylation maintenance [140].
The most abundant histone PTM is Nε-acetylation of lysine residues. Acetyl groups, installed and removed by HATs and HDACs, are recognized by several reader domains.

The bromodomains (BRDs) were the first identified acetyl-lysine binders. More recently, further readers for histone acetylations were found: The sequence-specific recognition of H3K14ac is described for the dual PHD finger domain (DPF) of DPF3b. This interaction is inhibited by another PTM - H3K4me. This regulated interaction results from the interaction of PHD1, anchoring H3K14ac in the binding pocket, and PHD2, that binds the first four N-terminal histone residues while non-methylated [141]. Histone chaperone Rtt106 consists of a double pleckstrin homology (PH) domain, forming two seven-stranded β-barrel that interact with H3K56ac.
A third group of acetylation readers are the YEATS-containing proteins, named by the first five discovered proteins containing this domain (Yaf9, ENL, AF9, Taf14, and Sas5) [142]. The YEATS domain is highly conserved and occurs in more than 100 proteins from over 70 eukaryotic species. AF9, ENL, and GAS41 are the best characterized human YEATS proteins and are involved in various cellular processes. YEATS domains are also linked to cancer, since AF9 and ENL are two common fusion partners of the mixed lineage leukemia (MLL) protein [143, 144]. Beside binding to acetyl-lysine, recent investigations of YEATS domains revealed binding of crotonyl-, propionyl- and butyryl-lysine. Since binding of these Nε-acylations was even preferred over acetyl-lysine, some YEATS domains might operate as binding modules for acyl-lysines [145].

1.3.1 Bromodomains as readers of histone lysine acetylations

The bromodomain protein family

Bromodomains (BRDs) are the major protein family of lysine Nε-acetylation readers. BRDs were first identified as a novel conserved domain in the *Brahma* (*brm*) gene of *Drosophila*. Similar motifs were also found in yeast and human regulatory proteins and grouped into the bromodomain family [146]. To date, 46 diverse proteins that contain 61 BRDs were identified in the human genome. As exclusive reader domains lacking any enzymatic activity, bromodomains are present in multidomain transcriptional co-regulators and chromatin remodeling enzymes, such as HATs, ATP-dependent chromatin-remodeling complexes, helicases and methyltransferases [147]. Furthermore, BRDs are arranged in multidomain proteins with various additional protein-protein interaction modules, including diverse zinc fingers such as PHD, ZZ-type, B-box and TAZ, PWWP domains, bromo-adjacent homology domains (BAH) and KID-interacting (KIX) domains (figure 7A) [148].

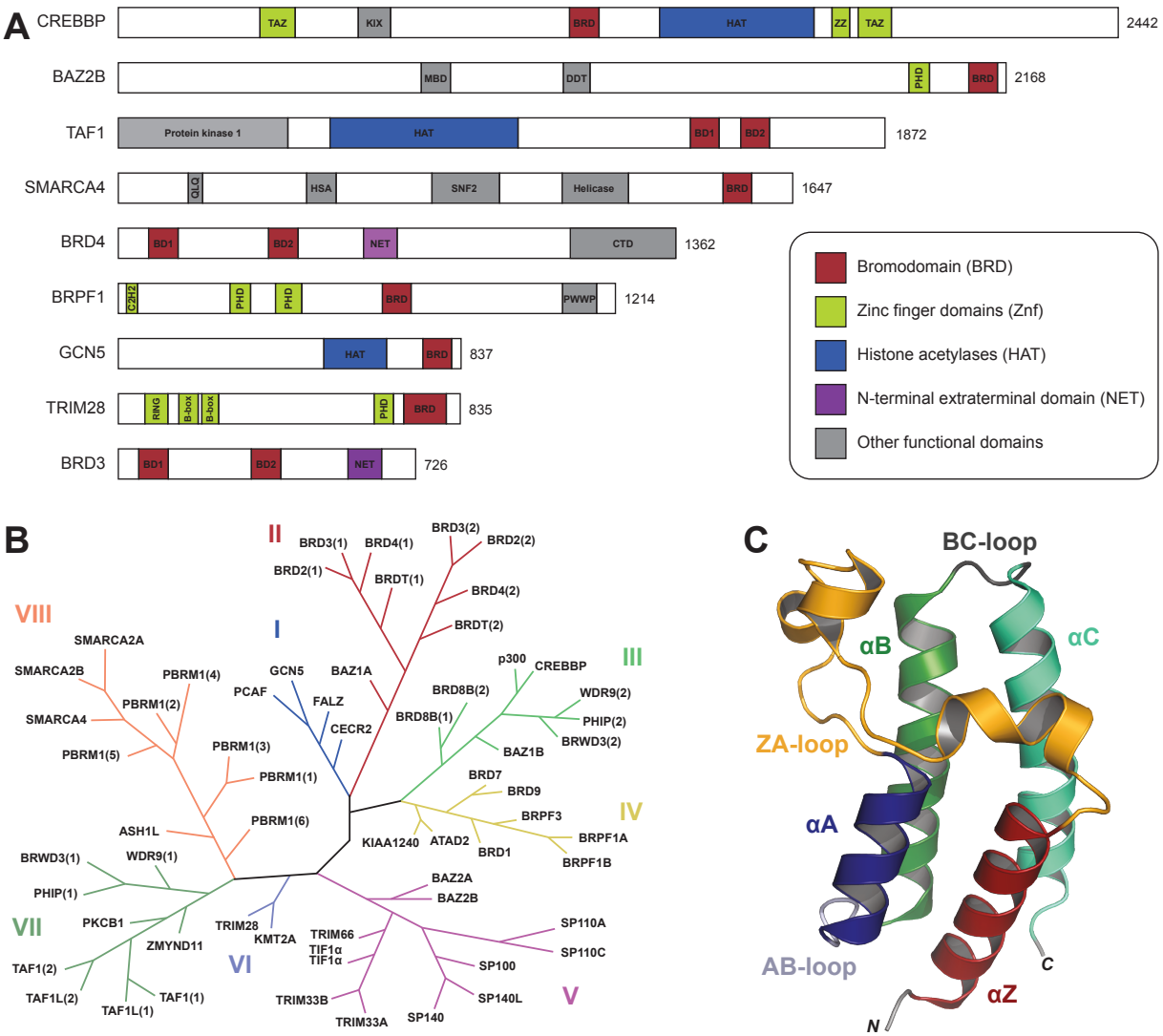


Figure 7: Overview of bromodomain protein families and their structural classification. **A** Bar charts showing the domain organization and length of selected proteins containing BRDs. BRDs are colored in red, zinc finger domains in green, HAT domains in blue, NET domains in purple and other functional domains in gray. **B** Phylogenetic tree of the human BRD family. Structural categorization bases on [149]. **C** Crystal structure of the second bromodomain of BRD3 (PDB code: 2OO1). Conserved structural elements of BRDs are labeled.

BRDs are highly evolutionary conserved protein interaction modules of about 110 residues. Despite large sequence variations, all BRDs share a conserved structural motif that comprises a left-handed bundle of four α-helices (αZ, αA, αB, αC), connected by loop segments of variable length (ZA- and BC-loop), which cover the acetyl-lysine binding site and determine binding specificity (figure 7C) [147, 149]. In most BRDs, acetyl-lysine substrates are recognized by a deep hydrophobic cavity, enclosing the aliphatic side chain of the acetyl-lysine and thereby aligning the N-acetyl group with the binding pocket. The acetyl group is anchored by a hydrogen bond between a conserved asparagine residue and the oxygen of the acetyl carbonyl group. The peptide backbone of the acetylated substrate is stabilized by a charged BRD surface at the binding site entry [150].

In general, binding affinities of BRDs and acetylated substrate sites are weak (K_D values

between 10 µM and 1 mM) toward mono-acetylated histone tail peptides *in vitro*. It is presumed, that recruitment of BRDs to acetyl-lysine motifs is supported and enforced by additional binding modules of other PTMs and DNA, present in the same protein or protein complex. It was also demonstrated, that binding affinities of some BRDs are significantly increased for di- or multi-acylated peptide substrates, indicating a role of BRDs in recognition of hyperacetylated histones [149, 151].

Subfamilies of human BRDs

Based on structural topology and sequence similarity, human BRDs were clustered into eight distinct subfamilies (subfamily I–VIII) as illustrated in the phylogenetic tree of figure 7B [149]. BRDs of subfamily II are in the focus of this thesis and will be discussed in a separate section of this chapter.
The most prominent BRD members of **subfamily I** are found in histone acetyltransferase GCN5 (also known as KAT2A or GCN5L) and the acetyltransferase containing proteins p300/CBP-associated factor (PCAF, also known as KAT2B) (see 1.2.1). Multiple binding sites for PCAF's bromodomain were identified on histone H3 (K9ac, K14ac, K36ac) as well as on histone H4 (K8ac, K16ac and K20ac) *in vitro*. The BRD of GCN5 was shown to interact with mono- and multi-acetylated peptides of histone H2A (K5ac), histone H3 (K9ac, K14ac, K9ac/K14ac) and histone H4 (K8ac/K14ac, K16ac, K5ac/K8ac/K12ac/K16ac) [147, 150, 152, 153].
BRDs of **subfamily III** are also part of HATs, in this case the paralogs E1A binding protein p300 (EP300 or p300) and the CREB-binding protein (CREBBP or CBP) are the most prominent examples (see also 1.2.1). Histone binding of p300/CBP by their BRDs is considered as means of targeting HAT activity to its substrate site on histones. Despite the high homology of both HATs, the BRDs show differences in substrate recognition. The BRD of CREBBP recognizes acetyl-lysines on histone H2B (K85ac), histone H3 (K9ac/K14ac, K36ac, K56ac) and histone H4 (K12ac, K20ac, K44ac). Non-histone substrates reported for this BRD are p53 (K382ac) and the myogenic transcription factor MyoD (K99ac, K99ac/K102ac) [154, 155]. P53 binding leads to p53-induced transcriptional activation of the cyclin-dependent kinase inhibitor p21, whereas binding of MyoD results in increased activity on muscle specific promoters. The BRD of p300 binds to histone H3 (K36ac, K56ac), histone H4 (K12ac, K20ac, K44ac) and to di-acetylated MyoD (K99ac/K102ac) [147].
The BRDs of **subfamily IV** are only poorly characterized so far. Members are participating in HAT recruitment (BRPF1-3), and chromatin remodeling complexes (BRD7, BRD9, and ATAD2A/B) [156]. The bromodomain and PHD finger-containing protein 1 (BRPF1, also known as Peregrin) consists of several zinc finger domains (two C2H2, two PDH) and a PWWP domain for binding of methylated lysine residues in addition to the BRD. BRPF1 serves as multivalent scaffold subunit of the HAT complexes MOZ/MORF and HBO1 [157], responsible for guiding and activation of the HATs. BRPF1 plays critical roles in different developmental processes including gene expression regulation as described for the *Hox* gene, where MOZ-dependent acetylation marks the gene for maintained ex-

pression throughout vertebrate development [158, 159].
Notable members of **subfamily V** are the bromodomain adjacent to zinc finger domain 2A and 2B proteins (BAZ2A and BAZ2B), which comprise a PHD–BRD tandem module at their C-termini. BAZ2A is involved in the formation of the chromatin remodeling complex NoRC (nucleolar remodeling complex) by binding of the ATPase SNF2h (sucrose nonfermenting protein 2 homolog), which mediates ribosomal DNA (rDNA) silencing [160]. Recent studies revealed that the PHD domain of BAZ2A enables binding to helical folded H3 tails, potentially expanding epigenetic regulation by an additional structural level [161].
Subfamily VI only consists of the histone-lysine N-methyltransferase 2A (KMT2A, also known as MLL1) and the transcription intermediary factor 1-β (TRIM28, also known as TIF1β). Interestingly, the BRDs of both proteins do not show any binding of acetyl-lysine residues, however the BRD of KMT2A modulates the function of an adjacent PHD finger domain, enhancing the affinity of the PHD finger for H3K4me3 by about 20-fold [162, 163].
Among the members of **subfamily VII**, the best characterized BRDs in terms of binding preferences to histone tails are found in the transcription initiation factor TFIID subunit 1 (TAF1). TAF1 possesses a tandem BRD module, able to cooperatively bind multiple acetylated histone H4 peptides (K8ac/K16ac, K5ac/K12ac, K5ac/K8ac/K12ac/K16ac) with enhanced affinity when compared to mono-acetylated H4 peptides (K8ac, K12ac, K16ac) [147, 164]. TAF1 is the largest subunit of the transcription factor IID (TFIID) complex, involved in the recruitment of further TAF proteins to promoter regions [165]. TAF1 also recognizes di-acetylated p53 (K373ac/K382ac), resulting in recruitment of TFIID subunits to the p21 promoter [166].
BRDs of the **subfamily VIII** are found in the histone-lysine N-methyltransferase ASH1L and in the chromatin remodeling factors SWI/SNF-related matrix-associated actin-dependent regulator of chromatin A2 and A4 (SMARCA2, SMARCA4). The SMARCA proteins are components of the SNF/SWI complex and displaying helicase and ATPase activities. Various binding sites on histone tails are described for BRDs of SMARCA2 (H3K9ac, H3K14ac, H4K8ac, H4K12ac, H4K16ac) and SMARCA4 (H2BK5ac, H3K14ac, H4K8ac, H4K12ac, H4K16ac) [147].

Subfamily II: The BET protein family

This subfamily contains the bromo and extra terminal (BET) proteins, to date, the most intensively studied BRD containing proteins. The four bromodomain-containing protein members (BRD2, BRD3, BRD4, BRDT) share a common domain organization with two conserved N-terminal tandem bromodomains (BD1 and BD2), an extra-terminal (ET) domain and a C-terminal domain (CTD). The CTD is only found in BRD4 and BRDT, enabling the recruitment of the positive transcription elongation factor (P-TEFb) to transcriptional start sites during mitosis, resulting in phosphorylation of serine residues of the RNA polymerase II [167, 168]. While the tandem BRDs are responsible for acetyl-lysine recognition,

the ET domain is responsible for the recruitment of specific effectors for transcriptional regulation, such as the histone-lysine N-methyltransferase NSD3 [169].
With exception of the testis-specific BRDT, BET proteins are ubiquitously expressed and localized in the cell nucleus. However, functional differences between the proteins are not well understood [170]. Possible target sites are primarily discovered by *in vitro* experiments, where the BET proteins showed preferred binding of multiple acetylated peptides with significantly increased binding affinities compared to mono-acetylated substrates. K_D values varied between 1 and 10 µM and were determined for tetra-acetylated histone H4 peptides (K5ac/K8ac/K12ac/K16ac) with single or tandem BRDs. Mono-acetylated histone H4 peptides (K5ac, K8ac, K12ac, K16ac) were also recognized by BET proteins and BD2 showed general higher affinities than BD1. Binding of acetylated histone H3 (K9ac, K14ac, K18ac) was detected as well, but the affinity was low when compared to the H4 substrates. Most effective recruitment of BRD2(1) and BRD4(1) was observed for di-acetylated histone H4 peptides (K5ac/K8ac) and crystal structures revealed, that the anchoring hydrogen bond with the conserved asparagine in the binding pocket is formed by K5ac [147].

The role of BET proteins in diseases

The biological functions of BET family members are extensively studied and multiple roles in cellular processes are described. These functions include regulation of cell growth and proliferation, inflammation, immunity, transcriptional machinery regulation, chromatin remodeling and histone modification [171, 172]. Furthermore, BET proteins are involved in the development of diseases, in particular various cancer types, inflammation, infection and autoimmune diseases [173].
In most cases, BET-driven diseases are induced by aberrant (over-)expression and mutation of BRD-containing proteins [148]. In addition, oncogenic fusions of BRD-containing proteins have been observed. BRD3 and BRD4 form abnormal fusion proteins with the *NUTM1* gene (nuclear protein in testis, *NUT*). This fusion event arises in midline structures of the human body and leads to the so-called NUT carcinoma (NC), also known as NUT midline carcinoma (NMC), which is driven by BRD4-NUT and BRD3-NUT [174, 175]. NC is a rare but highly aggressive squamous cell epithelial cancer, cytogenetically defined by a reciprocal translocation of *NUTM1* on the long arm of chromosome 15, with the *BRD4* gene on chromosome 19p13.1 (t(15;19)(q14;p13.1)) or with the *BRD3* gene on chromosome 9q34.2 (t(15;9)(q14;q34.2)), leading to the expression of BRD4-NUT or BRD3-NUT fusion proteins in NC cells [171]. The oncogenic mechanisms of those fusion proteins are not completely understood, but the recruitment of p300 to chromatin by the NUT domain of BRD4-NUT was discovered as a key event. This p300 recruitment leads to hyperacetylated chromatin regions, driving aberrant transcription of genes such as *Myc* [176]. The overexpression of *Myc* results in the undifferentiated and proliferative state of NC cells [177]. It was also shown, that BRD4-NUT is involved in p53 inactivation by disrupting p300/CBP-dependent p53 acetylation, altering transcription of p53 target genes [178].

1.3.2 Bromodomain inhibitors

The observed deregulation of BRD-containing proteins in different types of cancer renders them as therapeutic targets. The conserved and hydrophobic binding pocket of acetyl-lysine provides an attractive target for small molecule inhibitors [179]. Similar to HDAC inhibitors, bromodomain inhibitors (BRDis) alter gene expression levels, however in this case by modulating the recruitment to acetylated histones rather than acetylation levels itself.
Since the first small molecule inhibitors for BET proteins, (+)-JQ1 and I-BET762, have been described in 2010 [180, 181], numerous functional BRD inhibitors were developed and characterized. So far, described BRD inhibitors bind competitively to the acetyl-lysine binding pocket and share similar structural features, including an acetyl-lysine mimicking motif, which interacts with the conserved asparagine residue and a hydrophobic section, reaching into the hydrophobic cavity. Chemical structures of representative BRD inhibitors are shown in figure 8. The acetyl-lysine mimicking moieties cover thienotriazolodiazepines, dimethylisoxazoles, quinazolinones, thienopyridones and acetylindolizines [182].
Despite the conserved fold, the amino acid sequences of BRDs are diverse. This allows the development of chemical probes with high selectivity towards individual BRD subfamilies. Most of the reported small molecules target BRDs of the BET family (subfamily II) and are referred to as iBETs (BET inhibitors) [156]. The first described iBETs were JQ1 and I-BET762, small molecule ligands consisting of a thienotriazolodiazepine moiety, with a triazole structure serving as acetyl-lysine mimic. Binding affinities (K_Ds) below 100 nM towards both bromodomains (BD1 and BD2) of BRD3 and BRD4 were reported and JQ1 was initially tested for treatment of NUT carcinoma. Preclinical experiments demonstrated that JQ1 binds to the BRD4-NUT oncoprotein and displaces it from chromatin. Furthermore, JQ1 treatment leads to differentiation and growth arrest in NC cell lines, and also tumor regression and improved survival rates in a mouse xenograft model of NC, displaying the potential of iBETs in NC treatment [180]. In the following years, the efficacy of JQ1 was investigated in a large number of applications with BRD4 as key target. Studies included treatment of hematological malignancies, HIV infection and numerous tumors such as glioblastoma, hepatocellular carcinoma, colon cancer, pancreatic cancer and breast cancer [183]. Despite its anti-tumor activity, JQ1 has a short half-life of only 1 h and also showed a poor pharmacokinetic profile with low oral bioavailability [184]. Therefore, JQ1 is no longer used in clinical trials. Nevertheless, JQ1 is widely used in laboratory applications for the investigation of BET functions *in vivo*. To overcome the limitations of JQ1, novel inhibitors such as OTX015 (MK-8628) were developed. OTX015 is structurally similar to JQ1 but can be administered orally and is currently tested in clinical trials for the treatment of haematologic malignancies. However, severe dose-limited toxicity was observed [185].

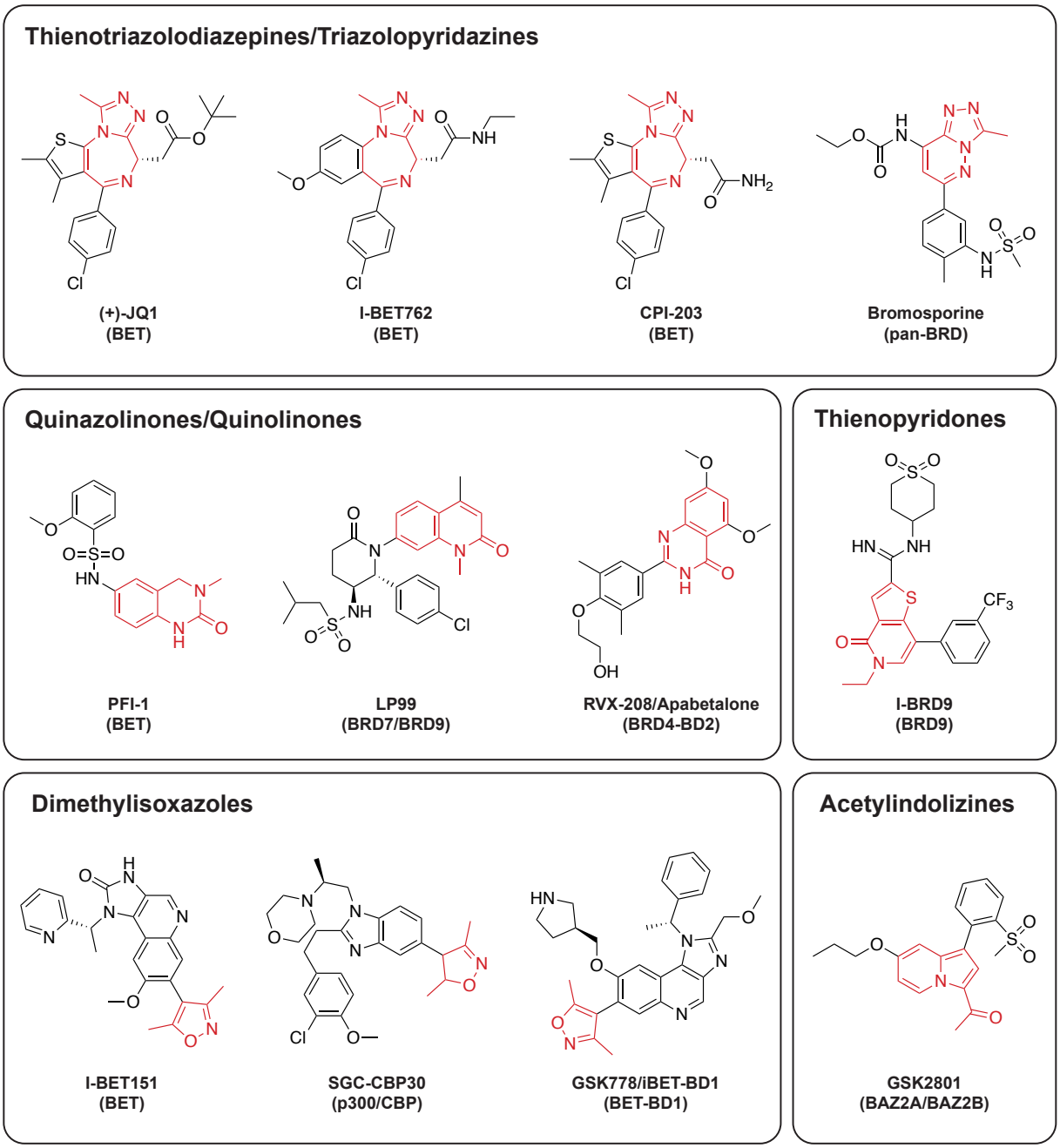


Figure 8: Chemical structures of selected small molecule bromodomain inhibitors. Inhibitors are grouped by their structural motif, responsible for acetyl-lysine mimicking (shown in red). Below inhibitor names, the target BRDs are displayed.

I-BET762 (Molibresib) shows similar affinity to BET proteins as JQ1 and downregulation of key inflammatory genes in bacteria-induced sepsis was observed during I-BET762 treatment [186]. I-BET762 shows similar efficacy as JQ1 against diverse cancer types in preclinical models, but also suffers from toxicity as observed for OTX015, indicating that these side effects might be a more fundamental problem of iBETs [184]. One approach for reducing toxicities is the specific targeting of single bromodomains (BD1 or BD2) of BET proteins. For example, the dimethylisoxazole-based GSK778 (iBET-BD1) [187] and the quinolinone-based RVX-208/Apabetalone [188] were already described as selective

ligands of BD1 or BD2, respectively. The recently reported iBETs SJ432 (BD1-selective) and ABBV-744 (BD2-selective) are showing potent anti-proliferation effects against cancer cell line models at significantly reduced doses when compared to other iBETs [189, 190].
While BET inhibitors have been extensively studied, less attention has been paid to the development and investigation of non-BET inhibitors. However, some non-BET BRD inhibitors are reported, including bromosporine as a multi-bromodomain inhibitor (pan-BRD) that targets a broad range of BRDs with nanomolar affinity [191]. I-CBP112 and SGC-CBP30 are inhibitors showing nanomolar activity and high selectivity towards the BRDs of p300/CBP (subfamily III). Due to the crucial roles of p300 and CREBBP in histone and non-histone acetylation, CREBBP inhibitors might find therapeutic applications [192, 193]. The highly homologous BRD7 and BRD9 (subfamily IV) can be targeted by the quinolone-fused lactam LP99, whereas the thienopyridone-based I-BRD9 is selective for BRD9 [194, 195]. GSK2801 is a BAZ2A/BAZ2B (subfamily V) inhibitor with K_D values of 136 and 257 nM for BAZ2B and BAZ2A, respectively. GSK2801 is used as versatile tool for studying biological functions of the BAZ2 bromodomains [196].

1.4 Chemical and biochemical tools to study PTMs

The investigation of post-translational modifications and their effects is essential for understanding cellular processes. To study the multifaceted effects of PTMs, access to proteins in a defined and homogenous modification state is required. Over the last decades, several chemical biologic and synthetic approaches have been developed, providing the tools for synthesis, identification and analysis of modified proteins [197]. This toolbox includes amino acid analogs and mimics of modified residues, protein semisynthesis, genetic code expansion and proteome-wide identification of PTM sites by mass-spectrometry-based proteomics.

1.4.1 Protein semisynthesis

Solid-phase peptide synthesis (SPPS) is nowadays a well established method for synthesis of peptides and allows the site-specific installation of single or multiple PTMs into amino acid sequences of interest. However, synthesis becomes more challenging with increased peptide length, limiting the method to peptide probes and small proteins. The production of larger proteins still requires recombinant expression and folding. To combine the advantages of SPPS with recombinant protein expression, chemoselective ligation techniques were developed to reconstitute proteins of larger size from synthetic and recombinant fragments (figure 9).

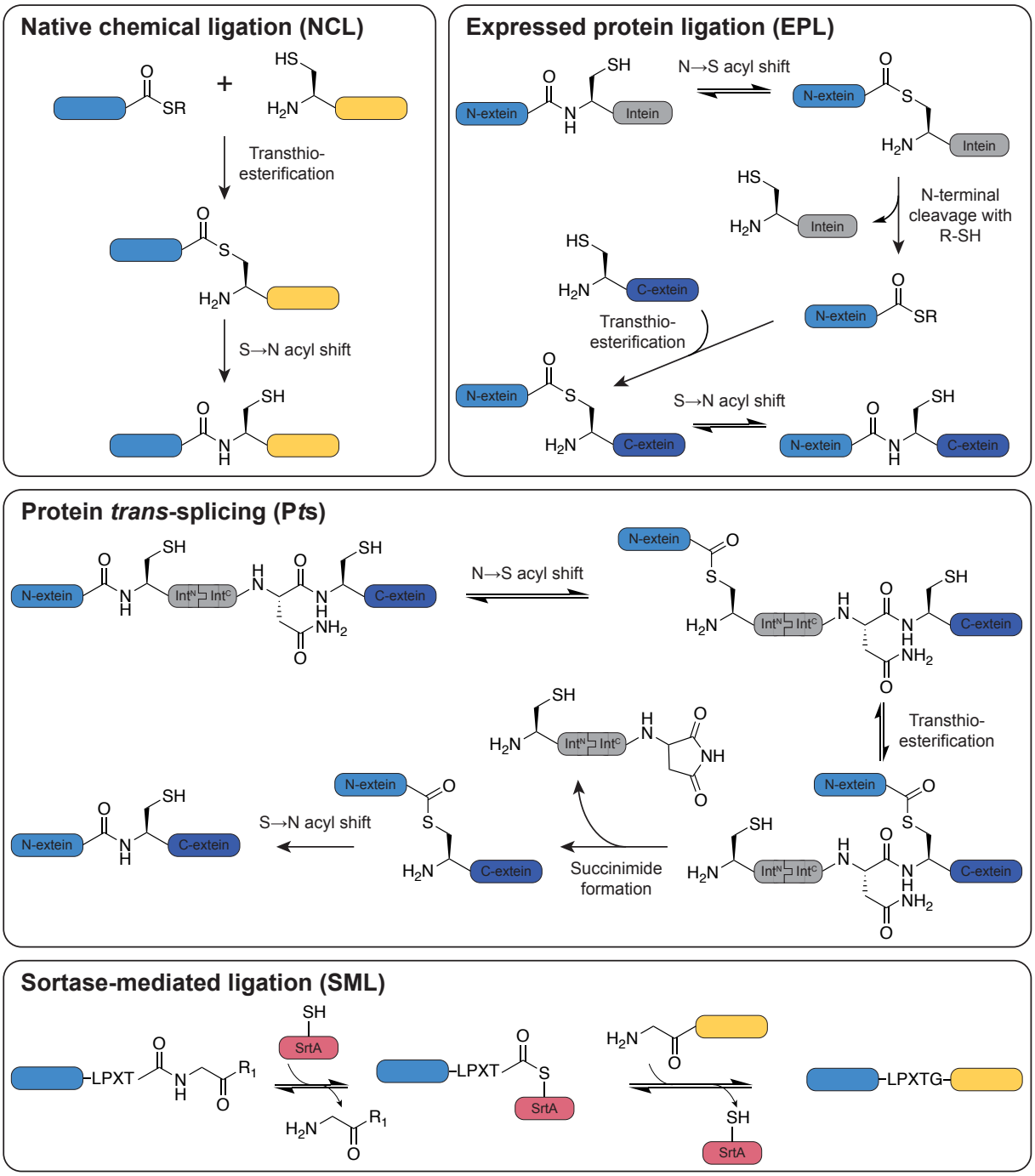


Figure 9: Schematic illustrations of selected protein semisynthesis strategies. The NCL, EPL and P*ts* strategies are used for amide-bond-forming ligations of synthetic peptides and recombinantly expressed proteins by a chemoselective reaction. SML bases on the Sortase A transpeptidase and allows ligation of LPXTG-containing sequences (X is variable) to N-terminal glycines of peptides or proteins. Figure adapted from [198].

The most widely used amide-bond-forming ligation method is native chemical ligation (NCL) [199]. NCL requires an N-terminal cysteine on one and a C-terminal thioester on the other fragment. The reaction takes place in two steps: First, a thioester intermediate is formed by transthioesterification, which is in the following irreversible rearranged by an S→N acyl shift, forming an amide bond (figure 9) [198]. Especially for the semisynthesis

of histones NCL is often the method of choice, due to the concentration of PTMs at the N-termini and the lack of cysteines in most histones. Desired modifications can be incorporated into histone tails by SPPS, whereas the globular domains are recombinantly expressed in bacteria. After successful ligation, the cysteine at the ligation site can be desulfurized to alanine, delivering the native histone [200].
The necessity of a C-terminal thioester for NCL, which has to be chemically incorporated, limits this approach to N-terminal ligations. The ligation of a synthetic C-terminus to a recombinant N-terminal fragment can be carried out by the so-called expressed protein ligation (EPL). This method exploits natural occuring internal protein sequences (inteins), that are able to excise itself out of the polypeptide and link the lateral fragments (N/C-exteins) under formation of an amide bond (protein splicing) [201]. A mutated intein is fused to the N-terminal protein segment, which catalyzes an initial N→S acyl shift. The addition of thiols releases the N-extein as a C-terminal thioester, which can undergo NCL with a protein fragment carrying an N-terminal cysteine [202].
A similar ligation approach is protein *trans*-splicing (P*t*s), which bases on split-inteins. In split-inteins, the intein domain is separated into two polypeptides (Int^N, Int^C), able to reassociate under formation of a fully functional intein. The intein fragments can be fused to exteins (N/C-extein), enabling the P*t*s reaction [203]. Depending on the size of selected split-intein fragments, the N-extein or the C-extein segment can be synthesized by SPPS. After association of Int^N and Int^C, a thioester intermediate is formed by an N→S acyl shift, followed by transesterification. Self-cyclisation of a C-terminal asparagine in the intein results in the release of a succinimide moiety, followed by amide-bond formation via an S→N acyl shift [204]. Compared to EPL, the usage of split-inteins allows the ligation of synthetic N-terminal or C-terminal fragments to recombinant segments, depending on the selected split-intein. Furthermore, the addition of a thiol reagent is not required for thioester release.
A different chemoenzymatic ligation strategy is sortase-mediated ligation (SML). Sortase A (SrtA) is a naturally occuring transpeptidase, which anchors surface proteins to the cell wall of the Gram-positive bacterium *Staphylococcus aureus* [205]. SrtA recognizes the so-called sorting motif LPXTG (where X is any amino acid) in the surface protein and cleaves the peptide bond between the threonine and glycine residue under formation of an enzyme-bound thioester intermediate. The ligation partner possesses an N-terminal glycine, serving as nucleophile for the surface protein. The free α-amino group of the glycine performs a nucleophilic attack on the thioester, forming an amide bond [206]. This ligation reaction can be used for semisynthesis approaches using two fragments of almost any size. For this purpose, the LPXTG motif is fused to the N-terminal fragments, whereas the C-terminal segment requires an N-terminal glycine. However, the ligated product also contains an LPXTG motif and furthermore, the leaving group also consists of an N-terminal glycine, thus making the SrtA reaction reversible. This reduces the overall ligation yields and limits SML application in protein semisynthesis. Several strategies were developed for making the reaction irreversible, including the usage of LPXTG-containing depsipeptides, bearing an ester bond between the threonine and

the glycine. After ligation, the N-glycolic acid of the leaving group can not serve as nucleophile for SML anymore and therefore, the reaction equilibrium is shifted towards the ligation product. An interesting application of SML is the semisynthesis of histone H3. The H3 tail contains an APATG motif (A29-G33) close to the globular domain, allowing the synthesis of H3 tails bearing an A29L mutation by SPPS, which can be ligated to recombinant Δ_{32}H3 proteins in the following [207, 208].

1.4.2 Genetic code expansion

Protein biosynthesis uses a set of 20 canonical amino acids, encoded by the triplet codons of the genetic code, which is highly conserved among all organisms. During gene expression, genetic information stored in DNA is transcribed into mRNA, which is decoded by tRNAs during translation. Transfer RNAs (tRNAs) are "charged" by aminoacyl-tRNA synthetases with their respective amino acids. Charged aminoacyl-tRNAs read the corresponding codons on mRNAs with their specific anticodon [198]. The discovery of selenocysteine (Sec) and later pyrrolysine (Pyl) extended the canonical set of amino acids to 22 genetically encoded (proteinogenic) amino acids. Whereas selenocysteine is naturally found in all kingdoms of life [209], pyrrolysine appears restricted to methanogenic archaea such as *Methanosarcina barkeri*, where it was found in the active site of methyltransferases [210]. Pyrrolysine is encoded by the amber stop codon TAG in DNA by overriding the stop function, which is enabled by usage of a specific aminoacyl-tRNA synthetase/tRNA pair named PylRS/tRNA^{Pyl}. The orthogonality of the PylRS system in other organisms allows hijacking the system for incorporating pyrrolysine during recombinant protein expression [211]. PylRS possesses a deep hydrophobic pocket accommodating the pyrrolysine side chain and the substrate specificity is mediated by only a few residues [212]. This renders PylRS an interesting target for protein engineering and directed evolution in order to encode non-canonical amino acids (ncAAs). This genetic code expansion constitutes a powerful tool for the installation of modified amino acids into recombinant proteins by either direct incorporation of PTM-bearing amino acids or by incorporation of amino acid precursors for subsequent chemical modification (figure 10).
The wild-type PylRS/tRNA^{Pyl} pair already accepts further substrates such as the protected lysine residues Nε-Boc-lysine (BocK) and Nε-Alloc-lysine (Alloc-Lys), which can be deprotected and modified chemically. Using BocK and the wild-type PylRS of *M. barkeri*, the generation of histone H3 with dimethylated K9 residue (H3K9me2) was already demonstrated [213]. However, the incorporation of most ncAAs of interest is not realizable by the wild-type PylRS, but by engineered mutants. To date, numerous PylRS mutants were engineered, accepting a wide range of ncAAs. An example is the PylRS(Y384F) mutant, which shows higher amber suppression efficiency with the ncAAs BocK and Alloc-Lys. The addition of another key mutation Y306A (PylRS^{AF}) enables the incorporation of ncAAs with larger site chains such as Nε-(o-azidobenzyloxycarbonyl)-L-lysine (AzZLys) [214]. Site-specifically installed ncAAs bearing azido or alkyne groups pro-

vide a bioorthogonal handle for conjugations by click chemistry via Staudinger ligation or azide–alkyne click chemistry [215]. For example, site-specific genetic incorporation of azidonorleucine (Anl) followed by Staudinger ligation with phosphinothioesters allowed the synthesis of histone H3 with any type of lysine acylation [216]. Besides the usage of amino acid precursors, several residues bearing native PTMs were genetically encoded directly including acetyl-lysine, phosposerine, its non-hydrolyzable analog phosphonomethylene alanine (see 1.4.3) and sulfotyrosine [217, 218].

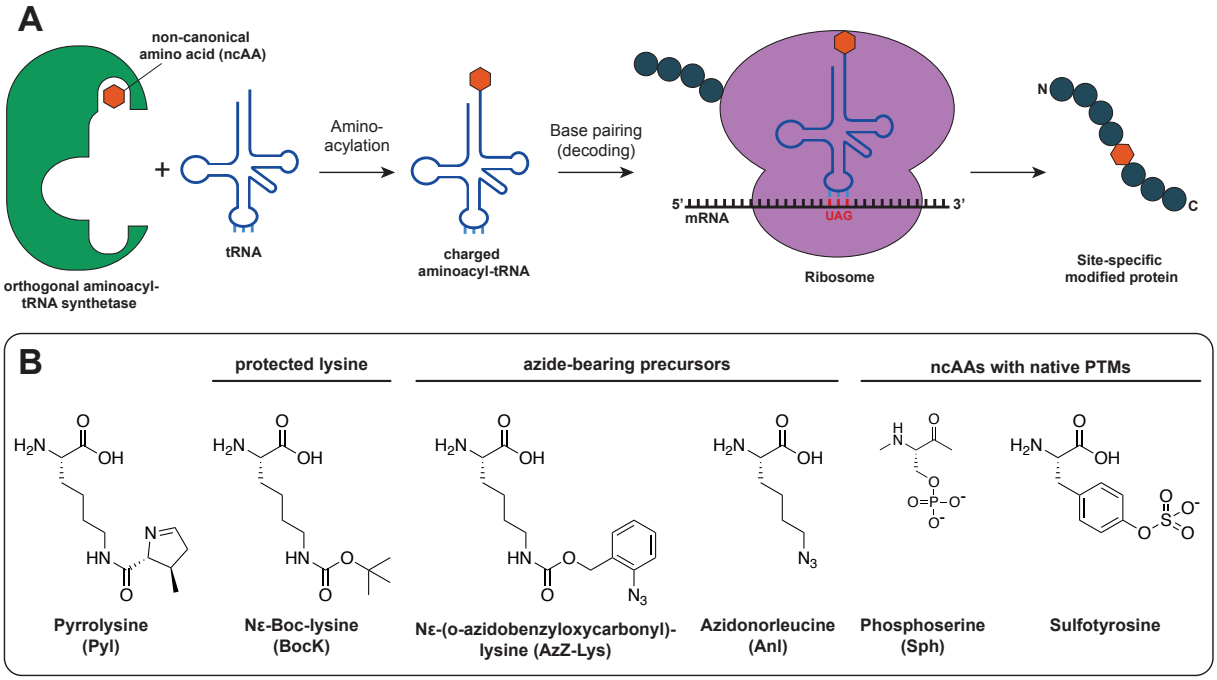


Figure 10: Genetic code expansion by an orthogonal aminoacyl-tRNA synthetase/tRNA pair. A Incorporation of non-canonical amino acids can be accomplished by usage of an orthogonal aminoacyl-tRNA synthetase, recognizing the ncAA as substrate. A tRNA bearing the anticodon (AUC) for the amber stop codon (UAG) is charged and incorporates the ncAA into proteins via ribosomal translation. **B** Examples of ncAAs that have been successfully incorporated into proteins via genetic code expansion, grouped by possible downstream applications.

Genetic code expansion is a powerful tool for the incorporation of one type of ncAAs into proteins. Current research approaches focus on the development of novel strategies for the incorporation of multiple distinct ncAAs into proteins in a single genetic code expansion approach. Those strategies include quadruplet codons, usage of non-natural DNA base pairs, synthesis of completely re-coded genomes and the generation of orthogonal translational components with reprogrammed specificities [219].

1.4.3 Investigation of PTM-based biological effects by mimics

A commonly used strategy to overcome difficulties in PTM stability and generation of site-specific modified proteins is the application of amino acid mimetics. Whereas the

already described methods are used for site-specific installation of native PTMs, the usage of mimics provides the advantage of stability towards demodifying enzymes.

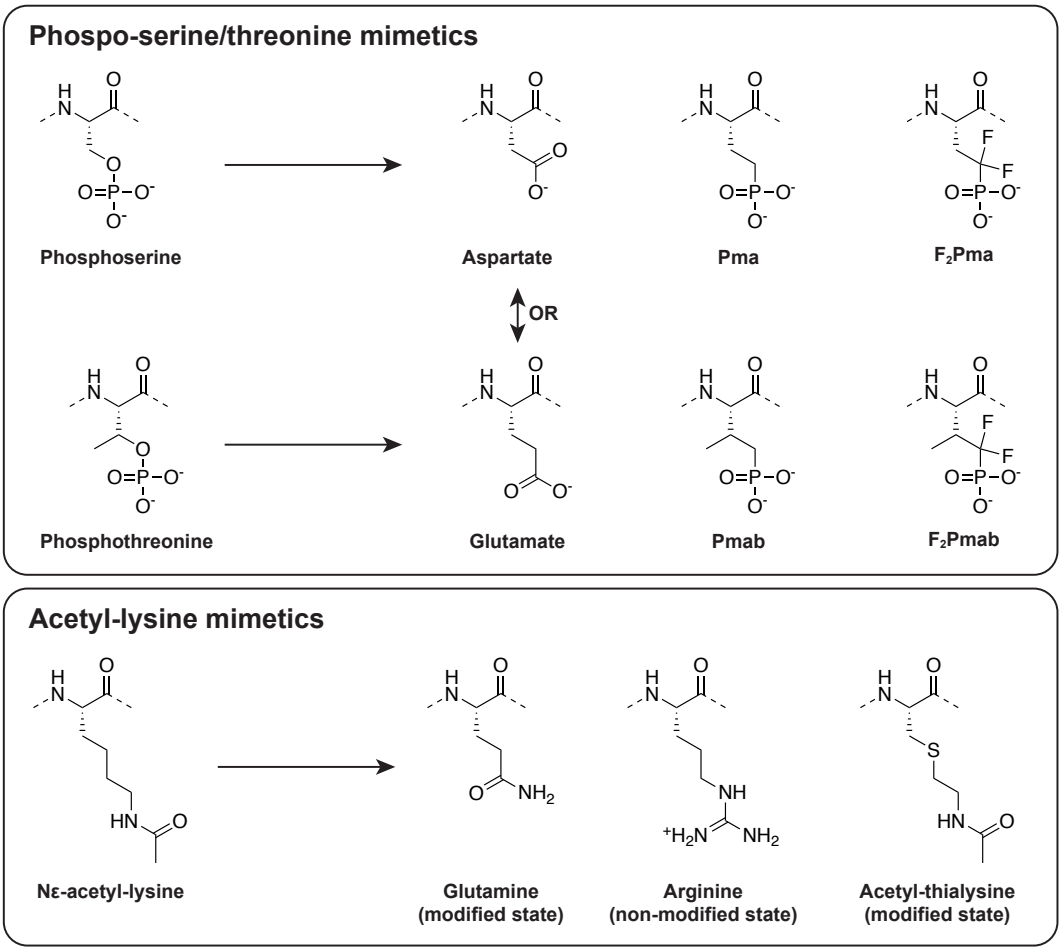


Figure 11: Overview of naturally phosphorylated and acetylated residues and their biological and chemical mimics. Phosphoserine is often imitated by incorporation of apsartate, glutamate or by the hydrolysis-stable mimic phosphonomethylene alanine (Pma) and the corresponding di-fluoro analog F_2Pma. Phosphothreonine is often imitated by aspartate/glutamate, the phosphonobutanoic acid Pmab or its di-fluoro analog F_2Pmab. Arginine and glutamine can be used to mimic unmodified and acetylated lysine, respectively. The acetyl-mimic analog can be generated by chemical modification.

Protein phosphorylation facilitates protein-protein interactions in signal transduction and modulates enzymatic activities. Consequently, several strategies for the incorporation of phosporylated amino acids were developed. Phospho-amino acids can be directly introduced during SPPS, and a semisynthetic ligation step results in the phosphorylated protein of interest. However, the native PTM is cleaved by phosphatases, limiting the utility of phospho-proteins in native lysates and living cells. A widely used and simple alternative is the substitution of the modified residue with a canonical amino acid mimic, which imitates the biophysical properties of the phospho-amino acid. The replacement of phosphoserine (Sph) and phosphothreonine (Tph) by aspartic or glutamic acid preserves the negative charge at the side chain and is the most commonly used mimic in molecular biology. The development of phosphonomethylene alanine (Pma), phos-

phonobutanoic acid (Pmab) and phosphonomethylene phenylalanine (Pmp) as non-hydrolysable substitutes of Sph, Tph and Yph offered valuable tools for the synthesis of peptides or semisynthetic proteins with potent mimics of phosphorylated residues (figure 11). Importantly, these mimics do not only preserve the negative charge at the PTM site, but also interact with the dedicated binding modules. Replacing the methylene group by difluoromethylene (F_2Pma, F_2Pmab, F_2Pmp) results in more physiological phosphorylation mimics that reflect the physiologic pK_as [220, 221].
Like phospho-amino acids, Nε-acetyl-lysine is often mimicked by canonical amino acids: Arginine serves as pseudo-deacetylated and glutamine as pseudo-acetylated lysine. The rationale behind these substitution is preservation of charge neutralization at the lysine residue. This substitution strategy is widely used, however arginine/glutamine mutations only reflect the charge states and might not recapitulate all biophysical properties of acetyl-lysine [222]. A more physiologic mimic of acetyl-lysine is acetyl-thialysine (K_SAc) (figure 11), obtained by modification of a genetically incorporated phenylselenocysteine. Dehydroalanine is generated by oxidative elimination, which can undergo a Michael addition with N-acetylcysteamine, delivering K_SAc. The replacement of the γ-carbon by a sulfur atom also changes the biophysical properties of thialysine when compared to lysine but not to the same degree as the glutamine substitution [223, 224]. The phenylselenocysteine incorporation also allows the generation of Nε-methyl-lysine analogs.

1.5 Aims of study

This work can be divided in two main parts: The first part focuses on the investigation of BRDs as readers of acetyl-lysine, whereas the second part focuses on HDACs and their role in chromatin activity.
The first objective was the optimization of a synthesis route for triazole-based amino acids as acetyl-lysine mimics. The development of novel synthesis strategies for additional triazole derivatives was also part of this project. After incorporating these amino acids into peptide-based probes derived from known BRD binding sequences, the selectivity and binding affinities of the amino acids should be characterized for various recombinant bromodomains in pull-down assays.
Most promising peptide probes, including triazole-based amino acids, can then be used for determination of binding affinities (K_D values) and co-crystallization with bromodomains.
The second part of this work focuses on the establishment of pull-down assays for endogenous HDACs using semisynthetic nucleosome core particles (NCPs). Therefore, the already established hydroxamate-based amino acid AsuHd [225–227] should be incorporated into histone H3 peptides and ligated to NCPs by sortase-mediated ligation (SML) as already described [207, 208]. In comparison to H3 peptide probes without nucleosomal context, the resulting designer nucleosomes should be applied in pull-down

experiments using western blot and proteomics analysis.
Both parts of this work are using non-canonical amino acids (ncAAs) for trapping and investigation of bromodomains (ApmTri) and HDACs (AsuHd) by pull-down experiments *in vitro*. A further objective of the study was the genetic encoding of ApmTri and AsuHD using the amber codon suppression method. Therefore, synthesis routes have to be established for accessing unprotected ncAA variants. Amber suppression should then be used for the development of interaction assays in mammalian cells by expressing ncAA-containing substrates and corresponding protein binders.

The aims of the study can be summarized as follows:

1. Triazole-based amino acids as acetyl-lysine mimics for BRDs
   (a) Synthesis of amino acids with triazole moieties as building blocks for solid-phase peptide synthesis (SPPS) and synthesis of peptide-based probes.
   (b) Cloning and expression of various recombinant bromodomain constructs and applying them in pull-down experiments with peptide probes.
   (c) Characterization of protein-peptide interactions.
2. Designer nucleosomes as probes for HDAC trapping
   (a) Synthesis of the AsuHd building block and incorporation into histone H3 tail peptides. Semisynthetic designer nucleosomes should be generated in the following by sortase-mediated ligation.
   (b) Employing histone H3 modified NCPs in pull-down experiments with mammalian cell extract, followed by western blot and proteomics analysis.
3. Genetic encoding of ApmTri and AsuHd
   (a) Synthesis of unprotected amino acids ApmTri and AsuHd.
   (b) Cloning of protein constructs for mammalian cell expression by utilizing an amber suppression system and establishing assays for investigation of the interactions between expressed recognition sites and binding proteins.

2 Results

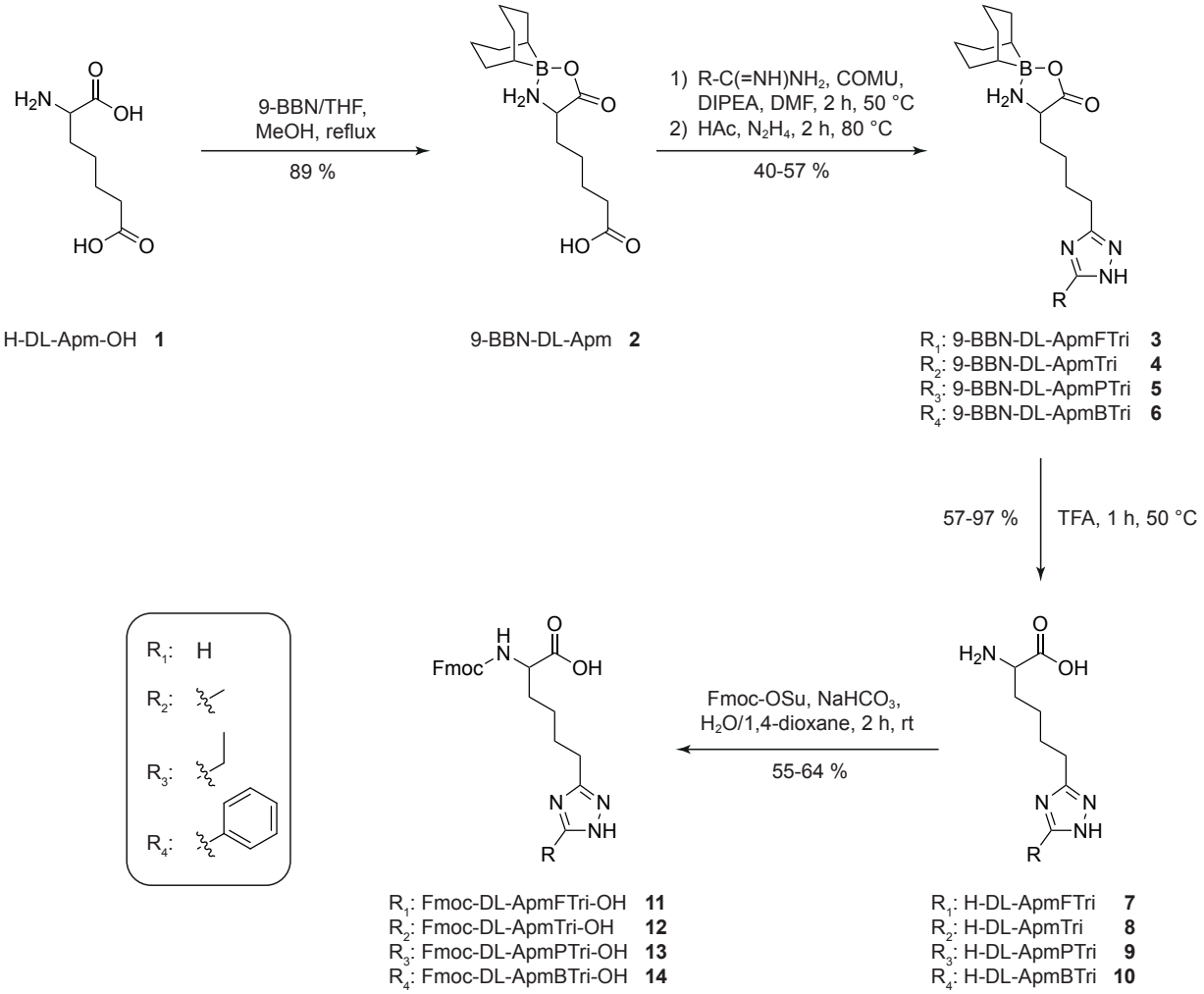
2.1 Synthesis of triazole amino acids

Small molecule inhibitors like JQ1 bind tightly to the binding pockets of bromodomains of the BET family. A triazole moiety mimicking the acetyl-lysine binding motif contributes strongly to this interaction. In the preceding master thesis, a 3-methyl-1,2,4-triazole structure was installed onto the side chain of 2-aminoheptanedioic acid (Apm) and 2-aminosuberic acid (Asu). The modified Asu amino acid was able to enrich bromodomains in pull-down experiments when imbedded into the context of a peptide sequence [228]. This finding demonstrated the potential of triazole-containing amino acids as acetyl-lysine mimics for BRDs and leads to the investigations of this thesis.
The synthesis of peptide-based probes for pull-down experiments requires building blocks compatible with solid-phase peptide synthesis (SPPS) by the Fmoc strategy. Different synthesis routes were established for accessing Nα-Fmoc protected amino acids carrying triazole moieties in the side chain, varying in length and substitution at the C-3 position of the triazole. Previous experiments already demonstrated stability and non-reactivity of the triazole moiety under SPPS conditions without further protection. All synthesized building blocks and abbreviations used in the following are listed in the appendix table A.1.

2.1.1 Synthesis of building blocks with triazole derivatives

A new synthesis route for the synthesis of a Fmoc-protected ApmTri building block was established as illustrated in scheme 1 [229]. In a first step, the α-amino and carboxy group of the racemic amino dicarboxylic acid H-DL-Apm-OH (**1**) were protected with 9-Borabi-cyclo[3.3.1]nonane (9-BBN) [230], forming 9-BBN-DL-Apm (**2**). The installation of the triazole was carried out in a one-pot reaction reported by *Castanedo et al.* [231], providing a general method for the synthesis of 1,3,5-trisubstituted 1,2,4-triazoles via the Einhorn-Brunner reaction. Different acylamidines were coupled to the Apm side chain with COMU as activation reagent, forming a primary acyl-amidine intermediate. Cyclization to the 1,2,4-triazole was achieved by addition of hydrazine and acetic acid [232]. Depending on the amidine applied in the coupling step, different C-3 substituents were introduced to the triazole. Formamidine resulted in 9-BBN-DL-ApmFTri (**3**), acetamidine in 9-BBN-DL-ApmTri (**4**), propionamidine in 9-BBN-DL-ApmPTri (**5**) and benzamidine in 9-BBN-DL-ApmBTri (**6**). After side chain modification, the 9-BBN group was removed by heating with trifluoroacetic acid (TFA), yielding the unprotected 1,2,4-triazole amino acid derivatives (**7-10**). In the last step, the α-amino group was protected with the Fmoc protection group by adding *N*-(9-Fluorenylmethoxycarbonyloxy)succinimid. The final building blocks were purified by preparative HPLC resulting in overall yields of 16 % for

Fmoc-DL-ApmFTri-OH (**11**), 23 % for Fmoc-DL-ApmTri-OH (**12**), 22 % for Fmoc-DL-ApmPTri-OH (**13**) and 19 % for Fmoc-DL-ApmBTri-OH (**14**). Figure 12 shows the LC-MS analysis of the Fmoc-DL-ApmFTri-OH (**11**) building block as representative example. All LC-MS traces of reaction products and intermediates (**2-14**) are shown in figure C.1-C.4 and NMR spectra are shown in figure C.10-C.21.



H-DL-Apm-OH **1**

9-BBN-DL-Apm **2**

R_1: 9-BBN-DL-ApmFTri **3**
R_2: 9-BBN-DL-ApmTri **4**
R_3: 9-BBN-DL-ApmPTri **5**
R_4: 9-BBN-DL-ApmBTri **6**

R_1: H-DL-ApmFTri **7**
R_2: H-DL-ApmTri **8**
R_3: H-DL-ApmPTri **9**
R_4: H-DL-ApmBTri **10**

R_1: Fmoc-DL-ApmFTri-OH **11**
R_2: Fmoc-DL-ApmTri-OH **12**
R_3: Fmoc-DL-ApmPTri-OH **13**
R_4: Fmoc-DL-ApmBTri-OH **14**

Scheme 1: Synthesis scheme of the building blocks Fmoc-DL-ApmFTri (11), Fmoc-DL-ApmTri (12), Fmoc-DL-ApmPTri (13) and Fmoc-DL-ApmBTri (14).

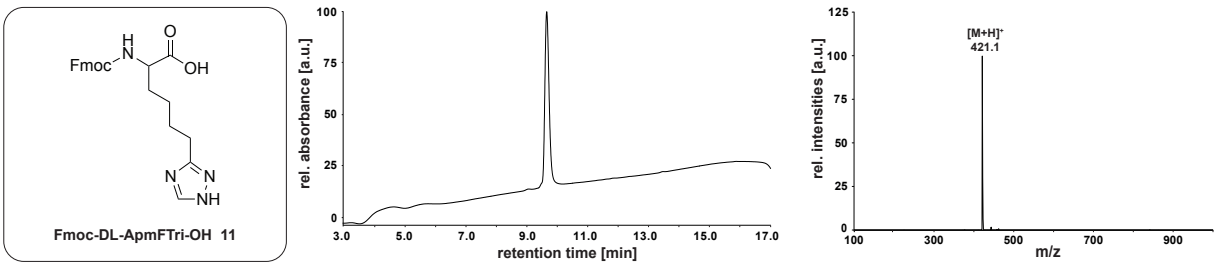


Figure 12: LC-MS analysis of building block Fmoc-DL-ApmFTri-OH (11). The building block was synthesized via scheme 1. UV absorption was measured at 218 nm. Assignable m/z values are labeled.

The side chain length of the triazole-containing amino acids were varied as well as shown in appendix scheme A.1. Following the same synthesis strategy, the building blocks Fmoc-L-AadTri-OH (**28**) and Fmoc-L-AadTri-OH (**29**) were synthesized starting from the amino dicarboxylic acids H-L-Aad-OH (**20**) and H-L-Asu-OH (**21**) with overall yields of 21 % and 29 %. LC-MS analysis of the final building block **29** is shown in figure 13 and of (**28**) in appendix figure C.6. Corresponding NMR spectra are shown in figure C.24-C.31.

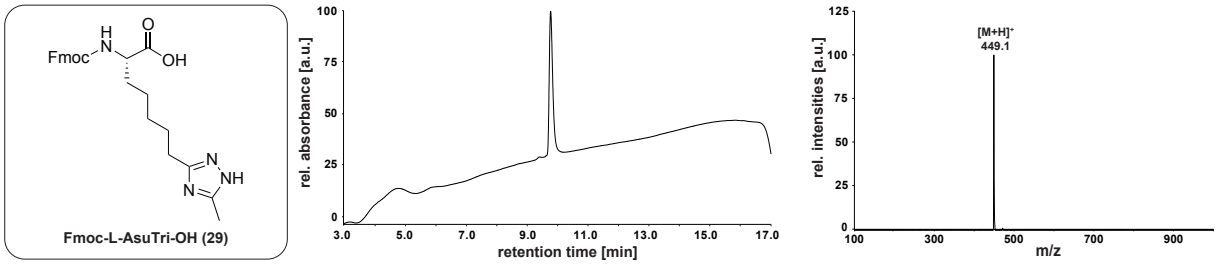
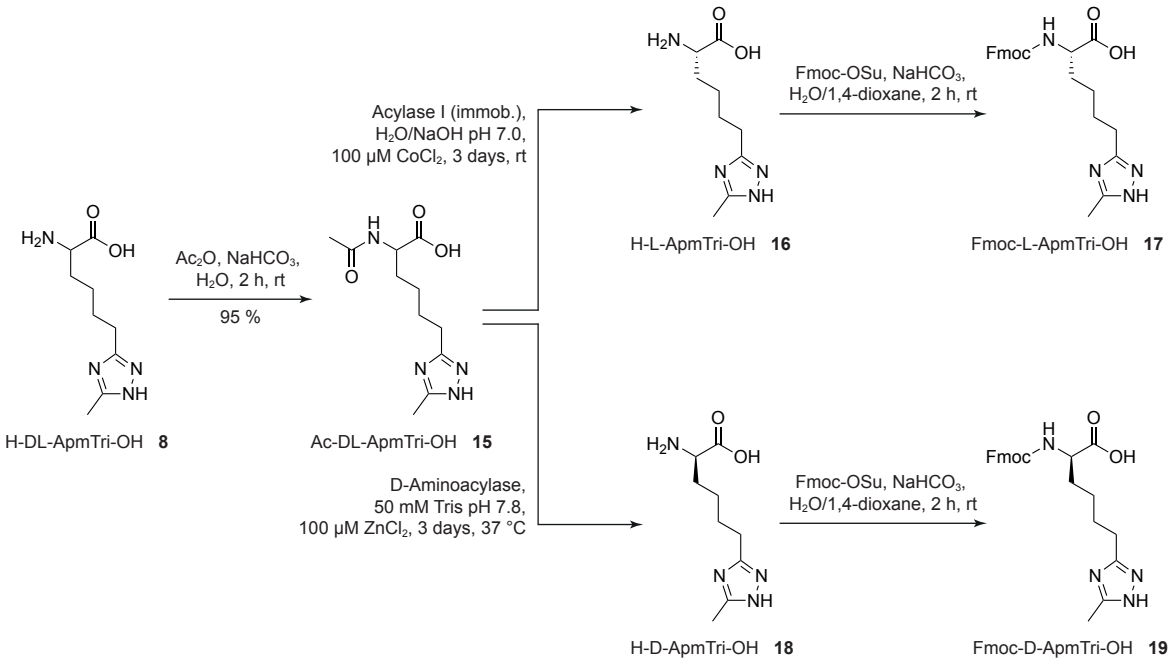


Figure 13: LC-MS analysis of building block Fmoc-AsuTri (29). The building block was synthesized via scheme A.1. UV absorption was measured at 218 nm. Assignable m/z values are labeled.

2.1.2 Synthesis of enantiopure ApmTri building blocks

Enantiopure ApmTri building blocks were generated by enzymatic chiral resolution using stereoselective aminoacylases catalyzing the hydrolysis of *N*-acetyl groups from *Nα*-acyl amino acids [233, 234]. Chiral resolution was carried out with the unprotected triazole-containing H-DL-ApmTri-OH (**8**) and with commercially available H-DL-Apm-OH (**1**) as described in 2.1.3. Following the chiral resolution, installation of the Fmoc protection group furnished the SPPS building blocks Fmoc-L-ApmTri-OH (**17**) and Fmoc-D-ApmTri-OH (**19**) as shown in scheme 2.



Scheme 2: Synthesis scheme of the building blocks Fmoc-L-ApmTri-OH (17) and Fmoc-D-ApmTri-OH (19).

Prior to the chiral resolution, the α-amino group of racemic H-DL-ApmTri-OH (**8**) (synthesized via scheme 1) was acetylated with acetic anhydride yielding Ac-DL-ApmTri-OH (**15**). Removal of the acetyl group by either Acylase I or D-Aminoacylase resulted in H-L-ApmTri-OH (**16**) and H-D-ApmTri-OH (**18**), respectively. The reactions were carried out successively. After deacetylation of the L-enantiomere, the Fmoc protecting group was installed exclusively on L-ApmTri by adding Fmoc-OSu. The Fmoc-L-ApmTri-OH building block (**17**) (yield of 70 % with respect to 50 % of **15**) and the remaining acetylated educt Ac-D-ApmTri (**15***) were separated by preparative HPLC. The residual Ac-D-ApmTri (**15***) was further treated with D-Aminoacylase. After Fmoc protection and preparative HPLC, Fmoc-D-ApmTri-OH (**19**) was obtained with a yield of 65 % with respect to **15***. Figure 14 shows the LC-MS analysis of the Fmoc-L-ApmTri-OH building block (**17**). All LC-MS traces of reaction products and intermediates of scheme 2 are shown in figure C.4. NMR spectra of **16** and **17** are shown in figure C.22 and figure C.23.

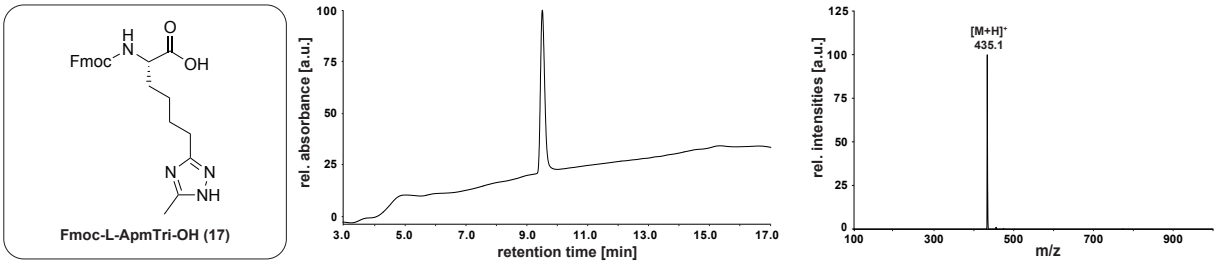
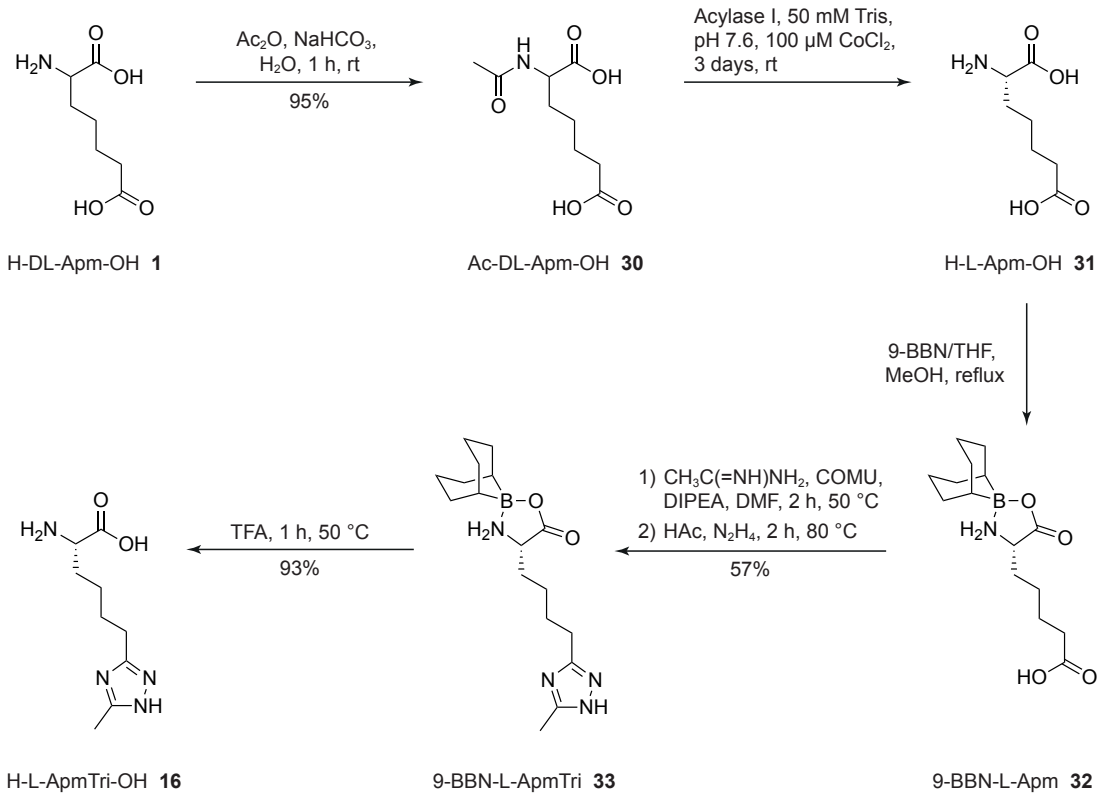


Figure 14: LC-MS analysis of building block Fmoc-L-ApmTri-OH (17). The building block was synthesized via scheme 2. UV absorption was measured at 218 nm. Assignable m/z values are labeled.

2.1.3 Synthesis of the unprotected amino acid H-L-ApmTri-OH

The application of ApmTri in amber suppression experiments required the synthesis of unprotected H-L-ApmTri-OH (**16**), which was obtained as intermediate in the synthesis route illustrated in scheme 2. Due to the challenging purification of the intermediate, the synthesis strategy was adapted as shown in scheme 3.
The chiral resolution was carried out at the level of the dicarboxylic amino acid H-DL-Apm-OH (**1**). Acetylated Ac-DL-Apm-OH (**30**) was treated with Acylase I and subsequently 9-BBN was added. Liberated H-L-Apm-OH (**31**) was protected to 9-BBN-L-Apm (**32**) and separated from the acetylated D-enantiomere. After triazole installation, the 9-BBN group was removed with TFA and extracted from the aqueous solution, yielding H-L-ApmTri-OH · TFA (**16**) with an overall yield of 39 %. The ESI-MS spectrum is shown in figure 15 and NMR spectra are illustrated in the appendix figure C.22.



Scheme 3: Synthesis scheme of the amino acid H-L-Apm(3-methyl-1,2,4-triazole)-OH (L-ApmTri) (16).

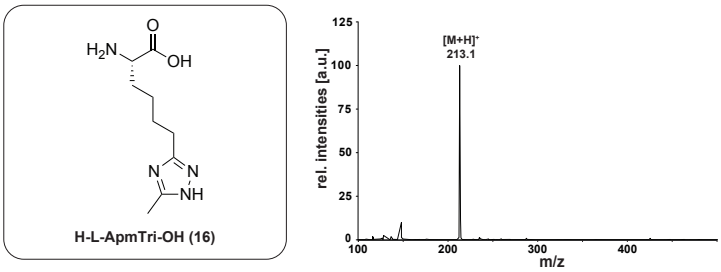


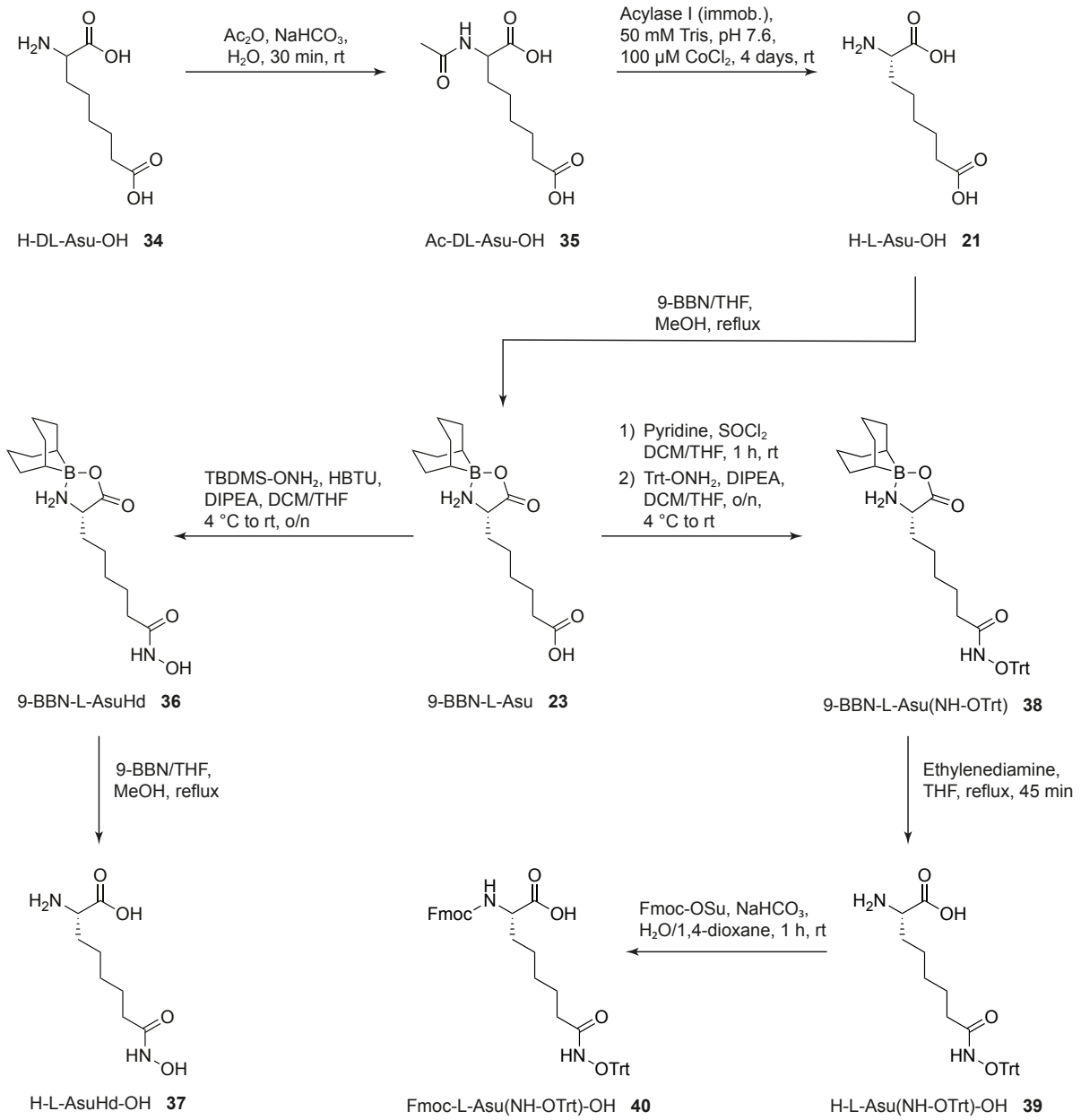
Figure 15: LC-MS analysis of synthesized amino acid H-L-ApmTri-OH (16). The building block was synthesized via scheme 3. UV absorption was measured at 218 nm. Assignable m/z values are labeled.

2.2 Synthesis of hydroxamate building block and unprotected AsuHd

The amino acid H-Asu(NH-OH)-OH (AsuHd)(**37**) and the building block Fmoc-L-Asu(NH-OTrt)-OH (**40**) were synthesized by *Annika Kühn* [235], following a protocol of *Julian Seidel* [236], which was extended by the prior chiral resolution of H-DL-Asu-OH.

The synthesis route of hydroxamate-containing amino acids is illustrated in scheme 4. In the first part of the synthesis the racemic amino dicarboxylic acid H-DL-Asu-OH (**34**)

was acetylated with acetic anhydride to Ac-DL-Asu-OH (**35**) and subsequently treated with Acylase I as described before. Deacetylated H-L-Asu-OH (**21**) was then protected with 9-BBN, resulting in 9-BBN-L-Asu (**23**) with a yield of 85 % with respect to 50 % of **34**. 9-BBN-L-Asu (**23**) was used as precursor of the unprotected H-AsuHd-OH and for the synthesis of an SPPS compatible building block.



Scheme 4: Synthesis scheme of the amino acid H-L-Asu(NH-OH)-OH (37) (L-AsuHd) and of the building block Fmoc-L-Asu(NH-OTrt)-OH (40).

Synthesis of unprotected AsuHd was carried out by coupling TBDMS-ONH_2 to 9-BBN-L-Asu (**23**) using HATU as coupling reagent. The TBDMS protection group of the hydroxamate was lost during over night reaction time, resulting in 9-BBN-L-Asu(NH-OH) (**36**). After 9-BBN deprotection by heating in TFA and subsequent extraction from the solution, H-L-

AsuHd-OH · TFA (**37**) was obtained with a yield of 75 % (starting from (**23**)). The ESI-MS analysis is shown in figure 16 and the NMR spectrum is shown in the appendix figure C.33. The SPPS-compatible protected AsuHd building block was generated by activating the carboxy group of 9-BBN-L-Asu (**23**) with thionyl chloride followed by coupling to O-tritylhydroxylamine (Trt-ONH_2), resulting in the trityl protected hydroxamic acid 9-BBN-L-Asu(NH-OTrt) (**38**). The acid lability of the trityl group required adjusted 9-BBN deprotection conditions and was achieved by heating the reaction mixture with ethylenediamine [237]. In the final step, the α-amino group was protected with Fmoc-OSu delivering Fmoc-L-Asu(NH-OH)-OH (**40**) in a yield of 57 % (starting from (**23**)). The UV chromatogram and the MS analysis are shown in figure 16. The LC-MS data of all intermediates are listed in figure C.8-C.9.

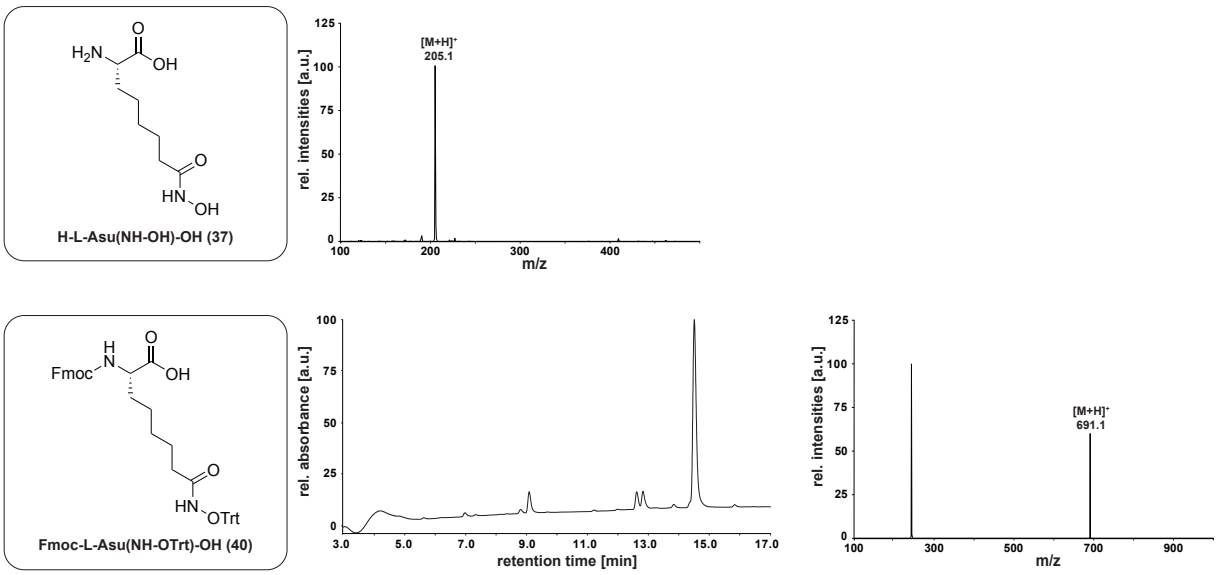


Figure 16: LC-MS analysis of synthesized amino acid H-L-AsuHd-OH (37) and of the building block Fmoc-L-Asu(NH-OTrt)-OH (40). The amino acids were synthesized via scheme 4. UV absorption was measured at 218 nm. Assignable m/z values are labeled.

2.3 Design and synthesis of peptide-based probes for pull-down experiments with bromodomains

Despite the evolutionary conserved structure of bromodomains, most BRDs exhibit high substrate specificity and recognize the sequence context adjacent to the acetyl-lysine moiety. Investigating the binding properties of different bromodomains (BAZ2B, BRD3(2), BRD4(1), BRD4(2) and CREBBP) towards peptide substrates commonly requires the synthesis of various sets of peptide probes. Acetylated tail regions of histone H3 and H4 are common binding sites of BRDs and synthetic peptides derived from these sites interact with BRDs with µM affinities [238]. Amino acid sequences of the binding sites for selected BRDs and the corresponding K_D values are listed in table 1.

Table 1: Overview of selected BRDs and their reported recognition sites. Sequences and affinity values were taken from [238]. The recognized acetyl-lysine residues are marked in bold.

BRD	Sequence		Affinity (µM)
BAZ2B	H3K14	KSTGG-**Kac**-APRKQY	7.6 ± 0.3
BRD3(2)	H4K20	AKRHR-**Kac**-VLRDNY	10.5 ± 0.2
BRD4(1)	H4K5K8	SGRG-**Kac**-GG-**Kac**-GLGK	6.8 ± 0.1
BRD4(2)	H4K20	AKRHR-**Kac**-VLRDNY	Not quantified
CREBBP	H3K56	IRRYQ-**Kac**-STELLY	13.8 ± 0.4

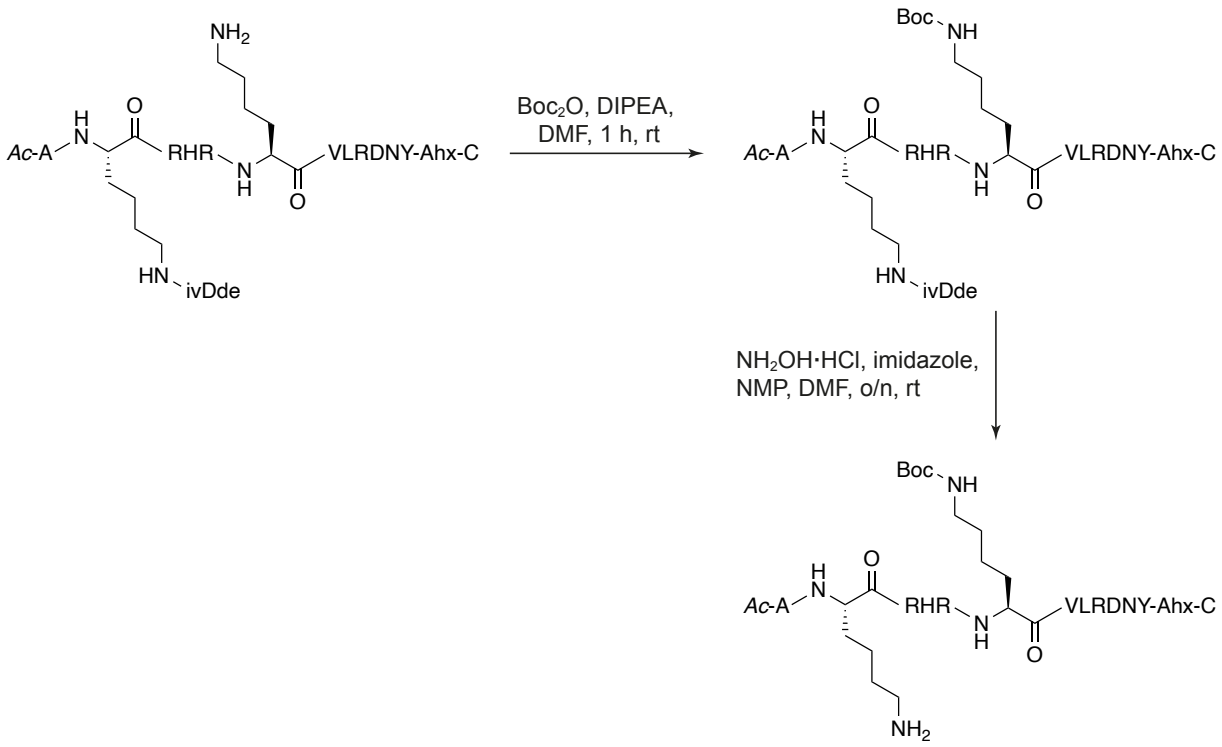
The published binding sequences consist of 4-5 residues flanking both sides of the acetylated lysine. An optional C-terminal tyrosine was added to peptide probes allowing quantification via UV absorption. In case of BRD4(1), the di-acetylated sequence of H4K5K8 showed the strongest binding affinity. The peptide probes for probing peptide-protein interactions were extended by a C-terminal cysteine for covalent immobilization on an agarose matrix and by a 6-aminohexanoic acid linker. The acetyl-lysine positions were replaced by potential acetyl-lysine mimics and N-terminal acetylation of the peptides furnished the final peptide probes used for immobilization. The general probe design was already established in [228]. Each sequence was synthesized with a lysine residue as negative and an acetyl-lysine residue as positive control. Di-modified H4K5K8 peptides were exclusively synthesized with enantiopure L-ApmTri (in the following referred to as ApmTri). All synthesized peptide probes and the corresponding purities are listed in table 2.

Table 2: Overview of synthesized peptide-based probes. Modifications at the recognized lysine positions are marked in bold. Different peptide probe sets are separated. Peptide purities were determined by integration of UV absorption in analytical HPLC chromatograms (see appendix figure C.34-C.43).

Peptide	Sequence	Purity
H4K20	*Ac*-AKRHR-**K**-VLRDNY-Ahx-C-NH_2	98.4 %
H4K20ac	*Ac*-AKRHR-**Kac**-VLRDNY-Ahx-C-NH_2	96.3 %
H4K20_L-AadTri	*Ac*-AKRHR-**L-AadTri**-VLRDNY-Ahx-C-NH_2	92.6%
H4K20_L-ApmTri	*Ac*-AKRHR-**L-ApmTri**-VLRDNY-Ahx-C-NH_2	97.4 %
H4K20_L-AsuTri	*Ac*-AKRHR-**L-AsuTri**-VLRDNY-Ahx-C-NH_2	95.5 %
H4K20Q	*Ac*-AKRHR-**Q**-VLRDNY-Ahx-C-NH_2	97.4 %
H4K20_D-ApmTri	*Ac*-AKRHR-**D-ApmTri**-VLRDNY-Ahx-C-NH_2	96.0 %
H4K20_DL-ApmTri	*Ac*-AKRHR-**DL-ApmTri**-VLRDNY-Ahx-C-NH_2	95.5 %
H4K20for	*Ac*-AKRHR-**Kfor**-VLRDNY-Ahx-C-NH_2	95.5 %
H4K20prop	*Ac*-AKRHR-**Kprop**-VLRDNY-Ahx-C-NH_2	93.0 %
H4K20benz	*Ac*-AKRHR-**Kbenz**-VLRDNY-Ahx-C-NH_2	89.2 %

Peptide	Sequence	Purity
H4K20_DL-ApmFTri	*Ac*-AKRHR-**DL-ApmFTri**-VLRDNY-Ahx-C-NH_2	93.6 %
H4K20_DL-ApmPTri	*Ac*-AKRHR-**DL-ApmPTri**-VLRDNY-Ahx-C-NH_2	98.0 %
H4K20_DL-ApmBTri	*Ac*-AKRHR-**DL-ApmBTri**-VLRDNY-Ahx-C-NH_2	95.3 %
H4K20Boc	*Ac*-AKRHR-**KBoc**-VLRDNY-Ahx-C-NH_2	96.3 %
H3K14	*Ac*-KSTGG-**K**-APRKQY-Ahx-C-NH_2	94.2 %
H3K14ac	*Ac*-KSTGG-**Kac**-APRKQY-Ahx-C-NH_2	96.7 %
H3K14_L-AadTri	*Ac*-KSTGG-**L-AadTri**-APRKQY-Ahx-C-NH_2	95.2 %
H3K14_DL-ApmTri	*Ac*-KSTGG-**DL-ApmTri**-APRKQY-Ahx-C-NH_2	96.7 %
H3K14_L-AsuTri	*Ac*-KSTGG-**L-AsuTri**-APRKQY-Ahx-C-NH_2	96.3 %
H3K14Q	*Ac*-KSTGG-**Q**-APRKQY-Ahx-C-NH_2	91.1 %
H3K56	*Ac*-IRRYQ-**K**-STELLY-Ahx-C-NH_2	94.1 %
H3K56ac	*Ac*-IRRYQ-**Kac**-STELLY-Ahx-C-NH_2	91.0 %
H3K56_L-AadTri	*Ac*-IRRYQ-**L-AadTri**-STELLY-Ahx-C-NH_2	92.5 %
H3K56_DL-ApmTri	*Ac*-IRRYQ-**DL-ApmTri**-STELLY-Ahx-C-NH_2	94.9 %
H3K56_L-AsuTri	*Ac*-IRRYQ-**L-AsuTri**-STELLY-Ahx-C-NH_2	91.8 %
H3K56Q	*Ac*-IRRYQ-**Q**-STELLY-Ahx-C-NH_2	88.8 %
H4K5K8	*Ac*-SGRG-**K**-GG-**K**-GLGK-Ahx-C-NH_2	98.2 %
H4K5/K8ac	*Ac*-SGRG-**Kac**-GG-**Kac**-GLGK-Ahx-C-NH_2	92.2 %
H4K5acK8	*Ac*-SGRG-**Kac**-GG-**K**-GLGK-Ahx-C-NH_2	94.6 %
H4K5K8ac	*Ac*-SGRG-**K**-GG-**Kac**-GLGK-Ahx-C-NH_2	95.9 %
H4K5/K8ApmTri	*Ac*-SGRG-**L-ApmTri**-GG-**L-ApmTri**-GLGK-Ahx-C-NH_2	96.6 %
H4K5ApmTriK8	*Ac*-SGRG-**L-ApmTri**-GG-**K**-GLGK-Ahx-C-NH_2	96.1 %
H4K5K8ApmTri	*Ac*-SGRG-**K**-GG-**L-ApmTri**-GLGK-Ahx-C-NH_2	97.4 %
H4K5ApmTriK8ac	*Ac*-SGRG-**L-ApmTri**-GG-**Kac**-GLGK-Ahx-C-NH_2	97.5 %
H4K5acK8ApmTri	*Ac*-SGRG-**Kac**-GG-**L-ApmTri**-GLGK-Ahx-C-NH_2	98.2 %

Peptides H4K5K8, H4K5/K8ac, H4K5acK8, H4K5K8ac, H4K5/K8ApmTri and H4K5ApmTriK8 were synthesized and purified by *Maren Scheffler* [239].
Lysine 20 acylations of H4K20 probes were installed as shown in scheme A.2 of the appendix. In brief, peptides were equipped with an Mtt-protected lysine building block at position K20 which was orthogonally deprotected using 1 % TFA in DCM. On-resin treatment of deprotected lysine with acetic formic anhydride, propionic anhydride and benzoic acid with PyOxim resulted in the corresponding acylated peptide (H4K20for, H4K20prop, H4K20benz).
The synthesis of the H4K20Boc peptide probe required the synthesis route illustrated in scheme 5.



Scheme 5: Synthesis route in solution for site-directed installation of a Boc group at the lysine 20 site of the H4K20 sequence.

Due to the acid-lability of the Boc group, this modification was installed after SPPS and cleavage of the peptide. The side chain of the second lysine at position 16 was protected with the acid-stable protecting group ivDde [240], which remained with the crude product after global deprotection. The crude peptide was purified by preparative HPLC and the Boc group was installed at lysine 20 by treatment with Boc anhydride in DMF. In a second step, the ivDde group was removed from K16 with a mixture of hydroxylamine/imidazole in NMP/DMF [241]. After an additional purification step by semi-preparative HPLC, the final H4K20Boc peptide was obtained.
Figure 17 shows structures and analytical data of mono-acetylated H4K20ac and di-acetylated H4K5/K8ac as representative examples for peptide probes applied in bromodomain binding assays. UV chromatograms and ESI-MS spectra of all peptides are shown in the appendix (figure C.34-C.43).

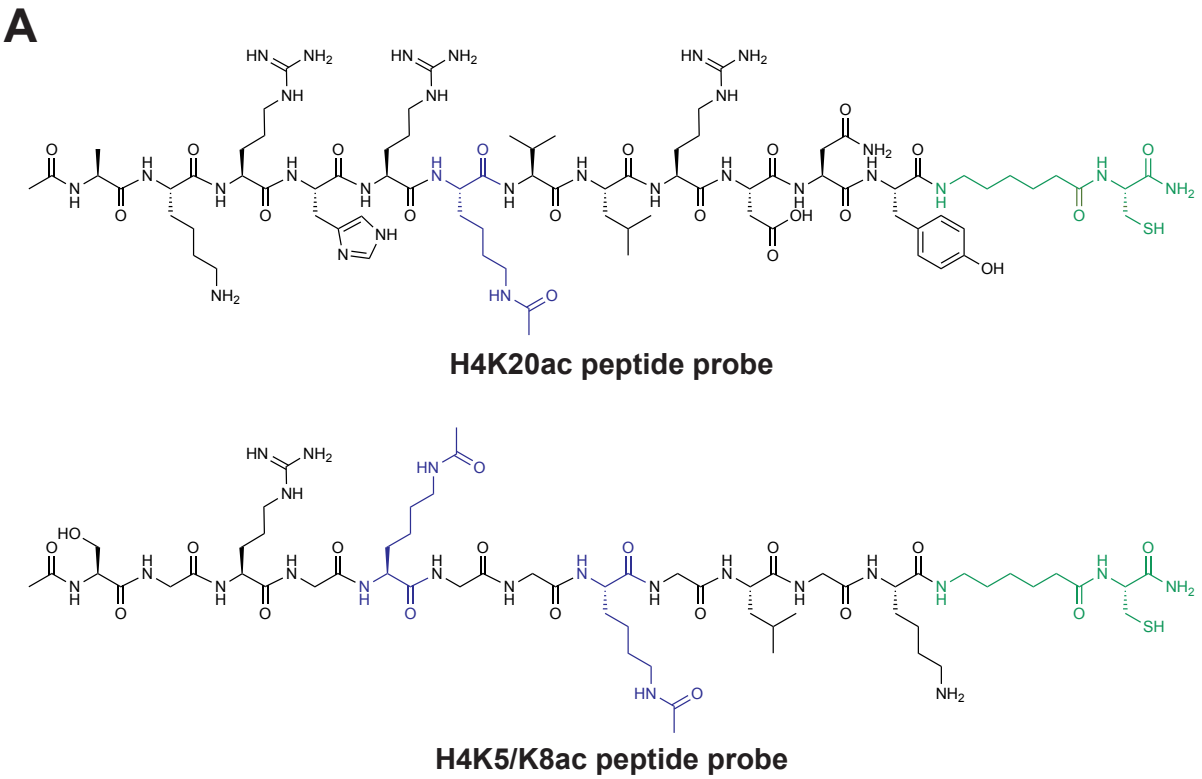


Figure 17: Structures and LC-MS analysis of synthesized peptide probes H4K20ac and H4K5/K8ac. A Structure of the synthesized H4K20ac (top) and H4K5/K8ac (bottom) peptide-based probes. BRD binding sites are highlighted in blue. The immobilization linker Ahx-Cys is colored in green. **B** Chemical characterizations of the peptides H4K20ac (top) and H4K5/K8ac (bottom). (Left panel) UV chromatograms at 218 nm of analytical HPLC analysis. (Right panel) ESI-MS spectra of LC-MS analysis, assignable m/z values are labeled.

2.4 General design of bromodomain constructs for recombinant expression

In order to investigate binding preferences and affinities of bromodomains to the synthesized peptide probes, several BRD constructs were cloned for recombinant protein expression in *E. coli*. The BRDs were selected from three BRD subfamilies (see 1.3.1): BAZ2B (PDB: 3GOL) from the subfamily V, CREBBP (PDB: 3DWY) from the subfamily III and three BRDs from the BET family: BRD3(2) (PDB: 2OO1), BRD4(1) (PDB: 2OSS) and BRD4(2) (PDB: 2OUO). The bromodomain amino acid sequences were taken from [242] and fused on genetic level to different fluorescent proteins. BAZ2B was fused to TagBFP [243], bromodomain 1 of BRD4 (BRD4(1)) to TagGFP2 [243] and bromodomain 2 of BRD3 (BRD3(2)), bromodomain 2 of BRD4 (BRD4(2)) and CREBBP were fused to TurboYFP [244]. Bromodomains and fluorescent labels were separated by a TEV protease cleavage site. Codon optimized synthetic genes coding for the different BRD fusion proteins were cloned into the pET28(+) expression vector. The constructs exhibit an N-terminal *His_6*-tag and a C-terminal *Strep II*-tag for purification. Plasmid maps are illustrated in appendix figure A.2. Schematic diagrams of the bromodomain constructs are shown in figure 18.

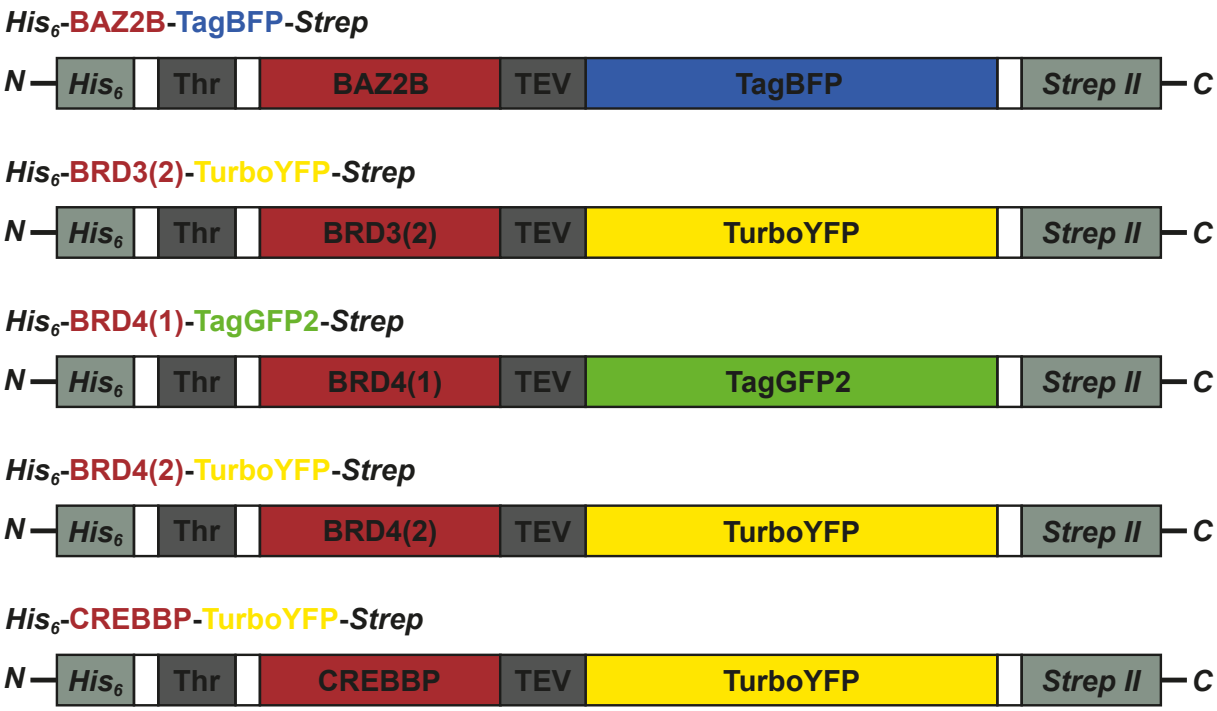


Figure 18: Schematic diagrams of the bromodomain constructs. Bromodomains are colored in red, fluorophores in their specific emission color, purification tags in light gray and protease cleavage sites (Thr: thrombin cleavage site; TEV: TEV protease cleavage site) in dark grey. Protein sequences are listed in the appendix, table B.1.

2.5 Pull-down experiments with bromodomains and peptide-based probes

2.5.1 Characterizing binding preferences of bromodomain BRD3(2) to triazole-based amino acids in H4K20 probes

Bromodomain interactions were investigated by incubating immobilized peptide probes with recombinantly expressed and fluorescently labeled bromodomain constructs. Previous experiments already showed weak interaction of the BET bromodomain BRD3(2) with the triazole amino acid AsuTri when incorporated into an H4 peptide probe (H4K20_L-AsuTri) [228]. In the following, the interaction between BRD3(2) and newly synthesized triazole moieties was characterized.

At first, the influence of the side chain length of triazole amino acids was investigated. For this purpose, a set of building blocks was synthesized with side chain lengths varying between three and five methylene groups (L-AadTri, L-ApmTri and L-AsuTri) and incorporated at the lysine 20 position of H4 peptide probes. Unmodified and acetylated lysines were used as negative and positive control. An additional peptide probe containing glutamine, a commonly used mimic of the acetylated lysines (see 1.4.3) was also tested for BRD binding.

The set of H4K20 peptide probes (H4K20, H4K20ac, H4K20_L-AadTri, H4K20_L-ApmTri, H4K20_L-AsuTri and H4K20Q) was immobilized on agarose beads enabling pull-down experiments with recombinant BRD3(2)-TurboYFP. The input concentration of BRD3(2)-TurboYFP was adjusted to the concentration of the published K_D value (10.5 µM for the H4K20ac sequence). In order to reduce unspecific interactions, BSA (10 mg/mL) was added to the BRD solution. Probe-bound proteins were analyzed by Coomassie-stained SDS-PAGE. A representative pull-down experiment is shown in figure 19.

The negative (H4K20) and positive (H4K20ac) controls showed the expected interaction profile with BRD3(2) enriched only on the acetylated probe. BRD3(2) was recruited to probes containing L-ApmTri and L-AsuTri, with a strong preference to the H4K20_L-ApmTri peptide. H4K20_L-AadTri and H4K20 did not show binding of BRD3(2). Thus, L-ApmTri with four methylene groups in the side chain appeared as the most potent acetyl-lysine mimic tested in this experiment. The pull-down experiment further indicated a weaker BRD binding to the L-ApmTri probe when compared to the native acetyl-lysine substrate. Based on this, the amino acid L-ApmTri was chosen for further investigations of bromodomain interactions with the triazole moiety.

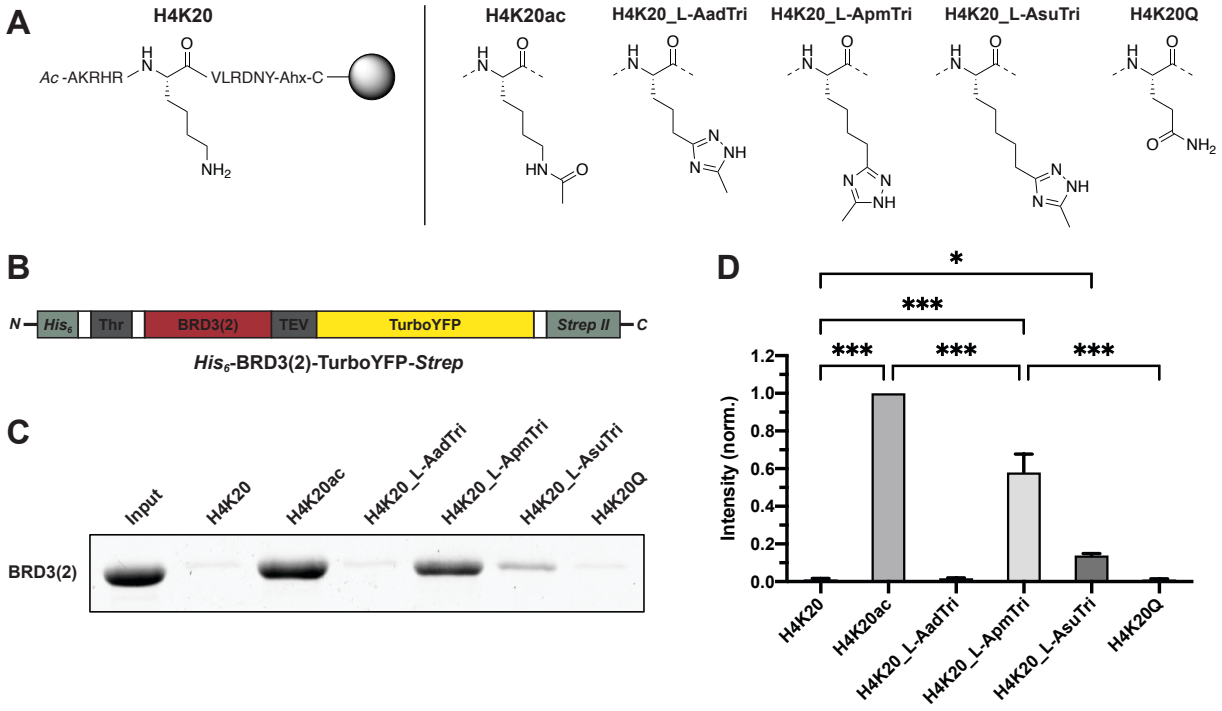


Figure 19: Pull-down results of BRD3(2) and triazole-containing H4K20 probes with varying side chain length. **A** Structures of immobilized H4K20 probes and their lysine 20 variations. **B** Schematic diagram of the BRD3(2)-TurboYFP fusion protein construct used in the pull-down assay. **C** Representative Coomassie-stained SDS-PAGE of a pull-down assay with BRD3(2) (10.5 µM) and H4K20 probes. **D** Quantification of the pull-down results. Values are normalized to the intensity of H4K20ac and shown as mean ± s.d (n=3). P values were calculated using one-way analysis of variance (ANOVA) with Tukey's post hoc test for multiple comparisons. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

In the following, the impact of ApmTri stereochemistry on BRD3(2) recruitment in the H4K20 peptide sequence context was investigated. Since ApmTri precursor Apm is only commercially available as racemic mixture, a synthesis strategy for enantiomeric resolution of DL-Apm was established (see 2.1.2) resulting in the enantiopure building blocks D-ApmTri and L-ApmTri.

The different ApmTri enantiomers and racemic DL-ApmTri were incorporated into the H4K20 sequence context (H4K20_D-ApmTri, H4K20_DL-ApmTri and H4K20_L-ApmTri) followed by pull-down assays under the same conditions as described above. The experiments are shown by representative coomassie-stained SDS-PAGE in figure 20.

In summary, BRD3(2)-TurboYFP showed strongest binding to the L-ApmTri probe while reduced binding to the racemic DL-ApmTri (34 % compared to L-ApmTri) and no binding to the enantiopure D-ApmTri was observed. This findings suggested that the triazole moiety does not bind to BRD3(2) independently of the amino acid context.

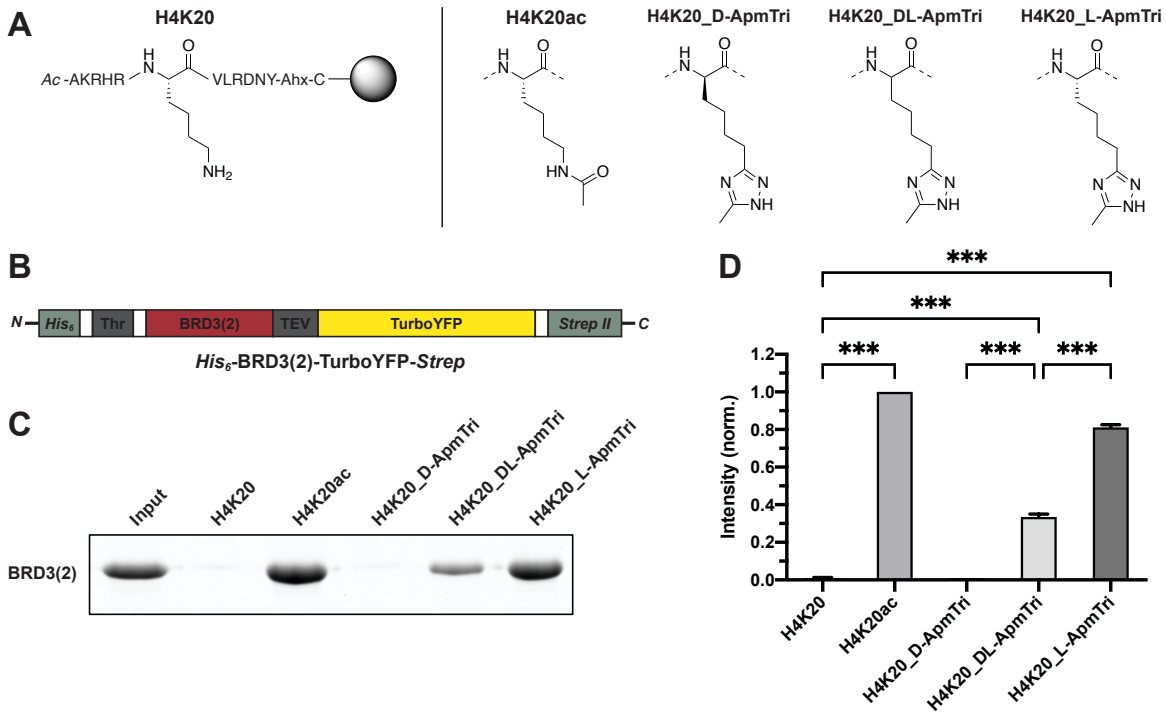


Figure 20: Pull-down results of examining the stereoselectivity of BRD3(2) to triazole-containing probes. **A** Structures of immobilized H4K20 probes and their lysine 20 variations. **B** Schematic diagram of the BRD3(2)-TurboYFP fusion protein construct used in the pull-down assay. **C** Representative Coomassie-stained SDS-PAGE of a pull-down assay with BRD3(2) (10.5 µM) and H4K20 probes. **D** Quantification of the pull-down results. Values are normalized to the intensity of H4K20ac and shown as mean ± s.d (n=3). P values were calculated using one-way ANOVA with Tukey's post hoc test for multiple comparisons. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

The experiments indicated a strong impact of the correct stereochemistry and linker length that is connecting the amino acid backbone with C-5 of the triazole. ApmTri carries a further methyl substitute at the C-3 position at the triazole installed as mimic of the corresponding methyl group in acetyl-lysine. It is known that bromodomains not only recognize acetyl-lysines, but also acylated residues including propionyl- and butyryl-lysine. In contrast, formylated lysines are not known to be bound by BRDs. The triazole synthesis route provides a comfortable way to vary the C-3 substituent by replacing acetamidine with other acylamidines in the coupling reaction prior to ring closure.
Three different triazole variants were selected for investigation and compared to the corresponding acyl-lysine in terms of BRD3(2) recruitment. The C3-methyl group was removed by using formamidine for triazole synthesis (ApmFTri), propionamidine resulted in a C3-ethyl substitution (ApmPTri) and benzamidine led to a bulky C3-benzyl group (ApmBTri). The triazole derivatives were only synthesized from the racemic DL-Apm. The variants were compared to the lysine acylations formyl-lysine (Kfor), propionyl-lysine (Kprop) and benzyl-lysine (Kbenz).
All amino acids were incorporated in the H4K20 peptide probes (H4K20for, H4K20prop, H4K20benz, H4K20_DL-ApmFTri, H4K20_DL-ApmPTri and H4K20_DL-ApmBTri) and tested for BRD3(2)-TurboYFP binding using the same conditions as described above. Representative Coomassie-stained SDS-PAGE analyses of the pull-downs are shown in figure 21.

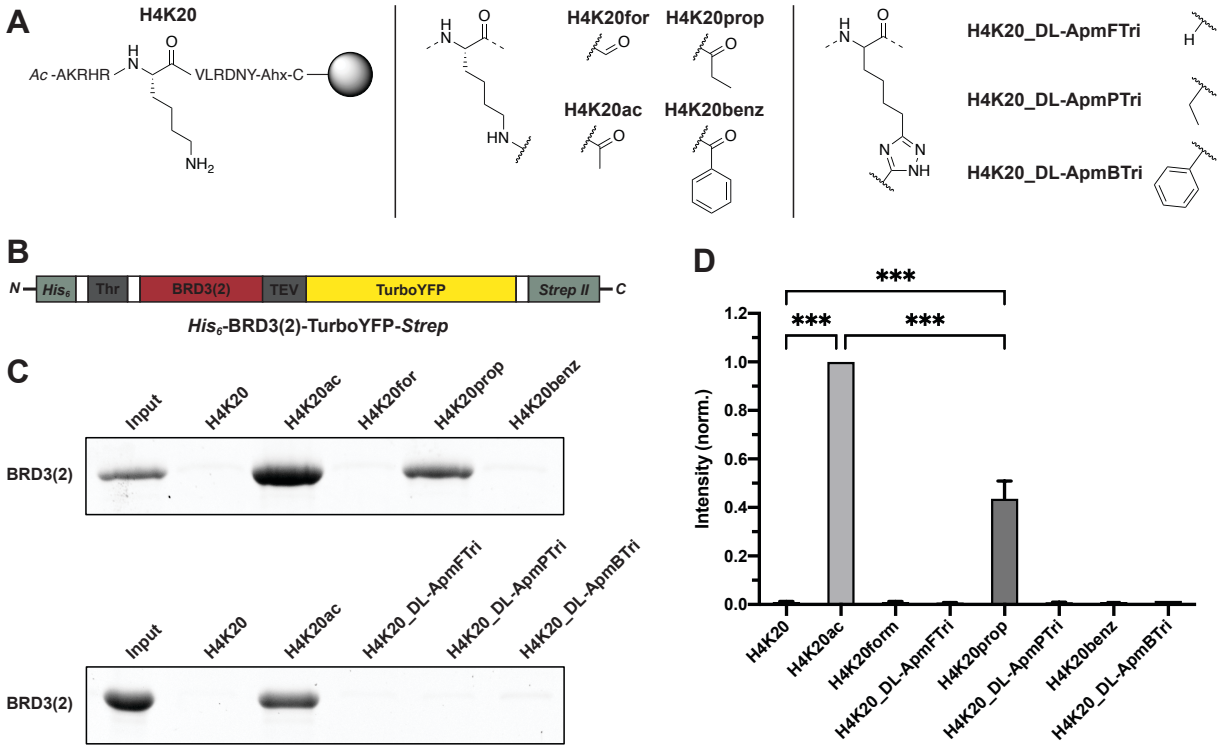


Figure 21: Pull-down results of BRD3(2) and acylated lysine probes in comparison to probes with modified triazole moieties. **A** Structures of immobilized H4K20 probes and their lysine 20 variations. **B** Schematic diagram of the BRD3(2)-TurboYFP fusion protein construct used in the pull-down assays. **C** Representative Coomassie-stained SDS-PAGEs of pull-down assays with BRD3(2) (10.5 µM) and H4K20 probes. **D** Quantification of the pull-down results. Values are normalized to the intensity of H4K20ac and shown as mean ± s.d (n=3). P values were calculated using one-way ANOVA with Tukey's post hoc test for multiple comparisons. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

The pull-down experiments showed, that BRD3(2) was only efficiently recruited to the acetyl-lysine (positive control) and propionyl-lysine probe whereas the formyl- and benzyl-acylation of lysine 20 as well as all triazole variants were not recognized as BRD3(2) substrates. Although propionyl-lysine was bound, the corresponding C3-ethyl triazole showed no interaction with BRD3(2). Removing of the methyl group (ApmFTri) resulted in a loss of binding as well. This findings pointed out the importance of the C3-methyl group for BRD3(2) binding. In summary, these discoveries indicate that L-ApmTri with the C-3 methyl substitution is the best triazole-based acetyl-lysine mimic for bromodomain BRD3(2). Therefore, subsequent experiments were carried out with the L-ApmTri amino acid, in the following also referred to as ApmTri.

2.5.2 Investigation of BRD4(2) binding to triazole-based amino acids

After characterization of BRD3(2) binding preferences, the synthesized H4K20 probes containing triazole-based amino acids with variable side chain length (L-AadTri, L-ApmTri, L-AsuHd) were used for probing the binding preferences of the homologous

BRD4(2) bromodomain. The second bromodomain of BRD3 and BRD4 share a high degree of structural homology and overlapping binding preferences.
Immobilized peptides were incubated with recombinant BRD4(2)-TurboYFP. The same input concentration as in BRD3(2) pull-down assays was used (10.5 µM) since no K_D value for the interaction between BRD4(2) and H4K20ac was reported. A representative pull-down experiment with the H4K20 peptide probe set is shown as coomassie-stained SDS-PAGE in figure 22.

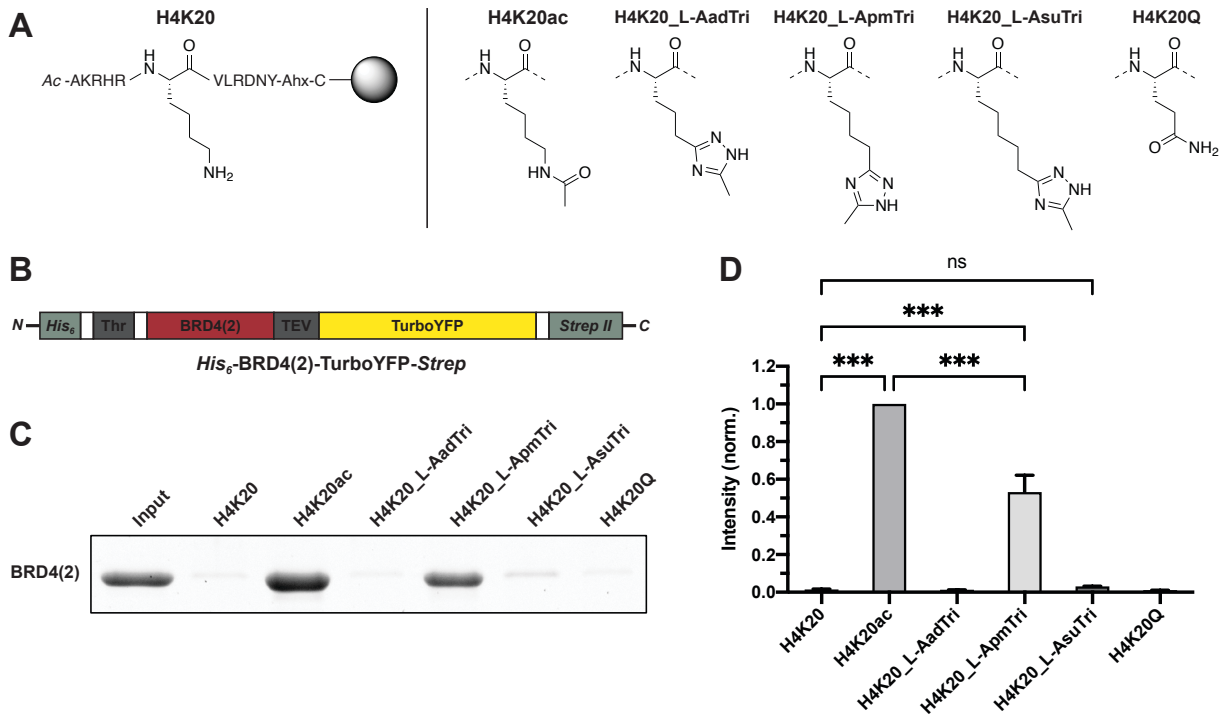


Figure 22: Pull-down results of BRD4(2) and triazole-containing peptide probes with varying side chain length. **A** Structures of immobilized H4K20 probes and their lysine 20 variations. **B** Schematic diagram of the BRD4(2)-TurboYFP fusion protein construct used in the pull-down assay. **C** Representative Coomassie-stained SDS-PAGE of a pull-down assay with BRD4(2) (10.5 µM) and H4K20 probes. **D** Quantification of the pull-down results. Values are normalized to the intensity of H4K20ac and shown as mean ± s.d (n=3). P values were calculated using one-way ANOVA with Tukey's post hoc test for multiple comparisons. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ns, not significant.

As expected, BRD4(2) showed a very similar binding pattern as BRD3(2). The enrichment of BRD4(2) on H4K20_L-ApmTri was strongly increased when compared with the negative probe H4K20, whereas binding to the extended side chain of H4K20_L-AsuTri was barely detected for BRD4(2) and no binding to H4K20_L-AadTri and H4K20Q was observed.

2.5.3 Examination of BAZ2B binding to triazole-based amino acids in H3K14 peptide probes

The previous pull-down experiments showed efficient recruitment of two members of the BET family (BRD3(2) and BRD4(2)) to the triazole moiety which reflects the specificity

of JQ1 for the BET family. The next objective was to characterize the binding specificity of triazole-based amino acids towards the non-BET bromodomains BAZ2B and CREBBP. Therefore, the triazole amino acids with variable side chain length were incorporated into the H3K14 sequence. This histone H3 derived sequence is a known binding site of the BAZ2B bromodomain with a reported K_D value of 7.6 µM. The H3K14 probe containing ApmTri was only synthesized with racemic DL-ApmTri amino acid. An unmodified lysine residue was used as negative and acetyl-lysine as positive probe resulting in a H3K14 peptide probe set containing H3K14, H3K14ac, H3K14_L-AadTri, H3K14_DL-ApmTri, H3K14_L-AsuTri and H3K14Q. The immobilized peptides were subjected to a pull-down assay with the recombinant BAZ2B-TagBFP fusion protein at an input concentration of 7.6 µM in presence of 10 mg/mL BSA. A representative Coomassie-stained SDS-PAGE of the pull-down assay is shown in figure 23.

The negative and positive probes showed the expected result by efficient recruitment of BAZ2B only to the acetyl-lysine probe, while other tested probes did not show any increased binding of BAZ2B. It can be concluded, that the assay with BAZ2B-TagBFP against the H3K14 peptide probes is applicable for the investigation of binding preferences of this bromodomain. However, none of the triazole amino acids AadTri, ApmTri and AsuTri could serve as acetyl-lysine mimic for BAZ2B.

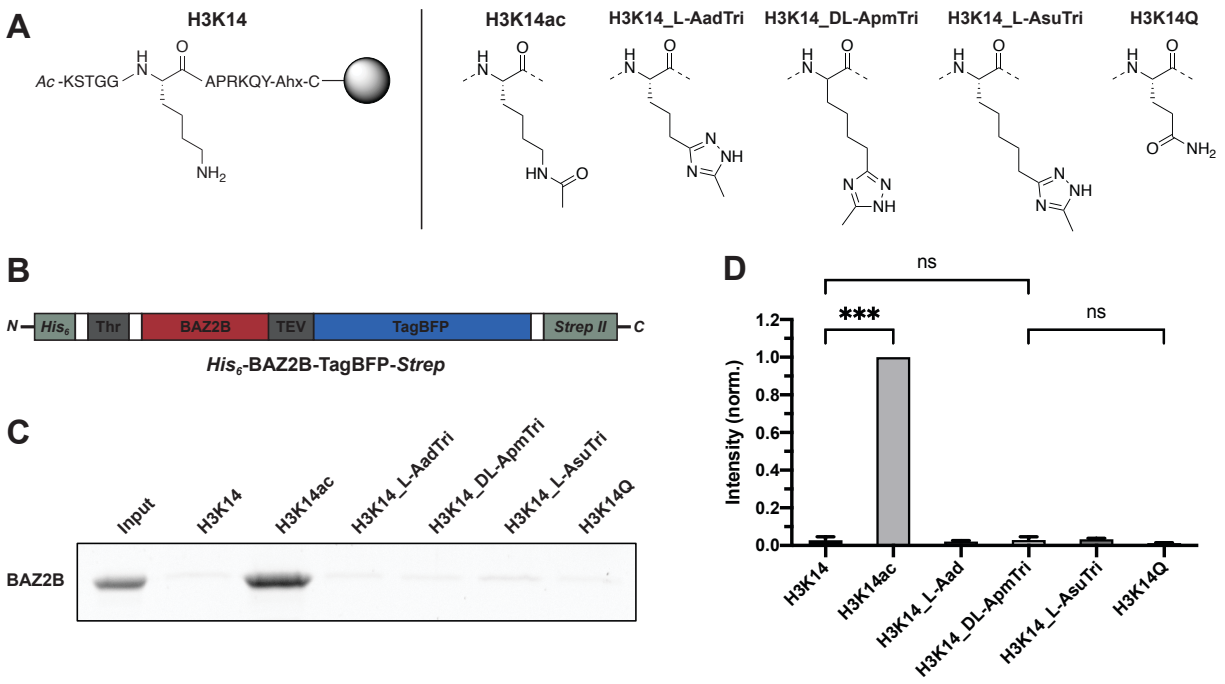


Figure 23: Pull-down results of BAZ2B and triazole-containing H3K14 probes varying side chain length. **A** Structures of immobilized H3K16 probes and their lysine 16 variations. **B** Schematic diagram of the BAZ2B-TagBFP fusion protein construct used in the pull-down assay. **C** Representative Coomassie-stained SDS-PAGE of a pull-down assay with BAZ2B (7.6 µM)and H3K14 probes. **D** Quantification of the pull-down results. Values are normalized to the intensity of H3K14ac and shown as mean ± s.d (n=3). P values were calculated using one-way ANOVA with Tukey's post hoc test for multiple comparisons. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ns, not significant.

2.5.4 Examination of CREBBP binding to triazole-based amino acids in H3K56 peptide probes

The results of the pull-down experiments with BAZ2B already indicated a selectivity of the triazole amino acids to specific bromodomains. For further clarification, the CREBBP bromodomain belonging to BRD subfamily III was tested with the triazole amino acid probes. A preferred binding site of CREBBP is acetylated lysine 56 of histone H3 with a reported affinity of 13.8 µM. The lysine residue was replaced with the triazole-containing amino acids of variable side chain lengths and glutamine. This resulted in a H3K56 peptide probe set comprising H3K56, H3K56ac, H3K56_L-AadTri, H3K56_DL-ApmTri, H3K56_L-AsuTri and H3K56Q.
The immobilized peptides were applied in a pull-down assay against the recombinant CREBBP-TurboYFP fusion protein at an input concentration of 13.8 µM supplemented with 10 mg/mL BSA. A representative Coomassie-stained SDS-PAGE of the pull-down assay is shown in figure 24.

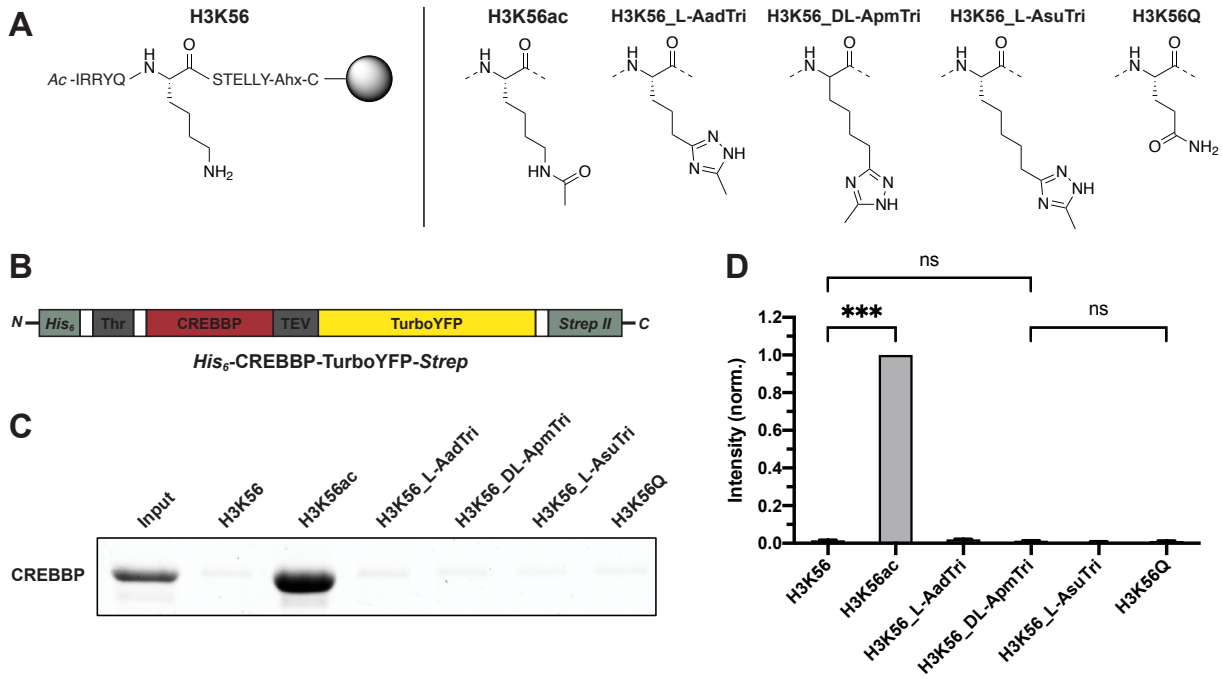


Figure 24: Pull-down results of CREBBP and triazole-containing H3K56 probes varying side chain length. **A** Structures of immobilized H3K56 probes and their lysine 56 variations. **B** Schematic diagram of the CREBBP-TurboYFP fusion protein construct used in the pull-down assay. **C** Representative Coomassie-stained SDS-PAGE of a pull-down assay with CREBBP (13.8 µM) and H3K56 probes. **D** Quantification of the pull-down results. Values are normalized to the intensity of H3K56ac and shown as mean ± s.d (n=3). P values were calculated using one-way ANOVA with Tukey's post hoc test for multiple comparisons. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; ns, not significant.

The pull-downs of CREBBP against H3K56 probes displayed a similar pattern as observed for the pull-down experiments of BAZ2B with the H3K14 peptide probe set. The positive

control H3K56ac recruited CREBBP, whereas none of the other H3K56 probes interacted with this bromodomain. This findings further support the notion that the triazole-based amino acids serve as selective acetyl-lysine mimics for bromodomains of the BET family.

2.5.5 Characterization of binding preferences of bromodomain BRD4(1) to dual-modified H4K5K8 peptide probes

After demonstrating binding of BRD3(2) and BRD4(2) to triazole-based amino acid variants with the highest enrichment found for the L-ApmTri amino acid, the following experiments are focussing on L-ApmTri and the interactions with BET bromodomains.
The small molecule inhibitor JQ1 is not only bound by the second bromodomains (BD2) of BET proteins, but also by the first bromodomains (BD1). An important feature of BD1 domains is the relatively weak binding of mono-acetylated substrates. The binding affinity is strongly enhanced with multi-acetylated substrates containing at least two acetyl-lysine residues in close proximity.
In order to study potential binding of L-ApmTri by BD1 domains, the first bromodomain of BRD4 (BRD4(1)) was recombinantly expressed as fusion construct BRD4(1)-TagGFP2. Binding of BRD4(1) to the di-acetylated H4K5/K8ac sequence is a confirmed interaction (K_D value of 6.8 µM). H4K5K8 peptide probes were synthesized with enantiopure L-ApmTri, acetyl-lysine and free lysine at position K5 and K8. An unmodified H4K5K8 peptide was used as negative and a di-acetylated H4K5/K8ac peptide as positive control. In addition, all combinations of non-acetylated, acetylated and L-ApmTri replacements at both positions were synthesized, resulting in the H4K5K8 peptide probe set (H4K5K8, H4K5/K8ac, H4K5/K8ApmTri, H4K5acK8, H4K5ApmTriK8, H4K5acK8ApmTri, H4K5K8ac, H4K5K8ApmTri and H4K5ApmTriK8ac).
Preliminary experiments showed that BRD4(1) was washed of the probes much more rapidly than BRD3(2) and BRD4(2). The washing protocol was adjusted accordingly by reducing washing time and salt concentration in the washing buffer. Cloning and expression of BRD4(1)-TagGFP2 and preliminary optimization of pull-down conditions were carried out by *Maren Scheffler* [239].
The H4K5K8 peptide probe set was examined for BRD binding by incubation with BRD4(1)-TagGFP2 at an input concentration of 6.8 µM, supplemented with a reduced BSA concentration of 0.1 mg/mL. Representative results of the pull-down experiments are shown as Coomassie-stained SDS-PAGEs in figure 25.
The di-acetylated control H4K5/K8ac showed significantly increased enrichment of BRD4(1) compared to the non-acetylated H4K5K8. The H4K5acK8ApmTri probe showed no significant difference in BRD4(1) recruitment when compared to the di-acetylated substrate. Binding to the probe with L-ApmTri at both positions (H4K5/K8ApmTri) was reduced when compared to the native substrate and the mixed H4K5acK8ApmTri probe. The mono-acetylated probes H4K5acK8 and H4K5K8ac were less efficient ligands of BRD4(1) and substitution of the acetyl-lysines with L-ApmTri resulted in even weaker bind-

ing. Less effective recruitment was also observed for the H4K5ApmTriK8ac probe with L-ApmTri at the lysine 5 position.

In summary, H4K5K8 peptide probes are suitable for characterizing BRD4(1) binding preferences with mono- and di-modified peptide substrates. Similar to the other tested BET bromodomains BRD3(2) and BRD4(2), BRD4(1) accepted L-ApmTri as acetyl-lysine mimic. Replacing K5 and/or K8 with L-ApmTri weakened the binding of BRD4(1) when compared to the corresponding acetyl-lysine probe. The combination of acetylated lysine 5 with substitution of lysine 8 by L-ApmTri (H4K5acK8ApmTri) led to BRD binding comparable to the native di-acetylated substrate. These results suggest, that L-ApmTri is a better substitute for acetyl-lysine at position K8 than at position K5 in the applied sequence context.

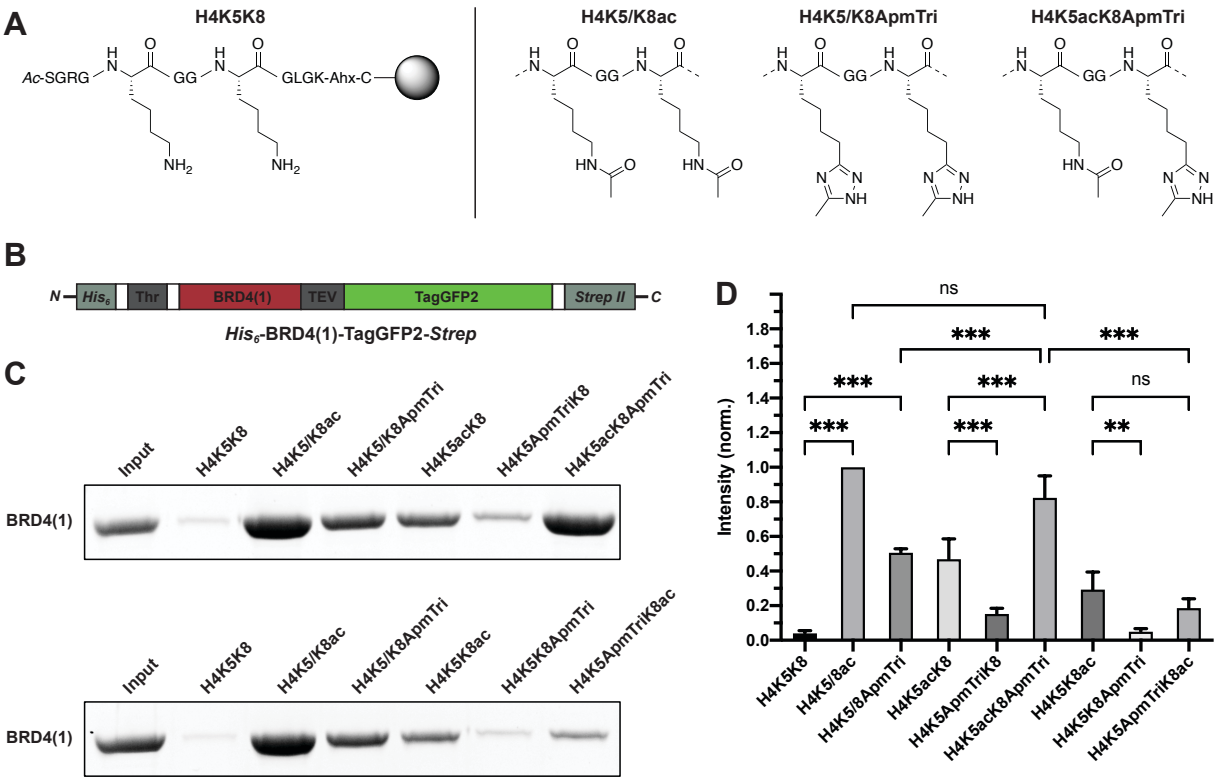


Figure 25: Pull-down results of BRD4(1) on (di-)acetylated and L-ApmTri-containing H4K5K8 probes. **A** Structures of immobilized H4K5K8 probes and three exemplary variations at lysine position 5 and 8. Peptides were synthesized with enantiopure L-ApmTri. **B** Schematic diagram of the BRD4(1)-TagGFP2 fusion protein construct used in the pull-down assays. **C** Representative Coomassie-stained SDS-PAGEs of pull-down assays with BRD4(1) (6.8 µM) and H4K5K8 probes. **D** Quantification of the pull-down results. Values are normalized to the intensity of H4K5/K8ac and shown as mean ± s.d (n=3). P values were calculated using one-way ANOVA with Tukey's post hoc test for multiple comparisons. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ns, not significant.

2.5.6 Inhibitor titration of JQ1 and GSK2801 in pull-down assays with BRD3(2) and BAZ2B

The previously described pull-down assays demonstrated effective enrichment of BRD3(2) and BAZ2B on their native acetylated substrate sequences (H4K20ac and

H3K14ac). In addition, BRD3(2) also showed binding to the triazole-based amino acid L-ApmTri. In a further step, the established pull-down assay was modified in order to perform competition assays with small molecule inhibitors and immobilized peptide probes. Two different small molecules were assayed: The BET inhibitor (+)-JQ1 as specific ligand of BET protein BRD3(2) (K_D of 82 nM) and GSK2801, a specific BAZ2B ligand with a K_D of 136 nM [196].

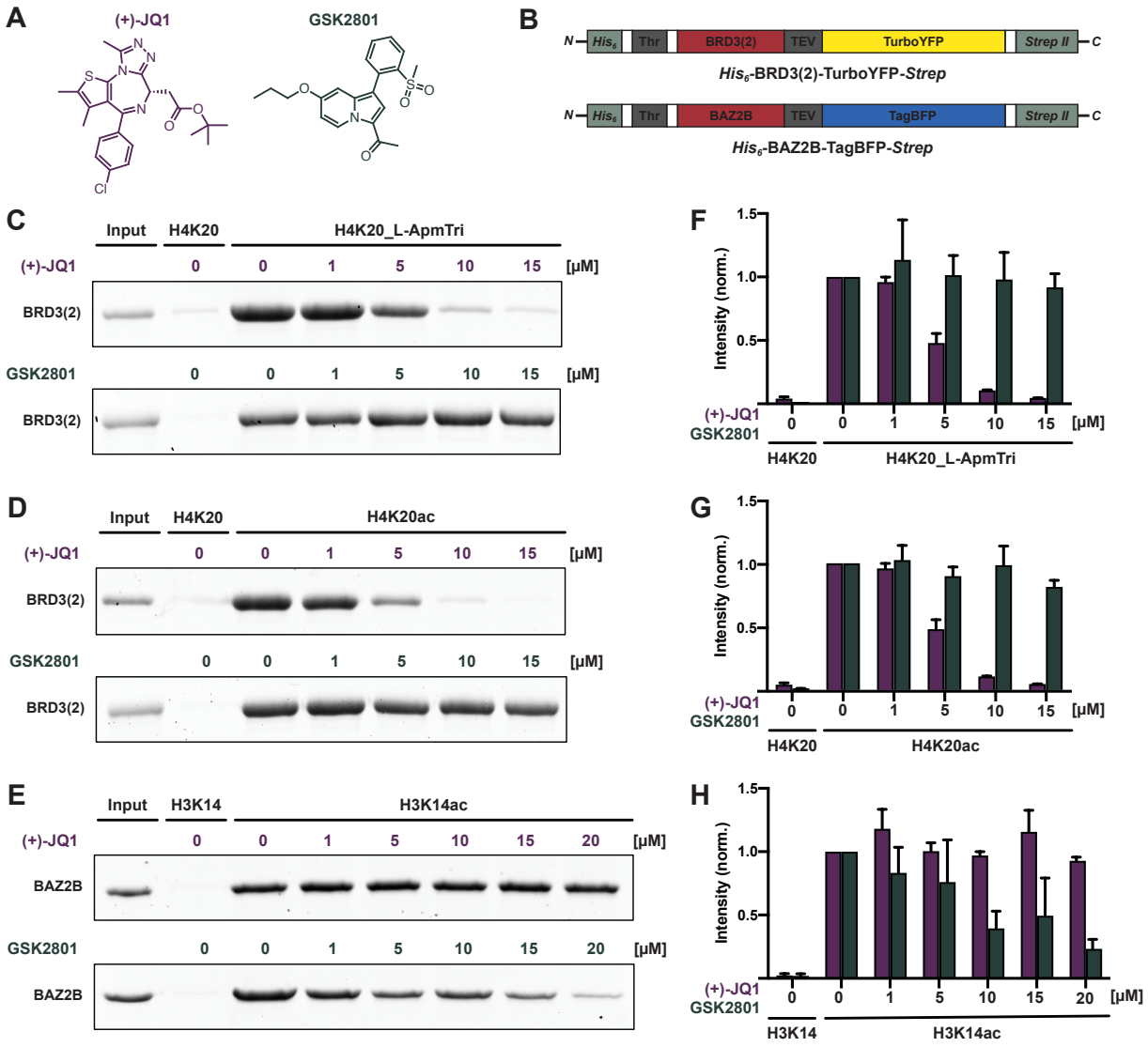


Figure 26: Pull-downs showing the titration of (+)-JQ1 and GSK2801 against BRD3(2) and BAZ2B to H4K20 and H3K14 probes. A Structures of the inhibitors (+)-JQ1 (purple) and GSK2801 (dark green). **B** Schematic diagrams of the BRD3(2)-TurboYFP and BAZ2B-TagBFP fusion proteins used in the pull-down assays. **C**-**E** Representative Coomassie-stained SDS-PAGEs of pull-down assays with BRD3(2) and BAZ2B (both 10.5 µM) on H4K20 and H3K14 probes. **F**-**H** Quantification of the pull-down results. Values are normalized to the intensity of the non-treated probe and shown as mean ± s.d (n=3). Results of JQ1 titration are colored in purple and results of GSK2801 titration in green.

Competition assays were carried out by incubation of the bromodomains with immobilized peptide substrates in prescence of the corresponding inhibitor at various con-

centrations. BRD3(2)-TurboYFP (10.5 µM) solution was supplemented with (+)-JQ1 or GSK2801 with concentrations ranging from 0 to 15 µM, followed by incubation with H4K20ac or H4K20_L-ApmTri probes. BAZ2B-TagBFP protein (10.5 µM) was incubated with H3K14ac probes in presence of GSK2801 at concentrations between 0 and 20 µM. Representative coomassie-stained SDS-PAGEs for the different inhibitor titration experiments are shown in figure 26.
Binding of BRD3(2) to H4K20ac as well as to H4K20_L-ApmTri was strongly affected by the presence of JQ1 during incubation. At equal concentrations of JQ1 (10 µM) and bromodomain (10.5 µM) the recruitment to the peptide substrates was barely detectable. Control experiments with equal concentrations of GSK2801 showed no significant inhibition of BRD3(2) binding. However, the addition of GSK2801 to pull-downs with BAZ2B and H3K14ac reduced the interaction between bromodomain and peptide probe efficiently. Compared to BRD3(2) and JQ1, higher concentrations (> 20 µM) of GSK2801 were required for competing BAZ2B off the probe. Titration of JQ1 to BAZ2B showed no effects on BAZ2B recruitment, confirming the selectivity of the small molecule inhibitors.

2.6 Quantification of bromodomain-peptide interactions by MST

2.6.1 Determination of BRD3(2) binding affinities with H4K20 peptides

The established pull-down assays allow fast screening of interactions between bromodomains and immobilized peptide substrates. The assays are also applicable for identification and quantification of bromodomain binding preferences with various probes of interest, but less suitable for determination of dissociation constants. In the following, the observed protein-peptide interactions between BRD3(2)-TurboYFP and H4K20_L-ApmTri and H4K20_L-AsuTri were quantified by microscale thermophoresis (MST).
MST measurements required fluorescent labeling of one interaction partner, here achieved by *His_6*-tag labeling of bromodomain BRD3(2) with the commercially available red dye NT647-tris-NTA [245]. BRD3(2) was expressed with an N-terminal *His_6*-tag lacking the fluorescent protein. Dilution series of the peptides H4K20ac, H4K20_L-ApmTri and H4K20_L-AsuTri were incubated with labeled BRD3(2) and MST was measured for each peptide dilution. The recorded MST traces and the resulting dose-response curves are shown in figure A.3 of the appendix. Data processing is described in 4.2.4.15. Figure 27 shows the determined dose-response curves and the obtained K_D values.
The MST measurement for BRD3(2) binding of the native substrate H4K20ac yielded a K_D value of 3.2 $\pm$ 0.3 µM. The determined K_D values of the triazole-containing peptides were 6.5 $\pm$ 0.5 µM for H4K20_L-ApmTri and 17.9 $\pm$ 1.7 µM for H4K20_L-AsuTri. The dissociation constants mirrored the preferred interaction of BRD3(2) with the L-ApmTri probe

already observed in the pull-down assays. The K_D of BRD3(2) to H4K20_L-ApmTri was approximately 2-fold higher than to the corresponding acetyl-lysine peptide. The K_D for H4K20_L-AsuTri was about 3-fold higher than for H4K20_L-ApmTri.

A

His_6-BRD3(2)

C

Peptide	Sequence	Affinity [µM]
H4K20ac	*Ac*-AKRHR-**Kac**-VLRDNY-Ahx-C-*NH_2*	3.2 ± 0.3
H4K20_L-ApmTri	*Ac*-AKRHR-**L-ApmTri**-VLRDNY-Ahx-C-*NH_2*	6.5 ± 0.5
H4K20_L-AsuTri	*Ac*-AKRHR-**AsuTri**-VLRDNY-Ahx-C-*NH_2*	17.9 ± 1.7

B

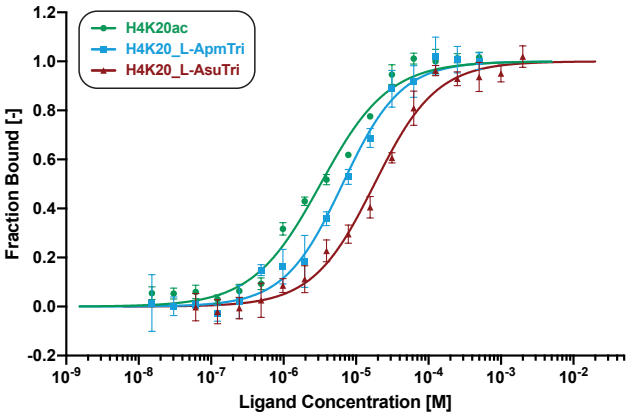
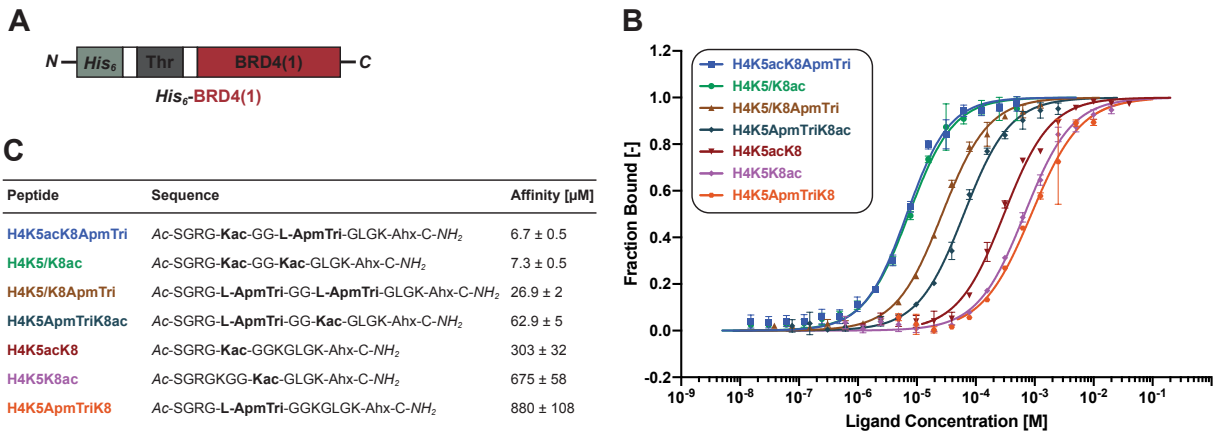


Figure 27: MST analysis of BRD3(2) interactions with H4K20 peptides. **A** Schematic diagram of the BRD3(2) protein construct with N-terminal *His_6*-tag used for labeling with RED-tris-NTA. **B** Dose-response curves of RED-tris-NTA-BRD3(2) towards H4K20ac (green), H4K20_L-ApmTri (blue) and H4K20_L-AsuTri (red). Fraction bound values are shown as mean ± s.d ($n \geq 3$). K_D values were determined by fitting the dose-response curves as described in 4.2.4.15. **C** Table listing the H4K20 peptides applied in MST measurements and the determined binding affinities (K_D values).

2.6.2 Determination of BRD4(1) binding affinities to H4K5K8 peptides

In accordance with the determination of binding affinities of BRD3(2), the interactions between BRD4(1) and mono- or di-modified H4K5K8 peptides were quantified by MST. BRD4(1) was expressed with an N-terminal *His_6*-tag lacking the fluorescent label and was labeled with NT647-tris-NTA afterwards. All peptides, which have shown detectable bromodomain enrichment in the pull-down assay (H4K5/K8ac, H4K5/K8ApmTri, H4K5acK8, H4K5ApmTriK8, H4acK8ApmTri, H4K5K8ac and H4K5ApmTriK8ac) were applied in MST experiments. The recorded MST traces and the resulting dose-response curves are shown in figure A.4 of the appendix. Figure 28 shows a comparison of the final dose-response curves and the determined K_D values for all peptides.

The lowest K_D values were measured for the mixed H4K5acK8ApmTri peptide with 6.7 ± 0.5 µM and for the native di-acetylated substrate H4K5/K8ac with 7.3 ± 0.5 µM. Both peptides also showed strongest BRD binding in the pull-down experiments. Lower affinities were determined for peptides with K5 replaced by L-ApmTri. BRD4(1) binding to double triazole-containing peptide H4K5/K8ApmTri resulted in a K_D of 26.9 µM and the mixed H4K5ApmTriK8ac peptide showed a K_D of 62.9 µM. The highest K_Ds were determined for mono-acetylated peptides H4K5acK8 (303 µM), H4K5K8ac (675 µM) and the mono-triazole-containing peptide H4K5ApmTriK8 (880 µM).



C

Peptide	Sequence	Affinity [µM]
H4K5acK8ApmTri	*Ac*-SGRG-**Kac**-GG-**L-ApmTri**-GLGK-Ahx-C-NH_2	6.7 ± 0.5
H4K5/K8ac	*Ac*-SGRG-**Kac**-GG-**Kac**-GLGK-Ahx-C-NH_2	7.3 ± 0.5
H4K5/K8ApmTri	*Ac*-SGRG-**L-ApmTri**-GG-**L-ApmTri**-GLGK-Ahx-C-NH_2	26.9 ± 2
H4K5ApmTriK8ac	*Ac*-SGRG-**L-ApmTri**-GG-**Kac**-GLGK-Ahx-C-NH_2	62.9 ± 5
H4K5acK8	*Ac*-SGRG-**Kac**-GGKGLGK-Ahx-C-NH_2	303 ± 32
H4K5K8ac	*Ac*-SGRGKGG-**Kac**-GLGK-Ahx-C-NH_2	675 ± 58
H4K5ApmTriK8	*Ac*-SGRG-**L-ApmTri**-GGKGLGK-Ahx-C-NH_2	880 ± 108

Figure 28: MST analysis of BRD4(1) interactions with H4K5K8 peptides. A Schematic diagram of the BRD4(1) protein construct with N-terminal *His_6*-tag used for labeling with RED-tris-NTA. **B** Dose-response curves of RED-tris-NTA-BRD4(1) towards H4K5acK8ApmTri (blue), H4K5/K8ac (green), H4K5/K8ApmTri (brown), H4K5ApmTriK8ac (blue-gray), H4K5acK8 (red), H4K5K8ac (purple) and H4K5ApmTriK8 (orange). Fraction bound values are shown as mean ± s.d ($n \geq 3$). K_D values were determined by fitting the dose-response curves as described in 4.2.4.15. **C** Table listing the H4K5K8 peptides applied in MST measurements and the determined dissociation constants (K_D values).

2.7 Determining crystal structures of BRD3(2) and BRD4(1) in complex with L-ApmTri peptides

Pull-down and MST experiments with bromodomains BRD3(2) and BRD4(1) demonstrated binding of acetyl-lysine and L-ApmTri substituted peptide probes. In order to study the interactions between bromodomains and ApmTri containing substrate peptides on molecular level, BRD3(2) and BRD4(1) were co-crystallized with the peptides H4K20_L-ApmTri, H4K5acK8ApmTri and H4K5/K8ApmTri. The peptides were synthesized as shorter variants of the pull-down probes by removing the immobilization linker consisting of a C-terminal Ahx and cysteine. Peptide sequences and achieved purities are shown in table 3.

Table 3: Peptides for co-crystallization with BRD3(2) and BRD4(1). Modifications at the recognized lysine positions are marked in bold. Peptide purities were determined by integration of UV absorption in analytical HPLC chromatograms (see appendix figure C.37 and C.43).

Peptide	Sequence	Purity
H4K20_L-ApmTri short	*Ac*-AKRHR-**L-ApmTri**-VLRDNY-NH_2	98.7 %
H4K5acK8ApmTri short	*Ac*-SGRG-**Kac**-GG-**L-ApmTri**-GLGK-NH_2	94.8 %
H4K5/K8ApmTri short	*Ac*-SGRG-**L-ApmTri**-GG-**L-ApmTri**-GLGK-NH_2	97.6 %

Crystal structures were determined in cooperation with the group of Prof. Dr. Thilo Stehle (University of Tübingen, Germany) by *Cengiz Koç*, *Natascha Bartlick* and *Michael Braun*. Crystallization conditions are listed in table 22. The crystal structure of BRD3(2) in com-

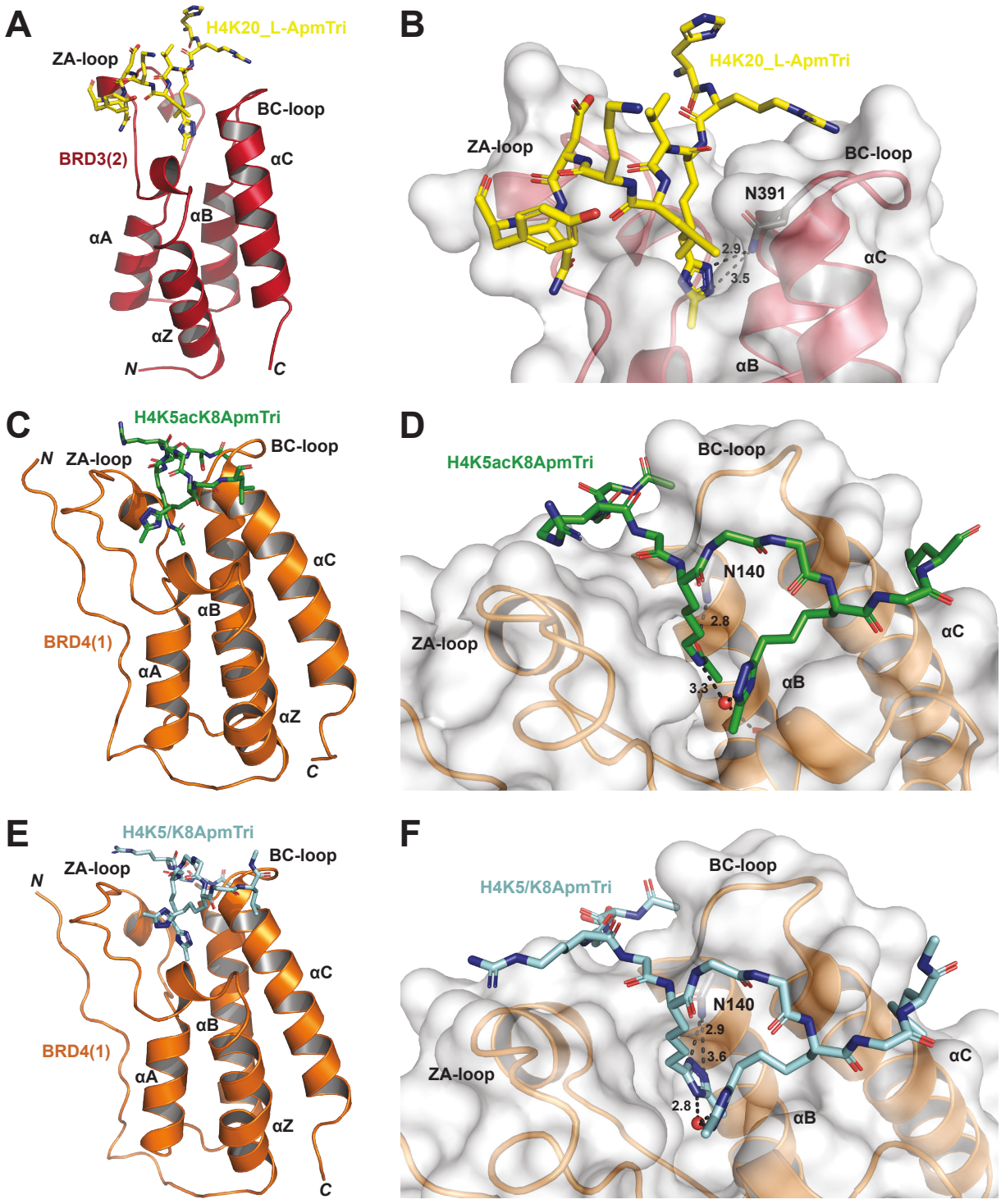


Figure 29: Crystal structures of BRD3(2) and BRD4(1) in complex with L-ApmTri substrate peptides. A, B Crystal structure of BRD3(2) (red) in complex with H4K20_L-ApmTri (yellow) at 1.85 Å. **C, D** Crystal structure of BRD4(1) (orange) in complex with H4K5acK8ApmTri (green) at 1.85 Å. **E, F** Crystal structure of BRD4(1) in complex with H4K5/K8ApmTri (light blue) at 1.58 Å. Bromodomain secondary structural elements and conserved asparagine side chains are labeled. Water molecules are illustrated as red spheres and hydrogen bonds (distance cut-off: 3.6 Å) are shown as grey dashes and labeled in Å.

plex with H4K20_L-ApmTri was solved at a resolution of 1.85 Å, the crystal structure of BRD4(1) in complex with H4K5acK8ApmTri was solved at 1.85 Å and in complex with

H4K5/K8ApmTri at 1.58 Å. Final protein structures are illustrated in figure 29. Electron density maps of L-ApmTri peptide ligands bound to BRD3(2) and BRD4(1) are displayed in figure A.5 and structure refinement data are shown in table A.3 of the appendix.
All three determined protein structures displayed the conserved fold of the bromodomains consisting of four α-helices (αZ, αA, αB, αC), connected by loop segments (ZA- and BC-loop). The crystal structure of BRD3(2) in complex with H4K20_L-ApmTri revealed, that the L-ApmTri amino acid is reaching into the BRD binding pocket of acetyl-lysine. The interaction is stabilized by anchoring hydrogen bonds between the aromatic triazole structure and the conserved asparagine residue N391, thereby mimicking the native acetyl-lysine binding mode. The peptide backbone is positioned in the surface cavity of BRD3(2), formed by the ZA- and BC-loop, responsible for substrate sequence selectivity.
In case of BRD4(1) in complex with H4K5acK8ApmTri and H4K5/K8ApmTri, the residue at position K5 (K5ac or L-ApmTri) is inserted into the binding pocket, forming the anchoring hydrogen bond to the conserved N140 residue. K8ApmTri is positioned at the protein surface in diagonal orientation, stabilized primarily through hydrophobic contacts. The residues K5ac as well as K5ApmTri are interacting with the triazole amino acid at position K8 by water-mediated hydrogen bonds, structurally linking both modified residues when bound to BRD4(1).

2.8 Verification of L-ApmTri stability in cellular environments

The next objective was the application of L-ApmTri in cell-based systems in order to study BRD interactions under more biological conditions. The stability of L-ApmTri in cellular lysate towards demodifying enzymes (especially deacetylases) was investigated in the following. A previously established MALDI-based assay for detecting deacetylase activity in cellular lysates [246] was adapted for testing the stability of the non-canonical amino acid (ncAA) L-ApmTri in human cell (HeLa) lysate.
Peptides H4K20, H4K20ac and H4K20_L-ApmTri were added in a concentration of 400 µM to 15 µg HeLa whole cell lysate and incubated for 4 h. Reactions were stopped after 0 h and 4 h by adding the pan-HDAC inhibitor trichostatin A and samples were analyzed by MALDI-MS. Figure 30 shows representative MS spectra at both time points. The non-acetylated peptide H4K20 (1) was used as control for assigning potential side products caused by proteolytic degradation. Deacetylation of H4K20ac (2) after 4 h led to an additional signal with a mass of -42 Da, which corresponds to deacetylated H4K20 (1). After 4 h, H4K20_L-ApmTri (3) was partially digested by proteolytic cleavage N-terminal to arginine 23 resulting in the peptide fragment ΔH4K20_L-ApmTri (4). Importantly, the triazole moiety of the amino acid was not affected in the cell lysate after 4 h and educt as well as digested side product showed the expected mass with an intact L-ApmTri amino acid.

The deacetylation of H4K20ac demonstrated the functionality of the MALDI-based stability assay for verifying the stability of L-ApmTri against deacetylases in HeLa whole cell extract. This findings allowed the subsequent application of L-ApmTri in cell-based experiments.

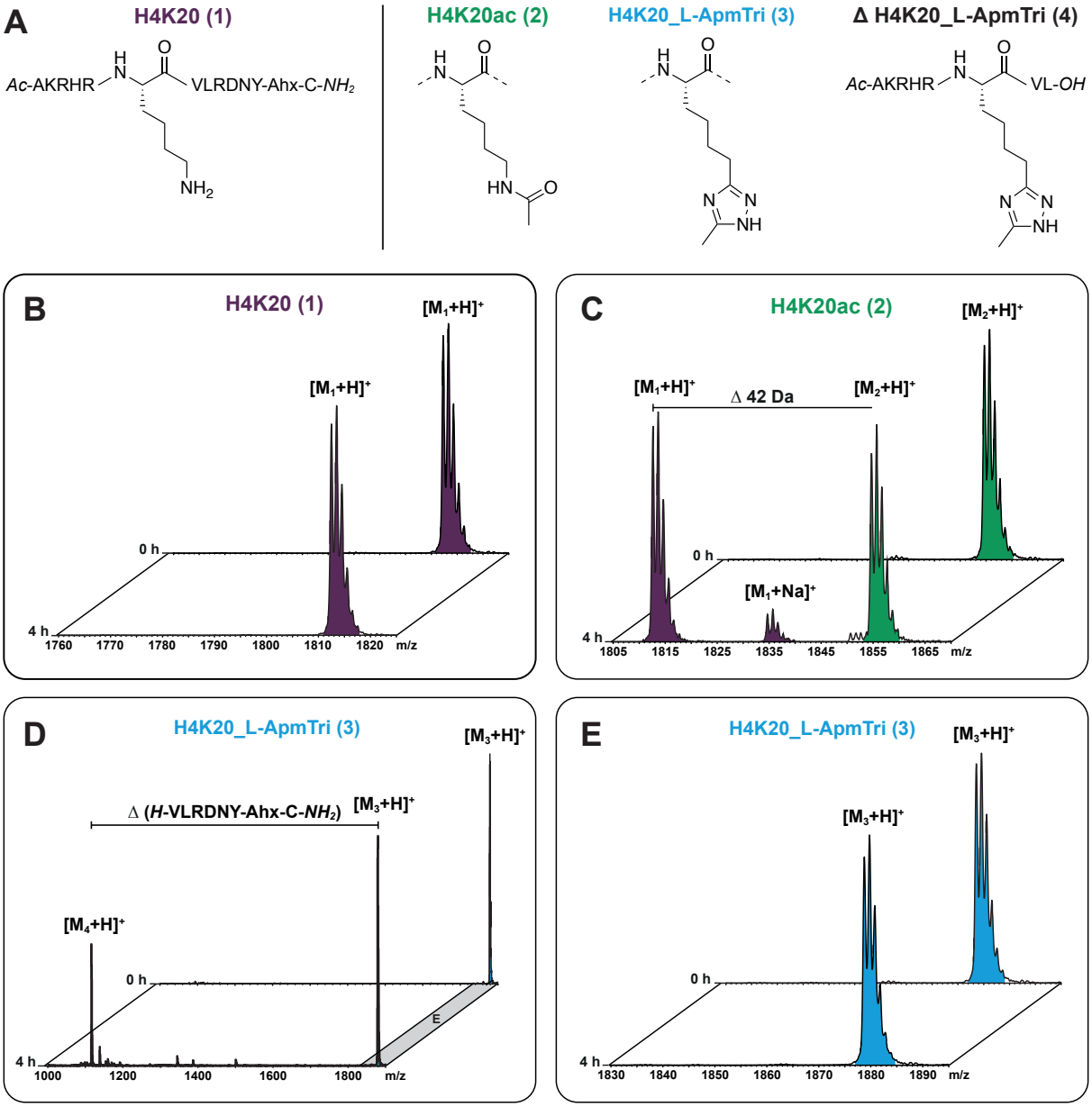


Figure 30: Stability results of the MALDI-MS based deacetylation assay with H4K20 peptides. A Structure of the H4K20 peptides H4K20 (purple), H4K20ac (green) and H4K20_L-ApmTri (blue) applied to the stability assay and the structure of an observed protease digestion product (Δ H4K20_L-ApmTri). **B**-**E** MALDI-MS spectra at the time points 0 h and 4 h. Assigned masses (M) are labeled with the number of the respective peptide. **E** shows the highlighted m/z range (gray) of the spectrum in **D**.

2.9 Genetic encoding of the non-canonical amino acid L-ApmTri

2.9.1 Initial investigation of an aminoacyl-tRNA synthetase and tRNA for L-ApmTri

In the following, L-ApmTri was genetically encoded using an amber codon suppression system. The concept of genetic code expansion allows site-specific incorporation of ncAAs at distinct positions during protein biosynthesis (1.4.2).
A pyrrolysyl-tRNA synthetase (PylRS) library screen was performed by *Christopher Reinkemeier* (group of Prof. Dr. Edward A. Lemke, Johannes Gutenberg University Mainz, Germany), identifying the wild type PylRS from *Methanosarcina mazei* as being capable of L-ApmTri incorporation in HEK293T cells. The plasmids used in the following experiments coding for the PylRS/$tRNA^{Pyl}$ pair and the reporter construct were provided by *Prof. Dr. Edward A. Lemke*.
After discovering L-ApmTri as substrate of the PylRS/$tRNA^{Pyl}$ pair, the incorporation efficiency and dependence on ncAA concentration was studied by fluorescence-based flow cytometry. The assay is based on expression of the amber suppression reporter iRFP-GFP^{TAG} in mammalian cells. The reporter is composed of the near-infrared fluorescent protein iRFP (iRFP713) [247] fused C-terminally to an amber mutant of EGFP (Y39TAG) [248, 249]. A nuclear localization signal (NLS) localizes the reporter to the nucleus and additional *FLAG*- and *His_6*-tags provide further options for detection and purification. A schematic diagram of the iRFP-GFP^{Y39TAG} reporter construct is shown in figure 31. Full-length iRFP-GFP was only expressed in presence of the PylRS/$tRNA^{Pyl}$ pair and by suppression of the amber stop codon with an encoded ncAA. Red fluorescence (iRFP) reported successful cell transfection with the reporter plasmid, whereas green fluorescence (EGFP) indicated the efficiency of amber suppression.
HEK293T cells were co-transfected with plasmids coding for the PylRS/$tRNA^{Pyl}$ pair and the iRFP-GFP^{Y39TAG} reporter construct. Transfected cells were either treated with 250 µM Nε-Boc-L-Lysine (BocK) or with different concentrations of L-ApmTri. BocK is a well-known substrate of wild type PylRS and was used as positive control in amber suppression experiments. Cells were analyzed by flow cytometry as described in 4.2.4.21 and by fluorescence microscopy (4.2.4.20). Representative flow cytometric analyses and fluorescent images of living HEK293T cells are shown in figure 31. Results of further control experiments are shown in figure A.7 of the appendix, including treatments with canonical lysine and additional concentrations of L-ApmTri.
Quantification of flow cytometric measurements was achieved by determining the GFP geometric mean fluorescence intensities (GMFI) of cells treated with BocK (250 µM) or L-ApmTri at concentrations between 0.5 mM and 10 mM. Single red positive cells (PC5.5-A^+) were transfected with the reporter plasmid, but showed no amber suppres-

sion, probably due to absence of the PylRS/tRNA^{Pyl} plasmid. Red and green positive cells (PC5.5-A$^+$/FITC-A$^+$) were expressing the full-length iRFP-GFP^{Y39TAG} protein, thus likely incorporated the ncAA into the mutated GFP.

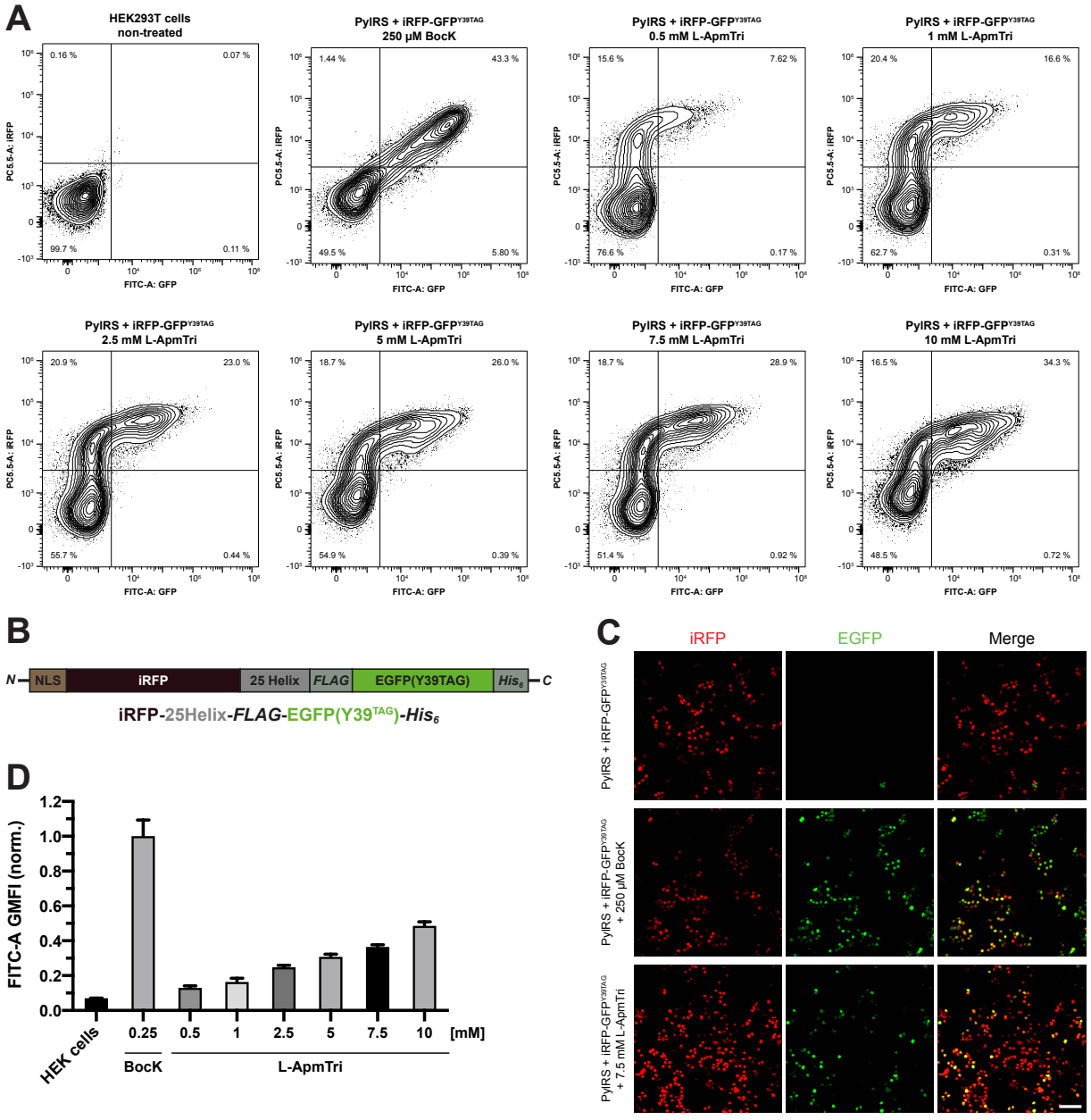


Figure 31: Flow cytometric and microscopic examination of L-ApmTri incorporation into a reporter construct using the wild type PylRS/tRNA^{Pyl} system. **A** Representative two-parameter contour plots showing near-infrared (iRFP) and green fluorescence (GFP) intensities for HEK293T cells expressing PylRS/tRNA^{Pyl} and iRFP-GFP^{Y39TAG}. Cells were treated with Boc-lysine (250 µM) or different concentrations of L-ApmTri. The gating strategy is shown in the appendix, figure A.6. **B** Schematic diagram of the iRFP-GFP^{Y39TAG} reporter protein construct. **C** Fluorescence microscopy images of living HEK293T cells transfected with plasmids coding for PylRS and iRFP-GFP^{Y39TAG} and treated with 250 µM BocK or 7.5 mM L-ApmTri. Scale bar, 100 µm. **D** Geometric mean fluorescence intensities (GMFI) calculated from the FITC-A channel (GFP) of flow cytometric analyses. Values are normalized to 250 µM BocK and are shown as mean ± s.d (n> 3).

Strongest expression of the full-length reporter (highest green fluorescence intensities) was found for cells incubated with BocK (250 µM). The efficiency of the L-ApmTri incorporation showed a strong concentration dependency. Cultures treated with 0.5 mM L-ApmTri showed only few FITC-A$^+$ cells, whereas incubation with 10 mM L-ApmTri led to substantially increased double positive cells. Considering the high demand of L-ApmTri at a concentration of 10 mM, the following experiments were conducted at 5 mM or 7.5 mM of L-ApmTri, which still showed robust amber suppression.
Treatment of non-transfected cells with up to 10 mM L-ApmTri was not impacting cell viability and fluorescence. Control experiments with cells expressing the iRFP-GFP^{Y39TAG} reporter in absence of PylRS showed red fluorescent cells due to the expression of iRFP and only few cells with weak GFP intensity, confirming efficient translation termination at the Y39TAG amber codon. The occurrence of FITC-A$^+$ cells without PylRS was probably caused by read-through of the amber stop codon by the cellular translation machinery (see appendix, figure A.7).
HEK293T cells transfected with PylRS and iRFP-GFP^{Y39TAG} plasmids were also analyzed by fluorescence microscopy. Transfected cells without ncAA exhibited nuclear localized red fluorescence, but no GFP signal. The addition of BocK (250 µM) or L-ApmTri (7.5 mM) led to red and green fluorescent cells, confirming the requirement of the ncAA for expression of the full-length reporter. The fluorescent images also showed higher efficiency for BocK incorporation when compared to L-ApmTri. Further fluorescent images of cells without co-transfection of PylRS are shown in figure A.7 of the appendix.

2.9.2 Establishment of an interaction assay based on the genetically encoded amino acid L-ApmTri

After genetic encoding of L-ApmTri the PylRS/tRNA^{Pyl} system was utilized to establish a binding assay for bromodomain BRD3(2) in a cellular approach. On the basis of the observed interaction between BRD3(2) and H4K20_L-ApmTri, two protein constructs were designed. Due to their potential application in fluorescence resonance energy transfer (FRET) experiments, the constructs are referred to as Donor1 (Don1) and Acceptor1 (Acc1).
Donor1 consists of the H4K20 binding sequence, where the lysine 20 codon is mutated to the amber stop codon (H4K20^{TAG}). The binding sequence is C-terminally fused to the fluorescent protein mCerulean3 [250]. An N-terminal *Strep*-tag and a C-terminal *His*$_6$-tag were added for detection and purification of the protein. Acceptor1 consists of the bromodomain BRD3(2) with fused EYFP [251] at the C-terminus, which forms a CFP/YFP FRET pair with mCerulean3. An additional *FLAG*-tag was fused C-terminally to the protein enabling detection and immunoprecipitation. Schematic diagrams of the designed protein constructs are shown in figure 32 and amino acid sequences are listed in the appendix, table B.2.

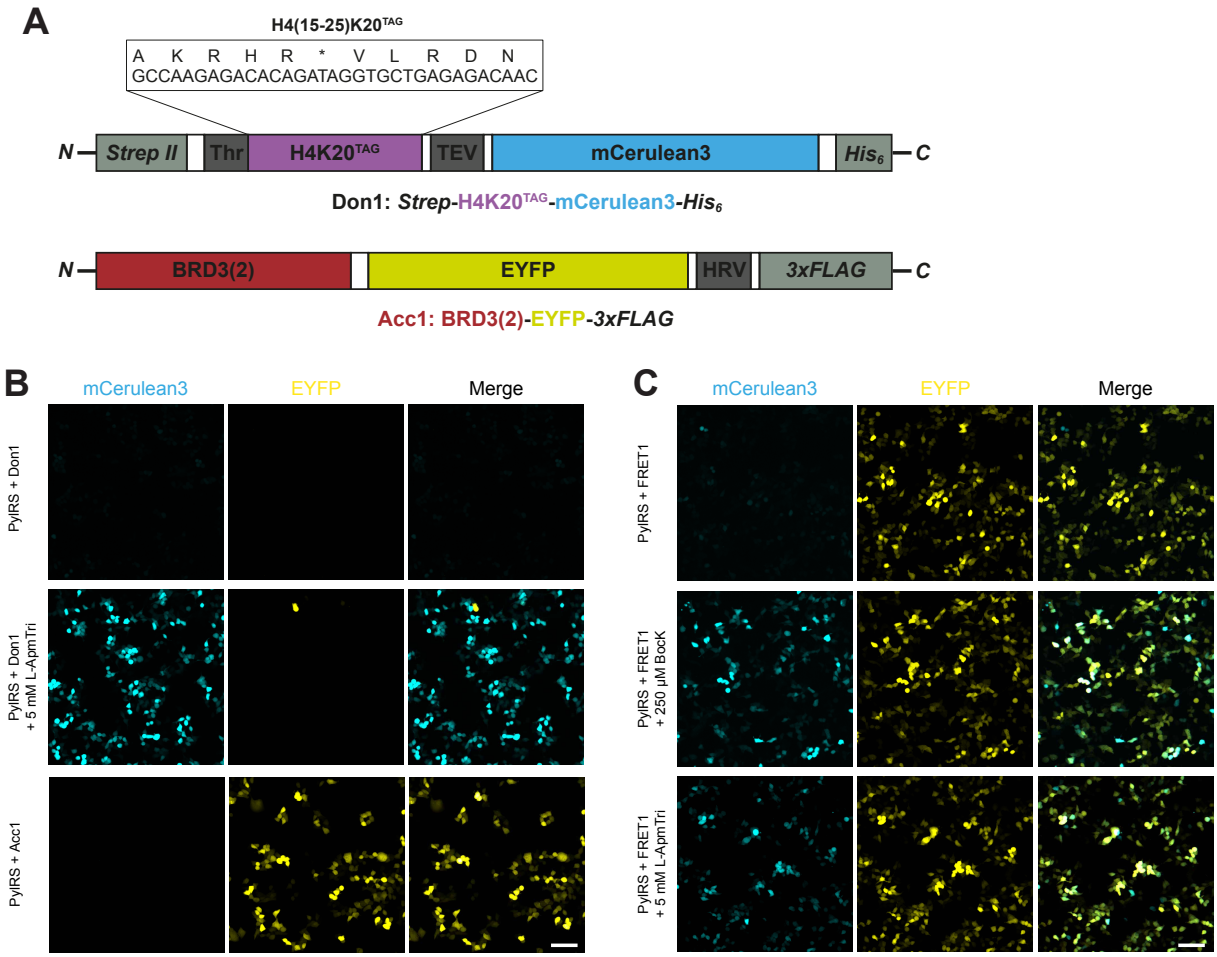


Figure 32: Expression of a Donor/Acceptor protein pair in HEK293T cells under incorporation of BocK and L-ApmTri at the H4K20 recognition site of BRD3(2). A Schematic diagrams of H4K20^{TAG}-mClover3 (Don1) and BRD3(2)-EYFP (Acc1) protein constructs. H4K20^{TAG} region (purple) is expanded showing DNA and amino acid sequence. Fluorophores are colored in their specific emission color, purification tags in light gray and protease cleavage sites (Thr: thrombin cleavage site; TEV: TEV protease cleavage site; HRV: HRV 3C protease cleavage site) in dark gray. Sequences are listed in the appendix, table B.2. **B, C** Fluorescence microscopy images of living HEK293T cells transfected with plasmids coding for PylRS and Don1/Acc1 protein constructs. The proteins were separately expressed (**B**) or co-expressed (**C**) using a bidirectional vector (FRET1) under treatment with 250 µM BocK or 5 mM L-ApmTri. Scale bar, 100 µm.

Synthetic and codon-optimized genes coding for Donor1 and Acceptor1 were cloned into the bidirectional vector pBI-CMV1, consisting of two independent expression sites. Expression vectors containing only one gene (Don1, Acc1) or both genes (FRET1) were cloned for individual or combined protein expression.

HEK293T cells were co-transfected with vectors coding for PylRS/tRNA^{Pyl} and Don1, Acc1 or FRET1. The ncAAs BocK (250 µM) and L-ApmTri (5 mM) were added after transfection as substrates for PylRS. Expressions were monitored by fluorescence microscopy and representative results are shown in figure 32. Additional control experiments for FRET1 expression in absence of PylRS are shown in figure A.7.

The C-terminal positioning of mCerulean3 to the amber codon site allowed verification of ncAA integration by fluorescence detection, since mCerulean3 was only expressed

after successful stop codon suppression. Cells transfected with PylRS and Don1 did not show any fluorescence for CFP or YFP, whereas addition of 5 mM L-ApmTri led to the expression of full-length Donor1 (H4K20^{TAG}-mCerulean3), visible in the CFP channel. Single expression of Acc1 (BRD3(2)-EYFP) resulted in yellow fluorescent cells, which did not show the CFP signal. After verification of appropriate expression of the individual constructs, cells were transfected with PylRS and FRET1 for co-expression of Donor1 and Acceptor1. Cells not supplemented with the ncAA showed only very weak CFP intensities. The addition of 250 µM BocK or 5 mM L-ApmTri resulted in cyan fluorescent cells, indicating incorporation of the ncAAs at the H4K20^{TAG} position. Fluorescent intensity for Donor1 in FRET transfected cells appeared weaker for L-ApmTri treated cells when compared to BocK treatment. The co-expression of Don1 and Acc1 with PylRS and ncAA also resulted in a reduced CFP Donor signal. All FRET1 transfected cells showed expression of the Acceptor1 protein BRD3(2)-EYFP, thus YFP expression was used as transfection control for the FRET1 plasmid.
In summary, the novel protein constructs were individually or simultaneously expressed in transfected HEK293T cells. Addition of the PylRS/tRNA^{Pyl} pair in prescence of BocK or L-ApmTri led to the incorporation of the respective ncAA into the H4K20 substrate sequence of BRD3(2).
The examination of the protein constructs Donor1 and Acceptor1 by fluorescence microscopy already indicated proper expression of the full-length proteins, but correct incorporation of L-ApmTri was further evaluated by immunoblot analysis. Therefore, HEK293T cells were transfected with Don1, Acc1 and FRET1 plasmids, co-transfected with the PylRS/tRNA^{Pyl} coding plasmid and treated with 250 µM BocK or 7.5 mM L-ApmTri. Cells were lysed 48 h after transfection and lysates were analyzed by western blot analysis (4.2.4.14). The Don1 protein H4K20^{TAG}-mCerulean3 was detected via the C-terminal *His$_6$*-tag and the Acc1 construct BRD3(2)-EYFP by incubation with an anti-*FLAG* antibody, targeting the C-terminally fused *FLAG*-tag. Cyclophilin B was detected in the different lysates as loading control. Representative immunoblots are shown in figure 33. As expected, Donor1 and Acceptor1 could not be detected in lysates of non-transfected cells. Transfection with vectors coding for Acceptor1 (Acc1, FRET1) led to expression of the full-length construct. Donor1 could be detected in lysates of cells transfected with PylRS and Don1 or FRET1 for BocK as well as for L-ApmTri treated cells. However, strong variations in expression levels were observed. Strongest western blot signals were detected for Donor1 in presence of BocK, whereas the co-expression of Acceptor1 was reducing the expression of Donor1. The same tendency was observed for Donor1 expression under incorporation of L-ApmTri. In general, protein yields for L-ApmTri were reduced when compared to BocK treatment and an additional band of a truncation product was observed. Since the truncated protein was detected by an anti *His$_6$*-tag antibody, the truncation site could be assigned to the N-terminus of the construct. Despite the lower protein levels for L-ApmTri-containing Donor1, detectable amounts of the full-length H4K20^{TAG}-mCerulean3 construct were expressed. Therefore, the established assay appeared to be suitable for investigation of BRD3(2) binding.

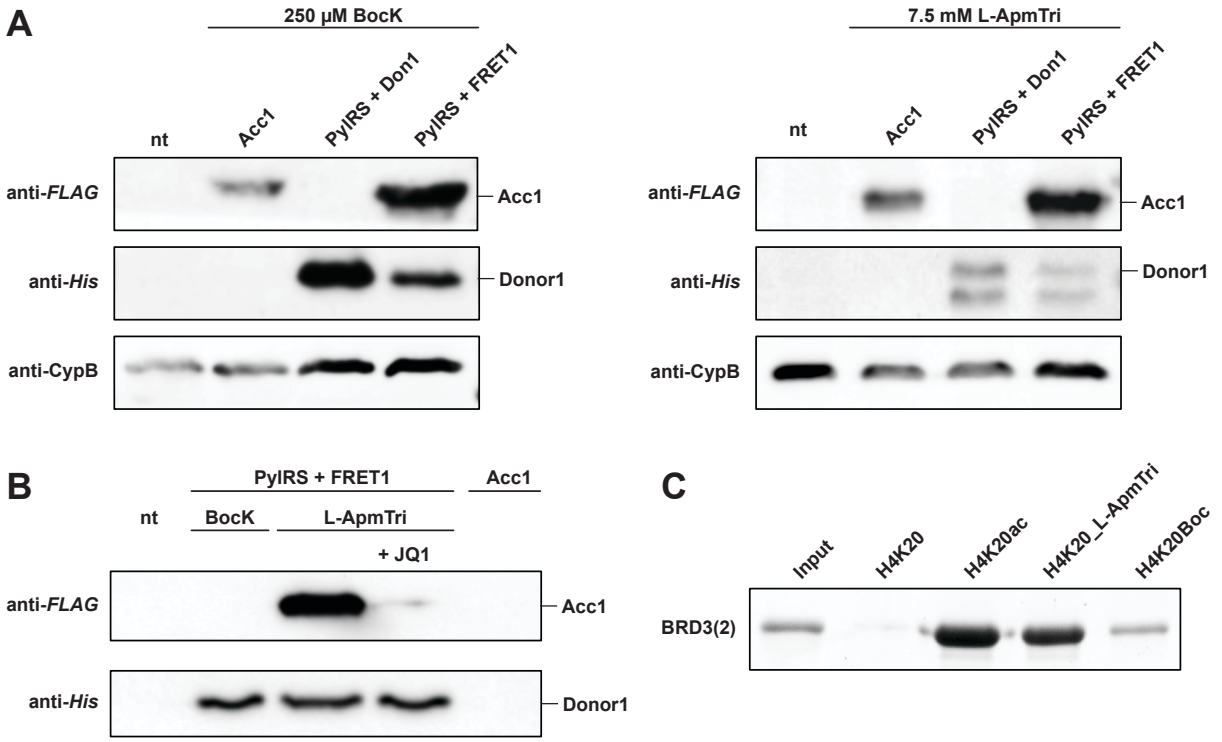


Figure 33: Immunoblot analysis of Donor1 and Acceptor1 protein expression and co-precipitation via *Strep*-tag of Donor1. A Immunoblot analysis of HEK293T cell lysates after single expression of Donor1 and Acc1 or co-expression (FRET1) using the PylRS amber suppression system. Non-treated (nt) cells were lysed and loaded as negative control. Concentrations of 250 µM BocK and 7.5 mM L-ApmTri were added for expression. BRD3(2) was detected with an anti-*FLAG*, Donor1 with an anti-*His* antibody and Cyclophilin B was used as loading control. **B** The lysates were used as input for co-precipitating (CoP) Don1 and Acc1 using the *Strep*-tag fused to Don1. FRET1/L-ApmTri lysate was further treated with 1 µM (+)-JQ1 prior to CoP. **C** Representative Coomassie-stained SDS-PAGE of a pull-down assay with BRD3(2) and H4K20 peptide probes.

In this experimental approach BocK was used as control, since BocK did not bind recombinant BRD3(2)-TurboYFP when incorporated into a H4K20 peptide probe (H4K20Boc, see scheme 5). A representative Coomassie-stained SDS-PAGE of the pull-down experiment is shown in figure 33.

In the following, FRET1 was expressed in presence of the PylRS/tRNA^{Pyl} pair and BocK or L-ApmTri. Co-precipitation (CoP) was performed with lysed HEK293T cells by trapping Donor1 via the *Strep*-tag label using immobilized Strep-Tactin (4.2.4.19). Lysate of FRET1 with L-ApmTri was further treated with 1 µM of the small molecule inhibitor JQ1 prior to CoP. Eluates were tested for Donor1 and Acceptor1 by western blot analysis. Inequal expression levels of Donor1 with either BocK or L-ApmTri were compensated by adjusting the sample amount loaded to obtain equal Donor1 signal intensities. A representative immunoblot analysis of the CoPs is shown in figure 33.

In eluates of the non-transfected control and Acceptor1 single expressing cells, Donor1 and Acceptor1 could not be detected after CoP, excluding unspecific binding to the Strep-Tactin matrix. In eluates of precipitations with lysates from FRET1 expressing cells, Donor1 was detected, proving efficient trapping of the expressed proteins by Strep-

Tactin. Furthermore, co-precipitation of Acceptor1 in the FRET1/L-ApmTri probe was detected, demonstrating the interaction of BRD3(2) to the L-ApmTri amino acid. Treatment of the same FRET1/L-ApmTri lysate with JQ1 inhibited bromodomain binding to Donor1 and resulted in a weaker signal for Acceptor1 after CoP.
In summary, the PylRS/tRNA^{Pyl} system in HEK293T cells was suitable for the incorporation of L-ApmTri into the H4K20 sequence as part of the Donor1 protein construct. Combined expression with the bromodomain led to binding of BRD3(2) to the L-ApmTri-containing protein, verified by co-precipitation of both proteins.

2.10 Genetic encoding of the non-canonical amino acid AsuHd

2.10.1 Initial screenings for incorporation of AsuHd in proteins using the iRFP-GFP reporter

After successful genetic encoding of L-ApmTri the established hydroxamate-containing non-canonical amino acid AsuHd was also genetically encoded in order to generate a new tool for investigating histone deacetylases. A pyrrolysyl-tRNA synthetase library screen for AsuHd incorporation was performed by *Christopher Reinkemeier* (group of Prof. Dr. Edward A. Lemke, Johannes Gutenberg University Mainz, Germany) uncovering the mutated PylRS^{AF}/tRNA^{Pyl} pair. PylRS^{AF} is derived from the wild type PylRS from *M. mazei*, engineered for incorporation of amino acids with bulky side-chains. The introduction of two mutations (Y306A and Y384F) resulted in the PylRS^{AF} enzyme [214]. The plasmid coding for the PylRS^{AF}/tRNA^{Pyl} pair was provided by *Prof. Dr. Edward A. Lemke*.
The amber suppression efficiency and dependency on AsuHd concentration was further studied using the fluorescence-based flow cytometric assay, already established for the L-ApmTri experiments. The general concept of the assay is described in 2.9.1. HEK293T cells were co-transfected with vectors coding for the PylRS^{AF}/tRNA^{Pyl} system and for the reporter iRFP-GFP^{Y39TAG}. Like amber suppression experiments with wild-type PylRS, BocK was utilized as known and well-accepted substrate of PylRS^{AF} serving as positive control. Cells were treated with lysine (5 mM), BocK (250 µM) and various concentrations of AsuHd (0.5 mM to 5 mM). Fluorescence intensities of iRFP (PC5.5-A) and GFP (FITC-A) were analyzed 24 h after transfection by flow cytometry. Representative contour plots are shown in figure 34.
As expected, non-transfected samples did not show any FITC-A or PC5.5-A positive cells. Transfection with iRFP-GFP^{Y39TAG} led to iRFP expression and also a small number of cells showed weak GFP expression. Treatment of PylRS^{AF}/iRFP-GFP^{Y39TAG} transfected cells with 5 mM Lysine did not significantly increase the number of FITC-A$^+$ cells, confirming orthogonality of the PylRS^{AF}/tRNA^{Pyl} pair. Treatment with BocK as well as AsuHd led to expression of the full-length reporter protein, represented by FITC-A$^+$/PC5.5-A$^+$

positive cells. The incorporation of AsuHd by PylRS^{AF} showed weak concentration dependency between the tested concentrations of 0.5 mM to 5 mM AsuHd, indicated by only a slightly increased percentages of double positive cells at higher concentrations. Based on this and on further experiments performed by *Annika Kühn* [235], the AsuHd concentration for further experiments was set to 1 mM, taking reduced cell viability at higher AsuHd concentrations into account.

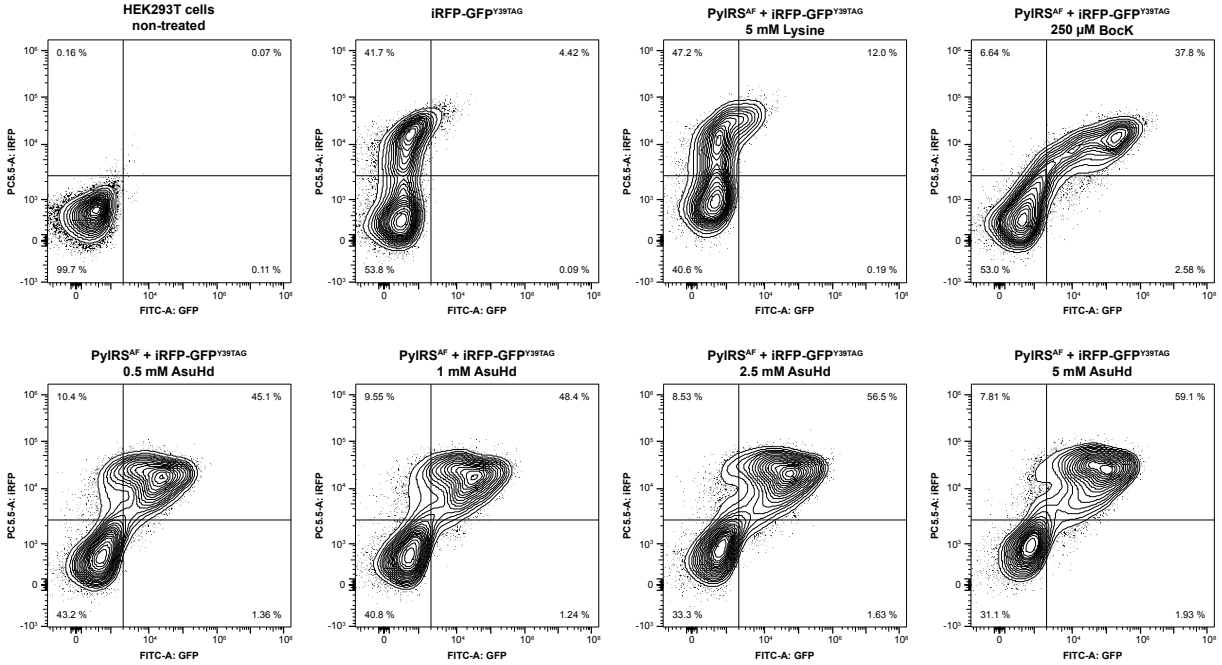


Figure 34: Flow cytometric analysis of AsuHd incorporation into a reporter construct using the PylRS^{AF}/tRNA^{Pyl} system. Representative two-parameter contour plots showing near-infrared (iRFP) and green fluorescence (GFP) intensities for HEK293T cells expressing PylRS^{AF}/tRNA^{Pyl} and iRFP-GFP^{Y39TAG}. Cells were treated with lysine (5 mM), Boc-lysine (250 µM) or different concentrations of AsuHd ranging from 0.5 mM to 5 mM. Gating strategy is shown in the appendix figure A.6.

2.10.2 Establishment of an HDAC interaction assay based on genetically encoded AsuHd

The successful integration of AsuHd into the reporter iRFP-GFP^{Y39TAG} provided the basis for the establishment of a cell-based interaction assay following a similar concept as described previously for L-ApmTri (see 2.9.2). Two different acetyl-lysine substrate sequences were investigated in this assay: p53(377-388)K382ac as a known HDAC1 substrate [66, 252] and K40ac of α-tubulin(32-48), a described deacetylation site of HDAC6 [253]. In addition, a probe with minimal sequence context (mini^{TAG}), consisting of an AsuHd flanked by glycines was used. The acetylation site codons of p53 and α-tubulin were replaced with an amber stop codon for site-specific incorporation of AsuHd, resulting in the p53^{TAG} and α-tubulin^{TAG} binding motifs. Three protein constructs were designed, consisting of the HDAC substrate sequence with C-terminally fused mClover3 [254] as fluorescent label. A nuclear export signal (NES) [255] was fused to the proteins

as well. An N-terminal *Spot*-tag and a C-terminal *Myc*-tag were added for identification and purification of the fusion proteins. A fourth protein construct was designed, consisting of the catalytic domain of HDAC1 with C-terminally fused mRuby3, which forms a green/red FRET pair with mClover3 [254, 256]. An NES was added for cytoplasmic localization of HDAC1. For purification and detection, an N-terminal *Twin-Strep*-tag [257] and a C-terminal *3xFLAG*-tag were fused to the HDAC1-mRuby3 protein. Due to their potential application in FRET experiments, $p53^{TAG}$-mClover3 is referred to as Donor2 (Don2), $mini^{TAG}$ and α-$tubulin^{TAG}$ as Donor3 (Don3) and Donor4 (Don4), respectively. HDAC1-mRuby3 is referred to as Acceptor2 (Acc2). Schemes of Donor2-4 are shown in figure 35 and of Acceptor2 in figure 37. Amino acid sequences are listed in the appendix table B.2.

Synthetic and codon-optimized genes coding for Donor2-4 were cloned into the bidirectional vector pBI-CMV1, which allowed co-expression of HDAC1 with the Donor constructs in further experiments. In a first step, the efficiency of AsuHd incorporation by the $PylRS^{AF}$/$tRNA^{Pyl}$ pair into the novel Donor proteins was validated. The fluorescent label, positioned C-terminally of the amber codon site, allowed tracking of efficient amber suppression by fluorecence detection.

HEK293T cells were co-transfected with plasmids coding for the $PylRS^{AF}$/$tRNA^{Pyl}$ pair and for the three Donor constructs. Cells were incubated without ncAA, with 250 µM BocK and 1 mM AsuHd for 24 h and analyzed by flow cytometry and fluorescence microscopy. Fluorescence intensities for mClover3 and mRuby3 were determined using FITC-A and PE-A filter settings. FITC-A^+ and PE-A^+ cells were defined by the gating strategy illustrated in the appendix figure A.6. Representative results are shown in figure 35.

The combined expression of Donor proteins and the $PylRS^{AF}$/$tRNA^{Pyl}$ pair in absence of an ncAA showed a small percentage of FITC-A^+ cells for $mini^{TAG}$ and α-$tubulin^{TAG}$ when compared to samples treated with BocK or AsuHd. Despite the absence of a ncAA, non-treated cells expressing $p53^{TAG}$ developed a high percentage of $FITC^+$ cells, indicating mClover3 expression independently of amber suppression. The p53(377-388) sequence contains an additional methionine at position 84 in close proximity to the amber codon. It appears plausible, that the expression of a truncated $p53^{TAG}$-mClover3 protein fragment led to the green fluorescence signal, resulted from miss-initiation at the second methionine codon. Treatment with BocK still led to a detectable increase in mClover3 expression. In case of $mini^{TAG}$-mClover3 and α-$tubulin^{TAG}$-mClover3, a defined $FITC^+$ cell population was detected, indicating successful AsuHd incorporation. However, green fluorescence intensities of AsuHd treated cells and consequently expression levels of Donor proteins were significantly lower when compared to BocK treated cells.

This was also observed in the fluorescent images of $PylRS^{AF}$/$tRNA^{Pyl}$ and α-$tubulin^{TAG}$-mClover3 (Don4) expressing cells. Non-treated cells did not show detectable fluorescence in the green and red channels, whereas addition of BocK and AsuHd led to green fluorescent cells with substantially lower intensities for AsuHd.

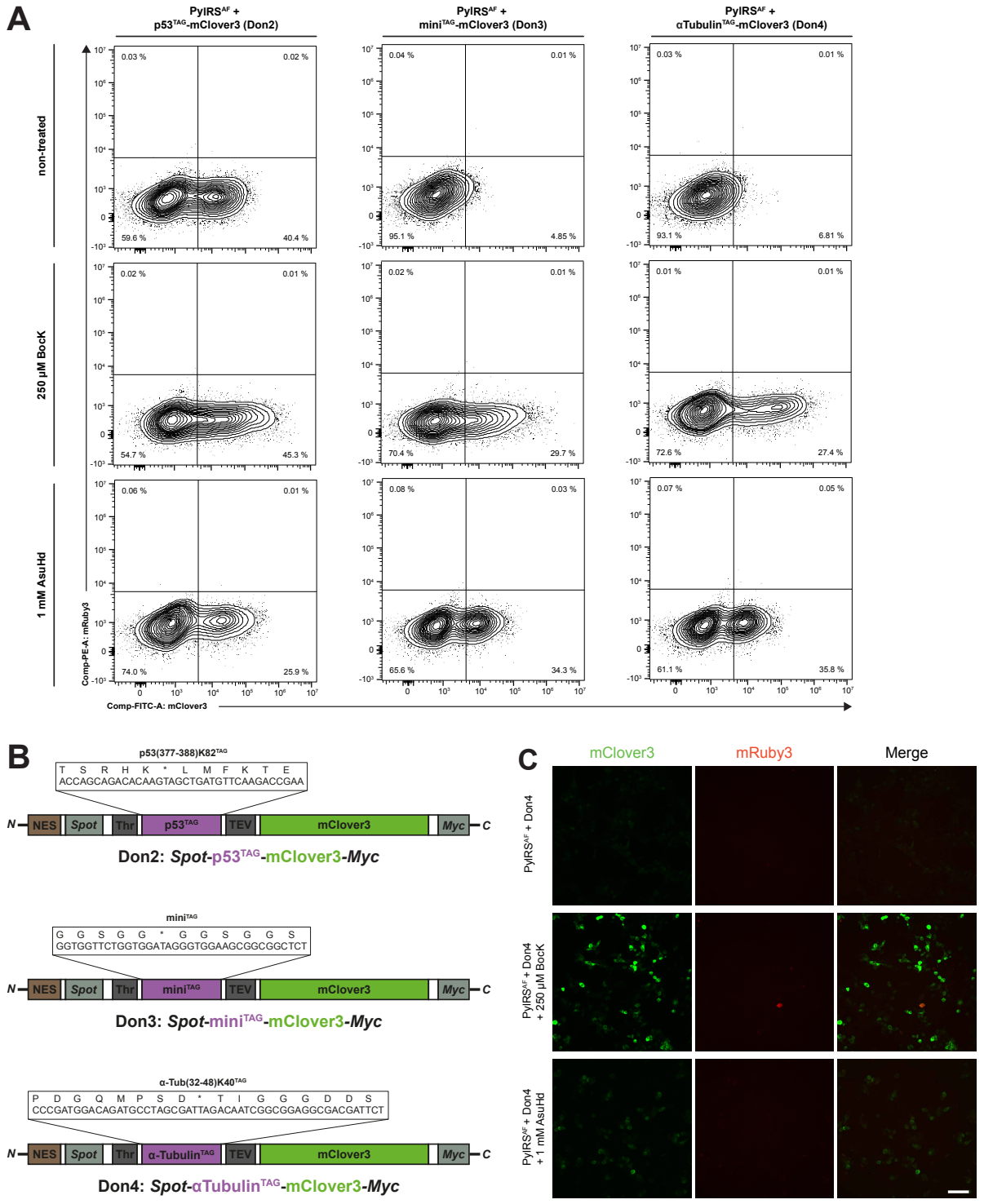


Figure 35: Flow cytometric and microscopic examination of AsuHd incorporation into the Donor protein constructs using the PylRS^{AF}/tRNA^{Pyl} system. A Representative two-parameter contour plots showing red (mRuby3) and green fluorescence (mClover3) intensities for HEK293T cells expressing PylRS^{AF}/tRNA^{Pyl} and Donor2-4. Cells were treated with 250 µM Boc-lysine or 1 mM AsuHd. The gating strategy is shown in the appendix, figure A.6. **B** Schematic diagrams of Donor2-4 protein constructs. HDAC binding regions (purple) are expanded showing DNA and amino acid sequences. mClover3 is colored in green, purification tags in light gray, NES in brown and protease cleavage sites (Thr: thrombin cleavage site; TEV: TEV protease cleavage site) in dark grey. Sequences are listed in the appendix, table B.2.**C** Fluorescence microscopy images of living HEK293T cells transfected with plasmids coding for PylRS^{AF}/tRNA^{Pyl} and Donor4 and treated with 250 µM BocK or 1 mM AsuHd. Scale bar, 100 µm.

In conclusion, the PylRS^{AF}/tRNA^{Pyl} pair is suitable for the incorporation of BocK and AsuHd into HDAC recognition sequences as part of the Donor3-4 protein constructs. Full-length expression of Don2 could not be verified by the flow cytometric assay, but increased fluorescent intensity for BocK treated cells suggested functional expression.

2.10.2.1 Co-IP of endogenous HDACs using proteins containing HDAC substrate sequences

Prior to co-expression of the HDAC1-mRuby3 construct with the Donor proteins, potential binding of endogenous HDACs to the AsuHd-containing recognition sites p53^{TAG}, mini^{TAG} and α-tubulin^{TAG} was analyzed. The PylRS^{AF}/tRNA^{Pyl} pair was expressed in combination with Donor2-4 in HEK293T cells, supplemented with 250 µM BocK as control or 1 mM AsuHd. After 24 h, cells were lysed and co-immunoprecipitation experiments (Co-IPs) were carried out by *Annika Kühn* [235]. Immunoprecipitation of Donor proteins was conducted by targeting the N-terminal *Spot*-Tag with Spot-Trap resin, composed of Spot-Tag binding nanobodies immobilized on agarose beads. Input samples were collected prior to immunoprecipitation and analyzed together with Spot-Trap eluates by immunoblotting. Donor proteins were detected using a Spot-label nanobody against the *Spot*-tag. Endogenous HDACs (class I: HDAC1, HDAC2, HDAC3, HDAC8; class IIa: HDAC4; class IIb: HDAC6) were detected with the respective anti-HDAC primary antibodies. The western-blot analysis of the Co-IP experiments are shown in figure 36.

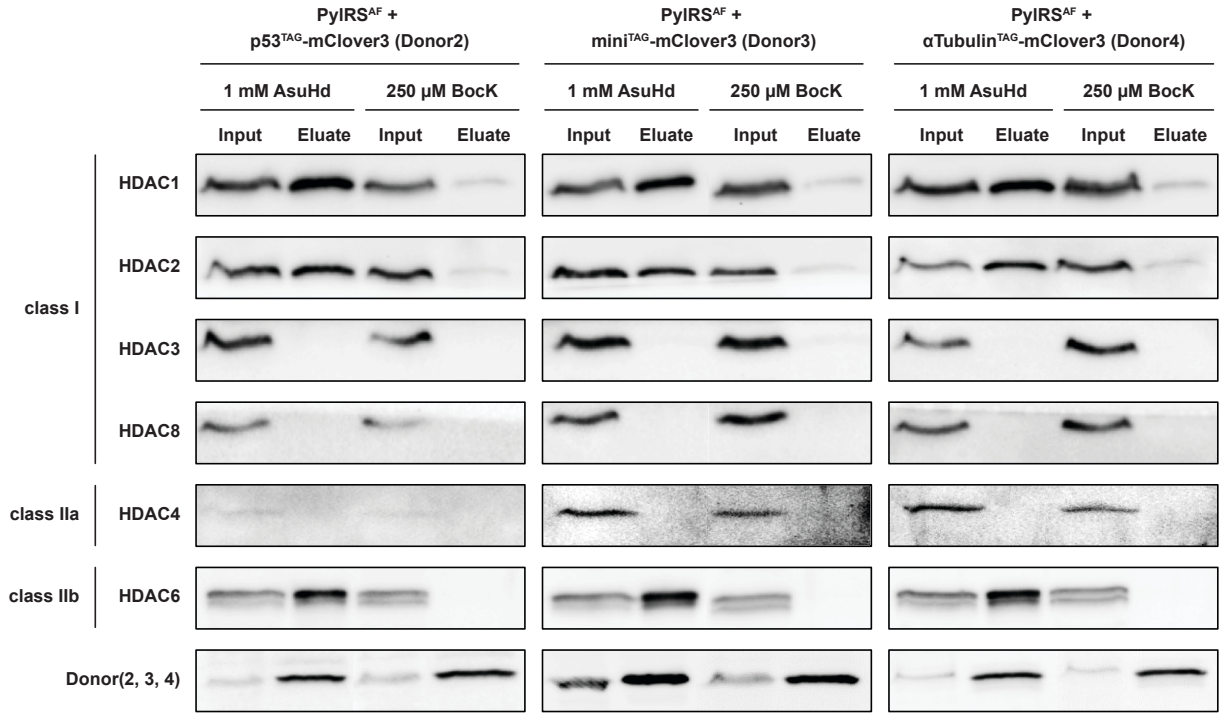


Figure 36: Immunoblot analysis of Co-IPs targeting *Spot*-tagged Donor proteins containing BocK or AsuHd. Donor proteins were expressed in HEK293T cells using the PylRS^{AF}/tRNA^{Pyl} system and treatment with 250 µM BocK or 1 mM AsuHd. Inputs were taken prior to incubation with the Spot-Trap.

Each of the HDAC substrate constructs (Donor2-4) was detected after Co-IP in the eluates, demonstrating successful expression of full-length proteins as well as efficient enrichment on Spot-Trap agarose beads. The HDAC binding patterns were very similar for the three expressed recognition sites p53, mini and α-tubulin. Co-immunoprecipitation was achieved for HDAC1, HDAC2 and HDAC6 in the lysates of AsuHd treated cells, whereas controls with incorporated BocK showed substantially reduced enrichment of HDACs. Binding of HDAC3, HDAC4 and HDAC8 was not detected.

2.10.2.2 Co-expression of proteins containing HDAC binding sequences and HDAC1-mRuby3

Combined expression of Donor and the HDAC1-mRuby3 acceptor construct (described in 2.10.2) was achieved by cloning the acceptor gene into the bidirectional vectors of Don2-4 for co-expression of both proteins, resulting in the vectors FRET2 ($p53^{TAG}$-mClover3/HDAC1-mRuby3), FRET3 ($mini^{TAG}$-mClover3/HDAC1-mRuby3) and FRET4 (α-$tubulin^{TAG}$-mClover3/HDAC1-mRuby3). HEK293T cells were transfected with FRET2-4 and $PylRS^{AF}/tRNA^{Pyl}$ vectors and cultivated for 24 h in absence of an ncAA or treated with 250 µM BocK or 1 mM AsuHd. Expression was determined by fluorescence intensities in flow cytometry and fluorescence microscopy as described in 2.10.2. Representative contour plots of the flow cytometric results and fluorescence images of FRET4 expressing cells are shown in figure 37.
Similar to single Don2 expression, FRET2 ($p53^{TAG}$-mClover3/HDAC1-mRuby3) expression in absence of an ncAA led to double positive (FITC-A^+/PE-A^+) cells. FRET2 expressing cells produced only a low percentage of single PE-A^+ cells, confirming the previous observation of amber suppression-independent mClover3 expression in Donor2. Expression of FRET3 ($mini^{TAG}$-mClover3/HDAC1-mRuby3) and FRET4 (α-$tubulin^{TAG}$-mClover3/HDAC1-mRuby3) showed the expected fluorescence signals: Without ncAA, only PE-A^+ cells could be detected, while addition of BocK or AsuHd led to an increased population of green fluorescent cells, indicating effective amber suppression. The amber suppression efficiency for FRET expression was significantly reduced for AsuHd when compared to BocK and also reduced when compared to single Donor expression experiments.
Fluorescent microscopy images showed the cytoplasmic localization of Donor4 and Acceptor2, resulting from the NES sequences fused to the proteins. Fluorescence microscopy images also confirmed the expression pattern for mClover3 and mRuby3 found in the flow cytometric assay. HDAC1-mRuby3 was expressed in all transfected cells. Higher green fluorescence (mClover3) was observed for cells treated with BocK when compared to Donor4 expression in prescence of AsuHd.
The general principle of combined expression of AsuHd-containing HDAC substrate sites and HDACs of interest seemed to be working, however the expression levels for the Donor proteins were very low. Co-immunoprecipitation experiments of Don4 and HDAC1 suggested an interaction between both proteins, but further experiments are required in order to verify these findings [235].

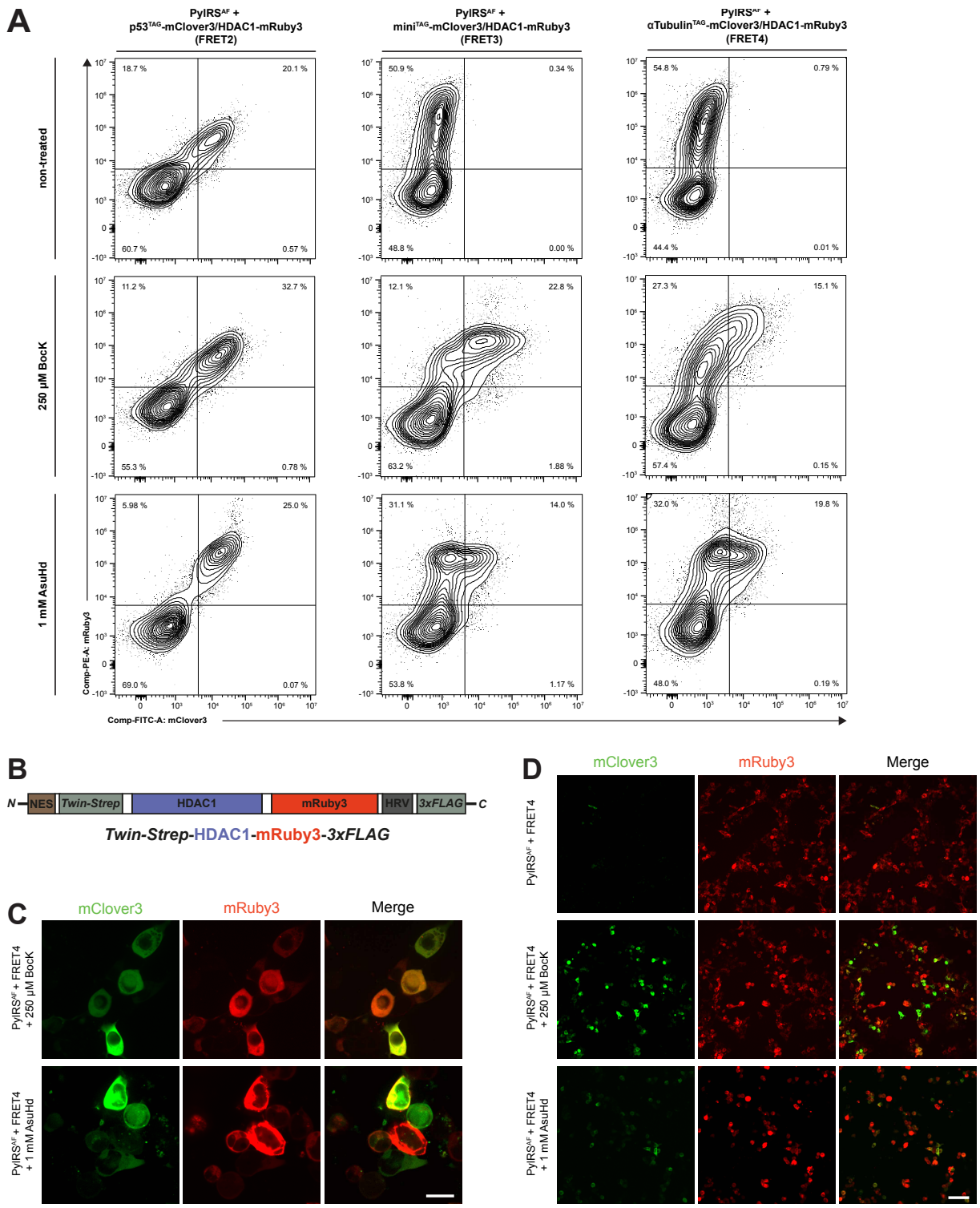


Figure 37: Flow cytometric and microscopic examination of AsuHd incorporation in cells transfected with FRET plasmids. A Representative two-parameter contour plots showing red (mRuby3) and green fluorescence (mClover3) intensities for HEK293T cells expressing PylRS^{AF}/tRNA^{Pyl} and FRET2-4. Cells were treated with 250 µM Boc-lysine or 1 mM AsuHd. The gating strategy is shown in the appendix, figure A.6. **B** Schematic diagram of the acceptor protein constructs. HDAC1 is colored in blue, mRuby3 in green, purification tags in light gray, NES in brown and HRV 3C protease cleavage site (HRV) in dark grey. Amino acid sequence is shown in the appendix table B.2. **C, D** Fluorescence microscopy images at 40x and 10x magnification of living HEK293T cells transfected with plasmids coding for PylRS^{AF}/tRNA^{Pyl} and FRET4, followed by treatment with 250 µM BocK or 1 mM AsuHd. Scale bars, 40x: 20 µm; 10x: 100 µm.

2.11 Semisynthesis of designer nucleosomes for characterization of HDAC binding

So far, HDAC research with the hydroxamate amino acid AsuHd was restricted to short synthetic peptide probes. However, lysine deacetylation is a highly regulated process, where large protein complexes are involved. Therefore, HDAC recruitment is likely influenced by various protein-protein interactions, especially on level of chromatin [258, 259]. The application of AsuHd in context of the chromatin structure requires a new approach for the incorporation of AsuHd into nucleosomal context.
The preparation of H3 ligation ready nucleosome core particles (NCPs) and their usage to generate semi-synthetic designer nucleosome libraries via sortase-mediated ligation (SML) was established by *Diego Aparicio Pelaz* and *Henriette Mahler* [207, 208, 260]. In brief, NCPs were obtained by wrapping the 147 bp long 601 widow DNA sequence [261] around recombinant histone octamers. The DNA sequence was extended by a cyanine 5 (Cy5) fluorescent label, a biotin tag for immobilization and three different restriction sites (HindIII, EcoRI and EcoRV) for release of immobilized nucleosomes. By inserting N-terminal truncated Δ(1-32)H3 histones (ΔH3) into the assembling process, ligation-ready NCPs were obtained. NCPs used in the following experiments were prepared by *Zhadyra Yerkesh* (group of *Prof. Dr. Wolfgang Fischle*, King Abdullah University of Science and Technology, Saudi Arabia).

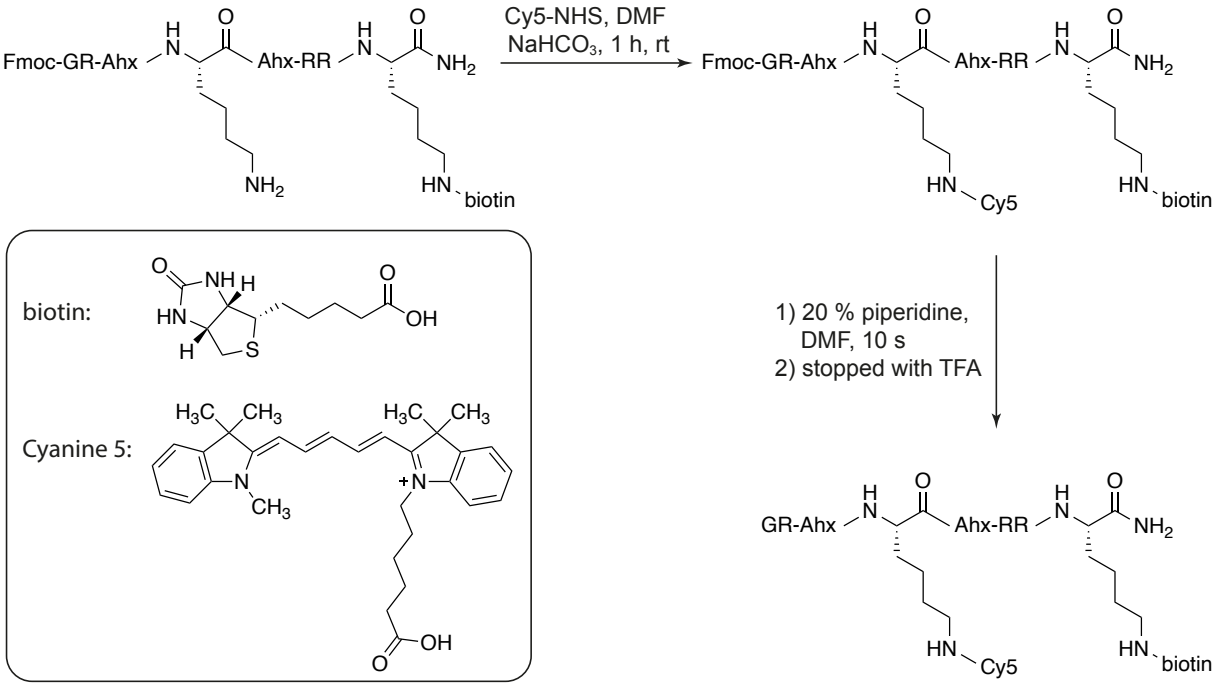
2.11.1 Synthesis of histone H3 depsipeptides

First, histone H3 derived peptides were synthesized for SML. Previous nucleosomal SML reactions showed highest ligation yields of ΔH3 with semi-synthetic full-length H3 by Sortase A with depsipeptides containing the LPATG motif. The synthesized H3 depsipeptides consisted of the first 33 residues of Histone 3 with an L29A mutation and an ester bond linking the threonine residue with glycolic acid in the sorting motif [262]. Selected histone H3 lysine acetylation sites were replaced with AsuHd (H3K4 [263], H3K9 [264] H3K27 [265]). Synthesis of the building block Fmoc-Asu(NH-OTrt)-OH is described in 2.2. An additional unmodified H3 depsipeptide (H3 unmod.) was added as binding control to the set of H3 depsipeptides (H3 unmod., H3K27AsuHd, H3K9AsuHd and H3K4AsuHd). Peptide sequences and purities are listed in table 4.

2.11.2 Synthesis of a Biotin-linker peptide as SML substrate

Pull-down experiments at NCPs need to be compared to binding patterns of H3 peptides lacking the nucleosomal context in order to study the impact of the nucleosomal structure. Therefore, a biotin-linker peptide (linker peptide) was synthesized as replace-

ment of the NCPs, consisting of an N-terminal glycine for SML and two lysine residues, modified with a Cy5 label and a biotin tag for immobilization. Scheme 6 illustrates the site-specific installation of the Cy5 label.



Scheme 6: Synthesis of the biotin-linker peptide in solution.

Lys(biotin) was incorporated during SPPS and after assembly, the peptide was cleaved off the resin with an N-terminal Fmoc protection group. Cy5 was coupled to the unprotected lysine residue in solution via an amine-reactive NHS ester, followed by deprotection of the Fmoc group. Due to low stability of Cy5 against piperidine, the deprotection was stopped by addition of TFA after 10 s. The final biotin-linker peptide was obtained after HPLC purification. The biotin-linker peptide sequence and obtained purity is shown in table 4.

Table 4: Synthesized peptides for H3AsuHd-NCP pull-down assays. Ester bonds are marked with (°). Peptide purities were determined by integration of UV absorption in analytical HPLC chromatograms (see appendix figure C.44).

Peptide	Sequence	Purity
Biotin-linker peptide	GR-Ahx-K(Cy5)-Ahx-RR-K(biotin)	98.3 %
H3(1-33) unmod.	ARTKQTARKSTGGKAPRKQLATKAARKSLPAT°G	94.1 %
H3(1-33)K27AsuHd	ARTKQTARKSTGGKAPRKQLATKAAR-**AsuHd**-SLPAT°G	94.2 %
H3(1-33)K9AsuHd	ARTKQTAR-**AsuHd**-STGGKAPRKQLATKAARKSLPAT°G	94.2 %
H3(1-33)K4AsuHd	ART-**AsuHd**-QTARKSTGGKAPRKQLATKAARKSLPAT°G	93.5 %

2.11.3 Sortase-mediated ligations of H3 depsipeptides

Sortase-mediated ligations were carried out on solid support, allowing simple removal of excess reaction components. Ligation-ready NCPs and the biotin-linker peptide were immobilized on streptavidin-coated magnetic beads by the biotin handles on DNA and peptide. The fluorescent Cy5 label was used to quantify and adjust immobilization input concentration in order to immobilize equal amounts of biotin-linker peptide and NCPs (see 4.2.4.22).

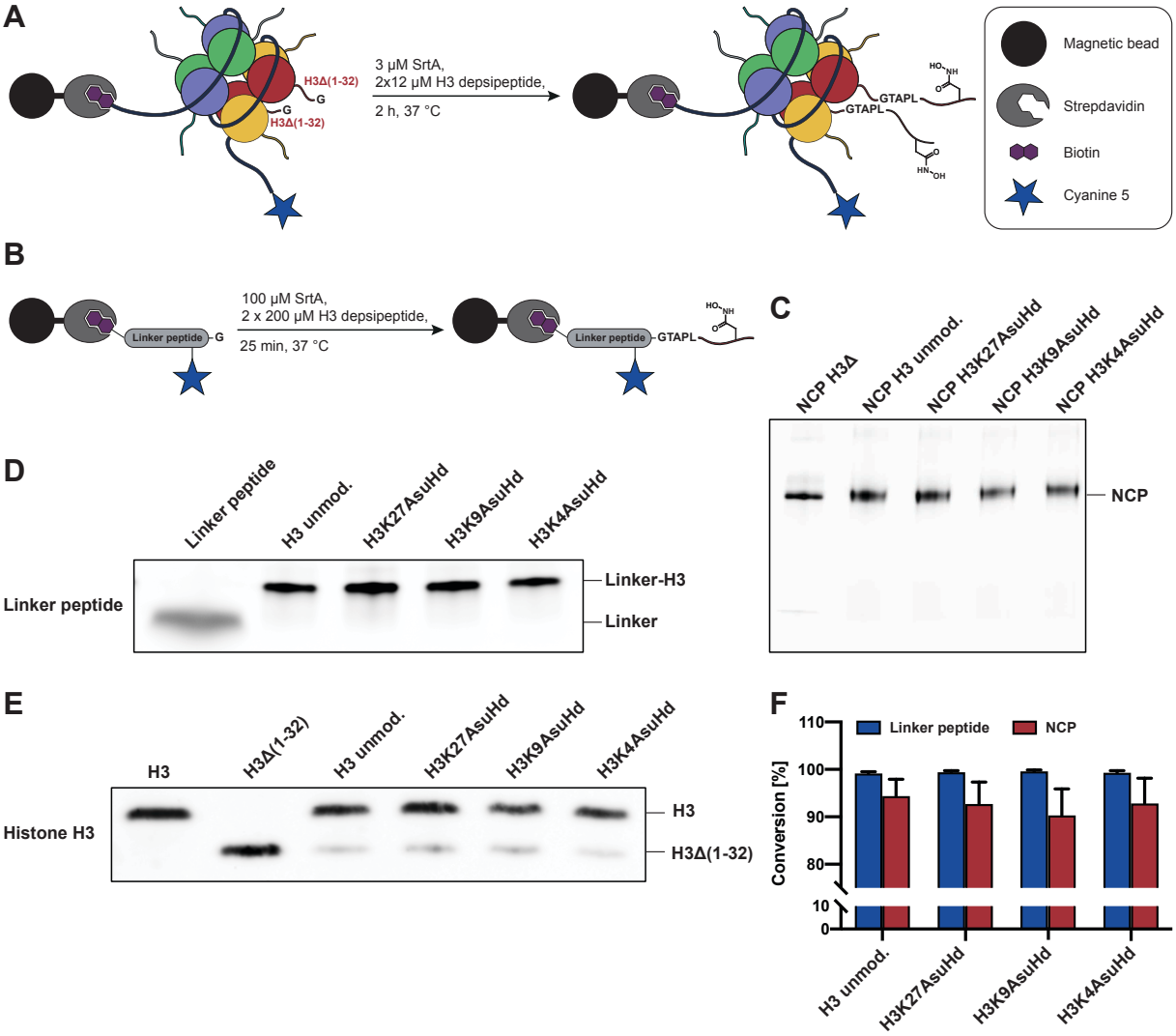


Figure 38: Results of sortase-mediated ligation of H3 depsipeptides and NCPs. **A** Reaction scheme illustrating SrtA-mediated ligation of H3 depsipeptides to immobilized ΔH3 NCPs. **B** Reaction scheme illustrating the SML of H3 depsipeptides to immobilized biotin-linker peptide. **C** Native PAGE of NCPs released from the beads by EcoRI digest after SML. NCPs were detected by Cy5 excitation. **D** Representative SDS-PAGE of eluted linker peptide after ligation. Peptides were detected via the Cy5 tag. **E** Representative western blot of SMLs with depsipeptides and immobilized ΔH3 NCPs. Full-length H3 and H3Δ(1-32) proteins were loaded as control. **F** SML conversion ratios of H3 depsipeptides and immobilized ΔH3 NCPs or linker peptide. Conversion ratios were determined by quantification of the western blot.

Two different SML protocols were used for peptide ligation to NCPs and the biotin-linker peptide, due to significant differences in ligation kinetics. Ratios for the conversion of ΔH3 to semi-synthetic full length H3 were determined by western blotting. Conversion ratios for the biotin-linker peptide were determined by quantification of Cy5 signal for non-ligated and ligated peptide after SDS-PAGE. Validation of NCP integrity after immobilization and SML was performed by releasing the NCPs off the resin by EcoRI restriction digest followed by native PAGE analysis. Reaction schemes and results of the SML reactions are shown in figure 38.
Sortase-mediated ligations of H3 depsipeptides to the immobilized linker peptide showed conversion ratios of >99 % for all peptide substrates. Conversion ratios for ligations on immobilized NCPs were slightly lower (>90 %). The released NCPs migrated as one distinct band on the native PAGE, indicating nucleosomal integrity after immobilization and SML.

2.11.4 Pull-down assays on AsuHd-containing H3 peptides and NCPs

Overall, four peptide-based and four NCP-based probes (H3 unmod., H3K27AsuHd, H3K9AsuHd and H3K4AsuHd) were synthesized and applied in the following pull-down assay. As further control, non-ligated ΔH3-NCPs (ligation-ready NCPs) were tested for unspecific binding of HDACs.
Equal amounts of magnetic beads for each probe were used for HDAC recruitment by incubating them with HEK293T whole cell extract at a concentration of 2 mg/mL (4.2.4.25). Input samples (30 µg whole cell lysate) and eluates were analyzed by immunoblotting (class I: HDAC1, HDAC2, HDAC3, HDAC8; class IIb: HDAC6). Representative immunoblots of pull-down experiments are shown in figure 39.
The pull-down experiments with H3 peptides showed efficient recruitment of all analyzed HDACs to H3K27AsuHd and H3K9AsuHd probes when compared to unmodified H3. HDAC binding was also detected with H3K4AsuHd, but showed overall weaker HDAC bands with exception of HDAC6. This indicated a preference of HDACs for the K27 and K9 sites. Pull-down experiments on the different NCPs showed enrichment of HDAC1, HDAC2 and HDAC6 at all three tested AsuHd incorporation sites. HDAC3 and HDAC8 were not recruited to the NCP probes at a detectable level. In general, pull-down experiments with nucleosomes as baits showed significantly lower unspecific binding to unmodified H3 when compared to the peptide probes. The most striking difference between peptide and NCP pull-downs was the exclusive recruitment of HDAC3 to AsuHd peptide probes. In addition, HDAC8 was enriched on the peptide probe H3K27AsuHd but not on the corresponding NCP probe.

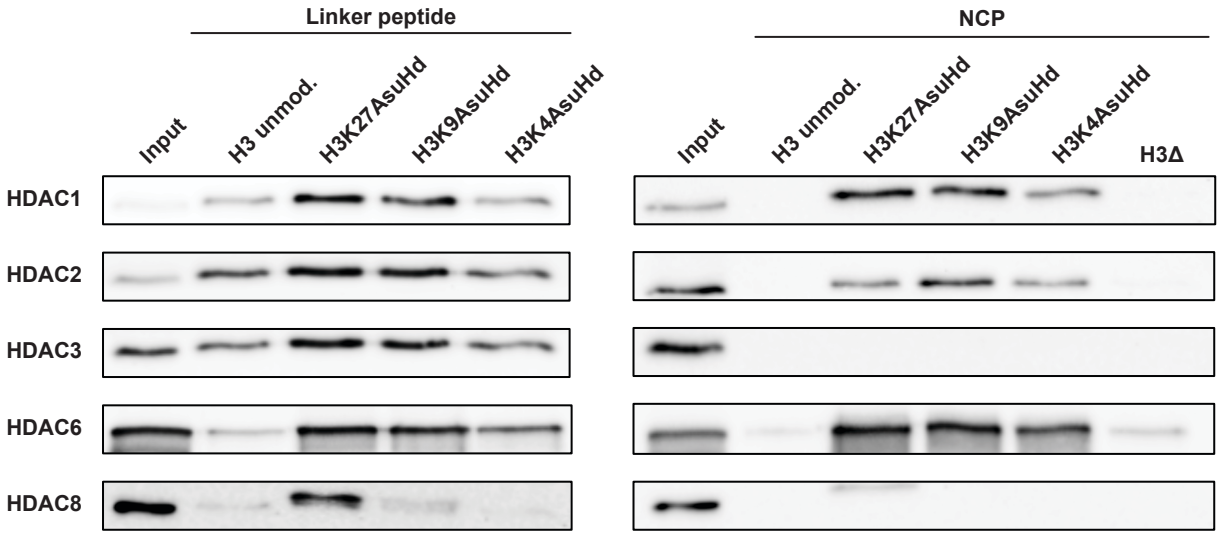


Figure 39: Pull-down results of H3-AsuHd peptide and NCP probes. Representative western blots are shown for pull-down experiments using ligated linker peptide probes (left panel) and ligated NCP probes (right panel) with HEK293T cell lysate (2 mg/mL). 30 µg lysate were used as input sample.

2.11.5 Proteome-wide pull-down analysis with H3 peptides and NCPs

After verification of successful HDAC enrichment on the AsuHd probes, additional HDAC interaction partners and potential unknown binders were investigated on a proteome-wide level. Label-free quantification (LFQ) of proteins bound to the probes by tandem mass spectrometry (LC-MS/MS) was used for this purpose. Pull-downs with immobilized and ligated peptide and NCP probes (H3 unmod., H3K27AsuHd, H3K9AsuHd and H3K4AsuHd) were performed with HEK293T whole cell extract at an input concentration of 2 mg/mL. Eluted proteins were alkylated, digested with LysC and trypsin, followed by desalting and concentrating via stop-and-go-extraction tips (StageTips) [266]. Nano-LC-MS/MS measurements, MaxQuant and Perseus data processing were carried out in cooperation with *Jürgen Eirich* (group of *Prof. Dr. Iris Finkemeier*, University of Münster, Germany).

LFQ values of identified proteins were $\log_2$ transformed and enrichment ratios were calculated between AsuHd-containing peptide or NCP probes and the unmodified H3 probes. Average $\log_2$ fold changes of identified proteins from three independent experiments were plotted against negative logarithmic scaled P-values determined by LIMMA analysis. Obtained volcano plots are shown in figure 40. Thresholds of $\log_2$ fold change ≥ 0.6 and $P \leq 0.05$ ($-\log_{10} \geq 1.3$) defined an area (highlighted in gray) of significantly AsuHd enriched proteins when compared to the respective unmodified probe. HDACs and known members of HDAC complexes are colored and labeled.

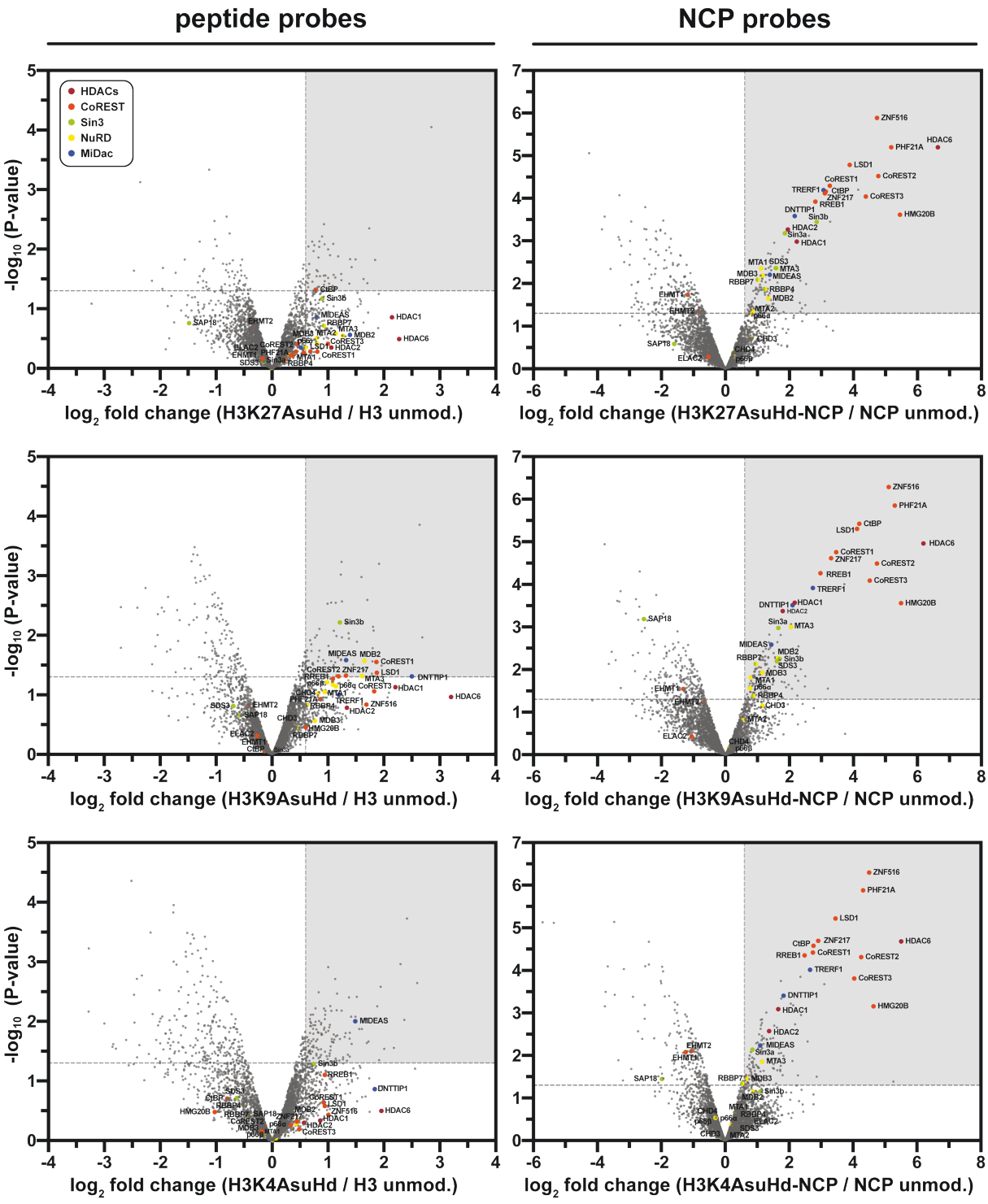


Figure 40: Volcano plots showing proteomics data of pull-down experiments with H3 peptide and NCP probes. The average $\log_2$ fold change of AsuHd (H3K27AsuHd, H3K9AsuHd and H3K4AsuHd) versus unmodified (H3 unmod.) peptide and NCP probes is plotted against negative logarithmic scaled P-values determined by LIMMA analysis. Dashed lines indicate significance thresholds of $P \leq 0.05$ ($-\log_{10} \geq 1.3$) and $\log_2$ fold change ≥ 0.6 and the area containing significantly enriched proteins is highlighted in gray. Pull-down experiments were performed as biological triplicates. HDACs are colored in red and HDAC complex components are colored by the respective complex: CoREST in orange, Sin3 in green, NuRD in yellow and MiDac in blue.

The three peptide probes H3K27AsuHd, H3K9AsuHd and H3K4AsuHd showed similar enrichment patterns. HDACs 1, 2 and 6 were identified but not significantly enriched. Interaction partners of the multiprotein HDAC complexes CoREST, Sin3, NuRD and MiDac (see 1.2.2) were also detected but most of them not at a significant level. Strongest enrichment of HDACs and complex partners was observed for the H3K9AsuHd peptide probe. Weakest enrichment was observed with H3K4AsuHd.
The volcano plots of proteomics-based pull-down experiments with NCPs showed much more robust enrichment patterns. Compared to the unmodified histone H3 NCP probe, all three AsuHd NCP probes showed strong enrichment of HDAC1, HDAC2 and HDAC6 and almost the entire CoREST, Sin3, NuRD and MiDac complex proteins. Components of the CoREST complex (CoREST1-3, PHF21A, LSD1, CtBP1, ZNF516, RREB1, HMG20B) were the strongest enriched proteins. Components of Sin3 (RBBP4, RBBP7, Sin3a, Sin3b, SDS3) and MiDac (TRERF1, DNTTIP1, MIDEAS) were slightly less abundant and lowest recruitment was detected for members of the NuRD complex (RBBP4, RBBP7, MTA1-3, MDB2 and 3, p66α). As already observed for peptide H3K4AsuHd, the weakest enrichment of HDAC complexes was found on the H3K4AsuHd NCPs, whereas H3K9AsuHd and H3K27AsuHd NCPs showed similar protein binding patterns. HDAC3 and the components of the N-CoR/SMRT complex were not identified or did not significantly enrich on NCP probes. The absence of HDAC3 on the NCP probes was already observed in the immunoblot analysis of the pull-down experiments.
The overall weak recruitment of HDAC complexes to the peptide-based probes led to the conclusion, that this experimental setup was less suitable for the identification of HDAC interaction partners. In comparison, the nucleosomal approach demonstrated solid and apparently more specific enrichment of HDAC1, HDAC2 and HDAC6 and their multiprotein complexes. The dataset of NCP enriched proteins was further processed by assigning the hits to protein groups of similar function. Due to the focus on nuclear proteins, non-nuclear proteins were excluded. Since the hydroxamate moiety is a general metal ion chelator, non-HDAC metal binding proteins were not further considered. The dataset further omitted known HDAC binding proteins. The final list of potential new HDAC interaction partners consists of 38 proteins, significantly enriched on at least one of the applied AsuHd-containing NCP probes. The newly identified potential binding partners were grouped into the following categories: DNA repair, histone modification, mRNA processing and transcriptional regulation (table 5).

Table 5: Selection of potential new interaction partners of HDACs 1, 2 and 6 identified by proteomics-based pull-down assays on AsuHd-NCP probes. Proteins with $\log_2$ fold change $\geq$ 0.6 and significant enrichment (LIMMA, $P \leq 0.05$) in pull-downs with AsuHd-NCP probes (H3K27AsuHd, H3K9AsuHd and H3K4AsuHd, abbreviated with Asu) versus H3 unmod. were filtered by function and nuclear localization. Known HDAC binders and non-HDAC metal binding proteins were excluded. Proteins were sorted in groups of similar function. Bold $\log_2$ fold change values indicate significant enrichment on the respective site.

UniProt ID	Protein	$\log_2$ fold change		
		K27Asu	K9Asu vs H3 unmod.	K4Asu
DNA repair				
Q96T60	Bifunctional polynucleotide phosphatase/kinase (PNKP)	**1.79**	**1.57**	**1.54**
P49916	DNA ligase 3 (LIG3)	**1.52**	**1.37**	**1.12**
P18887	DNA repair protein XRCC1 (XRCC1)	**1.43**	**1.25**	**1.16**
P09874	Poly [ADP-ribose] polymerase 1 (PARP1)	**1.03**	**0.94**	**0.75**
Q9UGN5	Poly [ADP-ribose] polymerase 2 (PARP2)	**0.96**	**0.85**	0.48
P13010	X-ray repair cross-complementing protein 5 (XRCC5)	**1.38**	**1.36**	**1.02**
P12956	X-ray repair cross-complementing protein 6 (XRCC6)	**1.27**	**1.26**	**0.85**
P46063	ATP-dependent DNA helicase Q1 (ReqQL)	**0.96**	**1.09**	**0.76**
P41208	Centrin-2 (CETN2)	**1.41**	**1.14**	**0.87**
Q01831	DNA repair protein complementing XP-C cells (XPC)	**0.80**	0.77	0.44
Q13426	DNA repair protein XRCC4 (XRCC4)	**1.17**	**1.10**	**0.66**
Q13889	General transcription factor IIH subunit 3 (GTF2H3)	**1.32**	**1.50**	**1.50**
Q92759	General transcription factor IIH subunit 4 (GTF2H4)	**1.28**	1.12	0.11
P19447	TFIIH basal transcription factor complex helicase XPB subunit (ERCC3)	0.52	**0.76**	0.29
P18074	TFIIH basal transcription factor complex helicase XPD subunit (ERCC2)	**1.08**	**0.76**	**1.22**
Q14191	Werner syndrome ATP-dependent helicase (WRN)	**1.23**	**1.35**	**1.14**
Histone modification				
Q9NSY1	BMP-2-inducible protein kinase (BMP2K)	**1.13**	**0.99**	**0.85**
Q13619	Cullin-4A (CUL-4A)	**2.21**	0.26	0.52
Q8TDB6	E3 ubiquitin-protein ligase DTX3L (DTX3L)	**1.27**	**1.05**	**0.88**

O60318	Germinal-center associated nuclear protein (GANP)	0.48	-0.40	**1.18**
Q9UI30	Multifunctional methyltransferase subunit TRM112-like protein (TRMT112)	-0.01	1.22	**1.80**
mRNA processing components				
Q15024	Exosome complex component RRP42 (EXOSC7)	**1.21**	-0.70	**0.98**
O60942	mRNA-capping enzyme (RNGTT)	**1.99**	**1.26**	**1.75**
Q8NAV1	Pre-mRNA-splicing factor 38A (PRPF38A)	0.97	**2.27**	1.65
Q9NXE8	Pre-mRNA-splicing factor CWC25 homolog (CWC25)	**0.84**	**1.02**	**1.03**
Transcriptional regulation				
Q9UHB7	AF4/FMR2 family member 1 (AFF4)	**1.56**	**1.12**	-0.25
Q12800	Alpha-globin transcription factor CP2 (TFCP2)	**2.39**	1.36	0.88
Q3L8U1	Chromodomain-helicase-DNA-binding protein 9 (CHD9)	**1.78**	1.42	0.75
Q14527	Helicase-like transcription factor (HLTF)	**1.15**	**0.93**	**0.65**
Q15475	Homeobox protein SIX1 (SIX1)	**3.10**	**3.20**	**2.86**
P42166	Lamina-associated polypeptide 2α (LAP2α)	**1.00**	**0.93**	**0.76**
P01106	Myc proto-oncogene protein (MYC)	0.85	**1.11**	**0.88**
P17480	Nucleolar transcription factor 1 (UBF-1)	**0.70**	**0.95**	**0.61**
Q06330	Recombining binding protein suppressor of hairless (RBPJ)	**1.35**	**1.24**	**0.82**
O00472	RNA polymerase II elongation factor ELL2 (ELL2)	0.89	**1.24**	**0.86**
Q15572	TATA box-binding protein-associated factor RNA polymerase I subunit C (TAF1C)	1.96	**2.10**	0.56
P05412	Transcription factor AP-1 (AP1)	0.30	**0.88**	**0.76**
Q969X6	U3 small nucleolar RNA-associated protein 4 homolog (UTP4)	-0.19	**1.56**	**2.19**

Most proteins of this list were associated with DNA repair processes and transcriptional regulation including transcription factors, transcriptional activators and repressors. The remaining proteins were grouped into histone modification enzymes and components, which are involved in mRNA processing. The majority of these proteins was significantly enriched at all three AsuHd sites with lowest enrichment on H3K4AsuHd. Several of these proteins showed distinct binding preferences towards specific sites, indicating potential site-specific protein recruitment by HDACs.

3 Discussion

The installation and removal of protein acetylation marks on lysine residues by HATs and HDACs is an essential regulation process of protein activity and function. In addition, lysine acetylation regulates gene expression by modulating chromatin states. Besides the direct effect of charge neutralization impacting protein function, acetyl-lysine also serves as binding motif for bromodomains. BRDs are usually found as integrated elements of multidomain proteins, containing a broad range of further reader domains and enzymatic subunits. Thereby, BRDs play a key role in regulation of chromatin modification levels by recruitment of additional chromatin remodeling factors. Alteration of protein acetylation levels correlates with the development of cancer and various other diseases, rendering HDACs and BRDs as promising drug targets. Understanding the role of HDACs and BRDs in regulation processes of the cell is fundamental for the development of HDAC and BRD inhibitors. The field of chemical biology provides a multitude of tools for HDAC and BRD research *in vitro* and *in vivo*, including strategies for protein semisynthesis, protein labelling, proteomics and non-canonical amino acids mimicking modified residues. Extending this toolbox offers new possibilities for future chromatin research potentially providing insights into essential cellular regulation mechanisms.

3.1 L-ApmTri as novel acetyl-lysine mimicking amino acid

Identifying the native binding sites of BRDs is an ongoing challenge of biochemistry. Early screening experiments with SPOT arrays demonstrated preferenced binding of BRDs to histone tails. In contrast, little is known about BRD binding to non-histone proteins [242]. Screening and validation of BRD substrates relied on acetylated peptide sequences and recombinantly expressed bromodomains. In cellular environments, acetyl-lysine is a limited BRD probe due to its instability towards deacetylases. This prompted the development of acetyl-lysine mimics exhibiting similar binding affinities as the native substrate and stability against deacetylases in this thesis. The discovery of JQ1 and I-BET762 as BET selective small-molecule inhibitors in 2010 opened new avenues in bromodomain research. The adaptation of small-molecule inhibitor moieties for the development of novel ncAAs was inspired by hydroxamate amino acids as HDAC traps [225] and recently, isoxazole-containing amino acids serving as BRD probes (see 3.1.5). Following the same design concept, the acetyl-lysine mimicking triazole moiety of JQ1 was transferred onto an amino acid scaffold and examined as stable acetyl-lysine mimic for BRDs.

3.1.1 Synthesis of triazole-containing amino acids

An efficient and high yielding synthetic access to ApmTri was the basis for its application as BRD probe. The developed synthesis route for Fmoc-L-ApmTri by enzymatic chiral resolution, followed by 9-BBN protection of the α-amino and carboxy group, triazole synthesis and Fmoc protection resulted in an overall yield of 26 %. The most challenging step was the synthesis of the 1,2,4-triazole or, more precisely, the coupling of acetamidine to the side chain of Apm. The coupling reagent and base were described as critical in the original protocol and various combinations were tested for this conversion. Most efficient coupling was reported with the combination HATU/DIPEA yielding >99 % conversion [231]. However, coupling of acetamidine to 9-BBN-Apm did result in much lower conversion rates. Testing of several coupling strategies revealed COMU/DIPEA as the most potent reagents. Acetamidine coupling with COMU/DIPEA resulted in conversion rates of >60 %, which was sufficient for delivering enough ApmTri for peptide probes and amber suppression experiments. However, additional activation strategies for the acetamidine coupling might increase synthesis yields even further. Strong activation of 9-BBN protected dicarboxy amino acids as acid chlorides has already been used to synthesize 9-BBN-L-Asu(NH-OTrt) (see scheme 4) and might facilitate this coupling step. The main advantage of the established synthesis route is the ease of modifying the triazole amino acid in terms of stereochemistry, side-chain length and substituents on the triazole moiety at the C-3 position (see scheme 1). The synthesis route further allows substituting the N-2 nitrogen by using monosubstituted hydrazines for triazole formation [231]. N-2 substituted ApmTri derivatives have not been synthesized and investigated yet, but this might allow the design of more specific BRD probes.

3.1.2 BRD binding to L-ApmTri-containing peptide probes

The synthesized triazole amino acids were tested for BRD binding after incorporation into peptide sequences derived from histone H3 and H4, which exhibit low micromolar binding affinities (table 1). Pull-down experiments with immobilized peptide probes and recombinant BRDs revealed a high selectivity for the recognition of triazole amino acids by BET bromodomains.

Efficient BRD3(2) binding is strongly dependent on the spacer length between triazole moiety and peptide backbone. Pull-down experiments with side-chain lengths of 3-5 methylene groups revealed a strong preference towards a spacer length of 4 methylene groups (L-ApmTri), whereas a 3 methyl group spacer (L-AadTri) showed no and a 5 methylene group spacer (L-AsuHd) weak enrichment (figure 19). The natural substrate acetyl-lysine also contains a 4 methylene group spacer between peptide backbone and Nε amide bond, indicating that this spacer length needs to be preserved in an acetyl-lysine mimic for BRDs. This could be explained by the anchoring hydrogen bond between the conserved asparagine residue and the oxygen of the acetyl carbonyl

group, which can only be formed if the hydrogen bond acceptor is at the correct position in the acetyl-lysine binding cleft of the bromodomain.
After optimizing spacer length, stereoselective BRD-ApmTri interaction was investigated. By using racemic DL-ApmTri and enantiopure D- and L-ApmTri in H4K20 probes for pull-down experiments with BRD3(2), the impact of the chiral center on BRD recognition was studied. As expected, the binding pattern revealed selective BRD3(2) binding to the L-enantiomer (figure 20). This finding suggested, that ApmTri is recognized as an acetyl-lysine mimicking amino acid by BRD3(2) and not as an inhibitory triazole moiety tethered to a peptide scaffold.
The impact of the C-3 methyl substituent of the triazole on BRD recognition was also investigated. A similar methyl group is also present in the small-molecule inhibitors JQ1 and I-BET762. Pull-down experiments with ApmTri derivatives without (L-ApmFTri), an ethyl (L-ApmPTri) and a phenyl (L-ApmBTri) C-3 triazole substituent in the H4K20 sequence context revealed the essential role of this methyl group for BRD3(2) binding (figure 21). Potent binding of BRDs to acyl-lysine substrates such as propionyl-, butyryl- and crotonyl-lysine was expected [267]. Propionyl-lysine was shown to enrich BRD3(2) in pull-downs. Surprisingly the corresponding ethyl-triazole moiety was not recognized by BRD3(2). However, the crystal structure of BRD3(2) with bound H4K20_L-ApmTri indicated that the ethyl group could be accommodated by the binding pocket of BRD3(2), if adopting the same conformation as the C-3 methyl group in ApmTri. No binding to L-ApmFTri and L-ApmBTri probes was detected, matching the inability of BRD3(2) to bind formyl- or benzoyl-lysine probes [267, 268].
In summary, investigating the relevance of different structural features of ApmTri for BRD3(2) binding revealed L-ApmTri as the best acetyl-lysine mimic for BRD3(2). Pull-down experiments with the non-BET bromodomains BAZ2B and CREBBP indicated a selectivity of L-ApmTri for BET protein recruitment. This finding correlates with the reported selectivity of JQ1 and I-BET762 for BRDs of the BET family [180, 186]. So far, only bromodomains of subfamily II, III and V were tested for L-ApmTri binding. The observed selectivity for subfamily II (BET proteins) could be further examined in future pull-down experiments with bromodomains of other subfamilies.
The interactions of L-ApmTri peptides (H4K20 and H4K5K8 probes) with BRD3(2) and BRD4(1) were quantified by MST (figure 27 and 28). The applicability of the MST setup for K_D determination was validated with the native acetylated substrates. The obtained K_D values of BRD3(2) and BRD4(1) binding to H4K20ac (3.2 $\pm$ 0.3 µM) and H4K5/K8ac (7.3 $\pm$ 0.5 µM) were in low micromolar range and agreed with reported K_D values for these interactions of 10.5 $\pm$ 0.2 µM and 6.8 $\pm$ 0.1 µM, respectively, which were determined by isothermal titration calorimetry (ITC) [242].
The binding affinity of BRD3(2) to H4K20_L-ApmTri (6.5 $\pm$ 0.5 µM) is approximately 2-fold lower as for H4K20ac. Slightly weaker affinity of BRD3(2) to ApmTri when compared to acetyl-lysine substrates was already indicated by the pull-down experiments (figure 19), but the MST results also confirmed these findings. Interestingly, the K_D value of 17.9 $\pm$ 1.7 µM recorded for H4K20_L-AsuTri binding to BRD3(2) indicated a stronger interaction

than expected from the pull-down data. This could be explained by differences in K_{on} and K_{off} rates. Due to extensive washing in pull-down experiments, the K_{off} is the primary determinante of the band intensity, whereas the K_D value is combining both rates. This suggests that the weaker binding of BRD3(2) to AsuTri substrates in pull-downs is primarily the result of a high K_{off}. In addition, it should be noted, that intensive staining of H4K20_L-ApmTri and H4K20_L-AsuTri beads by the fluorophores of the BRD constructs was detected directly after the pull-down. However, during the washing steps BRD3(2) was only weakly retained on H4K20_L-AsuTri when compared to H4K20_L-ApmTri.
Determined binding affinities of BRD4(1) to H4K5K8 substrates correlate with the binding preferences already observed in pull-down experiments (figure 25). The only exception is H4K5ApmTriK8ac, which exhibited a lower K_D value (62.9 µM) than expected. MST measurements of H4K5acK8ApmTri (6.7 $\pm$ 0.5 µM) and H4K5/K8ApmTri (26.9 µM) also confirmed the finding, that L-ApmTri seems to be an efficient mimic of acetyl-lysine at the K8 site, whereas substituting acetyl-lysine at the K5 site with L-ApmTri led to reduced BRD binding.
The titration experiments of JQ1 in BRD3(2) pull-downs on H4K20_L-ApmTri probes demonstrated successful blocking of BRD binding to the ncAA by competitive inhibition. The BAZ2B specific inhibitor GSK2801 was not affecting BRD3(2) recruitment to L-ApmTri, confirming the described substrate selectivity of the inhibitors towards members of the respective BRD subfamilies [180, 196]. Utilizing L-ApmTri for inhibitor screening assays on peptide scaffolds might provide a tool for development and characterization of novel small-molecule inhibitors for BET proteins. Due to the stability of L-ApmTri against deacetylases (figure 30), this type of displacement assay could also be carried out in cellular lysates.

3.1.3 Interaction of L-ApmTri and BRDs on molecular level

Crystal structures of BRD3(2) in complex with H4K20_L-ApmTri and BRD4(1) bound to H4K5acK8ApmTri or H4K5/K8ApmTri were determined in order to investigate ApmTri-BRD interactions on molecular level (figure 29).
The crystal structures revealed successful substitution of acetyl-lysine with L-ApmTri by preserving the essential anchoring hydrogen bond between the triazole and a conserved asparagine residue in the binding pockets of BRD3(2) and BRD4(1). With distances of 2.9 Å for N-1 and 3.5 Å for N-2 of the triazole to the conserved asparagine, both nitrogens could serve as hydrogen bond acceptor. The shorter distance of N-1 leads to the suggestion, that the anchoring hydrogen bond between N-1 and N391 is favoured. An alignment of BRD3(2) structures in complex with H4K20_L-ApmTri and H3K18ac (figure 41A) revealed, that N-2 is occupying the same position as the amide oxygen, serving as acceptor for the anchoring hydrogen bond in acetyl-lysine. So far, the formation of this bond between L-ApmTri and N391 could not be completely clarified but might be further investigated with an N-2 substituted triazole amino acids to eliminate N-2 as potential hydrogen bond acceptor.

The structure of BRD3(2) in complex with JQ1 was reported as well and was aligned to the structure of BRD3(2) with bound H4K20_L-ApmTri (figure 41B). This alignment revealed that the 1,2,4-triazole moieties of JQ1 and L-ApmTri adopt similar conformations in the acetyl-lysine binding pocket of BRD3(2).
The structure also illustrates the relevance of the side-chain length. AadTri most likely does not provide a long enough spacer for the triazole to approach N391 and adopt the conformation required for establishing the anchoring hydrogen bond. The electron density of the triazole and the C-3 methyl group is well defined in the BRD3(2)-H4K20_L-ApmTri structure in contrast to the 4 methylene groups of the side chain (see appendix figure A.5). The apparently high flexibility of the side chain might explain why the H4K20_-L-AsuTri probe binds BRD3(2) as well. However, the longer linker might not be optimal for simultaneous interactions of the triazole with N391 and the peptide part of the probe with the BRD3(2) surface. Alternatively, the limited space of the binding cleft provides less opportunities to accommodate a 5 methylene spacer. Both notions agree with the weaker binding of H4K20_L-AsuTri to BRD3(2) when compared to H4K20_L-ApmTri.

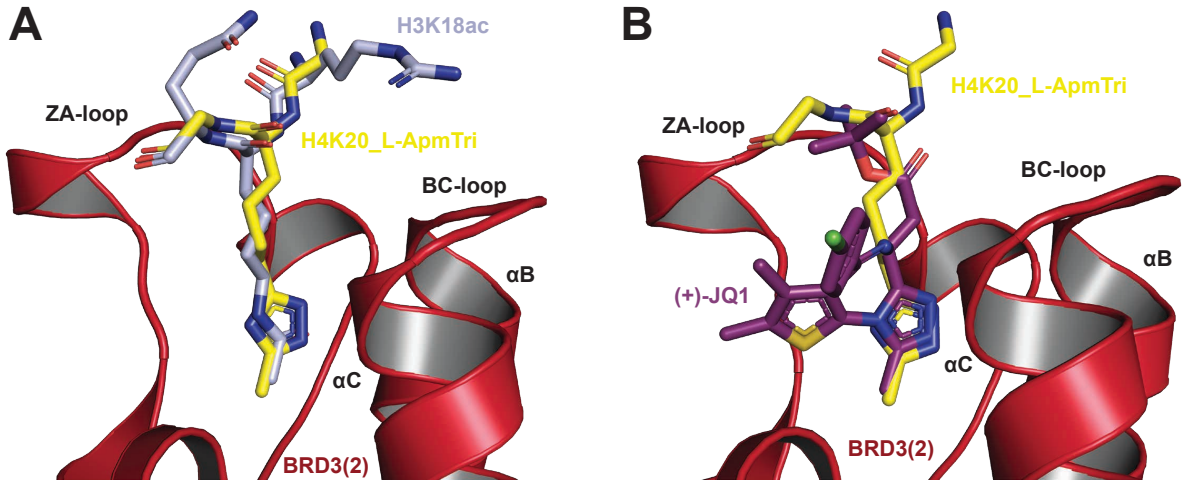


Figure 41: Structural overlay of H4K20_L-ApmTri with H3K18ac and JQ1 to the binding site of BRD3(2). A Structural overlay of H4K20_L-ApmTri (yellow) and H3K18ac (light blue, PDB code: 5HJC) in complex with BRD3(2) (red). **B** Structural overlay of H4K20_L-ApmTri and JQ1 (purple, PDB code: 3S92) in complex with BRD3(2). For reasons of clarity, the peptide backbones were shortened.

The crystal structures of BRD4(1) in complex with H4K5acK8ApmTri and H4K5/K8ApmTri provide insights into the L-ApmTri mediated interaction with this BRD (see figure 29). The N-terminal BRDs of BRD3 and BRD4 have evolved the unusual simultaneous binding mode of two acetyl-lysine residues. As reported for bromodomain 1 of BRD4, the N-terminal acetyl-lysine residue of di-acetylated substrates like H4K5/K8ac forms the anchoring hydrogen bond with the conserved asparagine N140. The C-terminal H4K8 acetyl-lysine is positioned at the protein surface and approaches H4K5ac in a diagonal orientation. Water-mediated hydrogen bonds stabilize the interaction between K5ac and K8ac [242, 269]. This binding mechanism could also be observed for K8ApmTri in combination with K5ac and K5ApmTri. An alignment

of BRD4(1) structures with bound H4K5acK8ApmTri and H4K5/K8ApmTri showed, that K5ac and K5ApmTri reach into the conserved acetyl-lysine binding cleft of BRD4(1) (figure 42C). The amide and triazole occupy the same positions, allowing K5ApmTri to establish the anchoring hydrogen bond with N140. ApmTri at the K8 site adopts a similar conformation as K8ac and approaches the K5 residue along the surface. The ligand electron density maps for both BRD4(1) bound peptides (H4K5acK8ApmTri and H4K5/K8ApmTri) define a distinct conformation of the acetyl group and the triazole in the BRD4(1) binding cleft for K5ac (see appendix, figure A.5). The N-2 of the triazole replaces the amide oxygen of acetyl-lysine in a similar fashion as observed for BRD3(2) with bound H4K20_L-ApmTri. The conformation of the K8 triazole appears more flexible between the substrates H4K5acK8ApmTri and H4K5/K8ApmTri. The positioning of the triazole ring of K8ApmTri is well-defined by the electron densities, but not the orientation of the C-3 methyl group. This allows two possible triazole conformations of K8ApmTri in H4K5acK8ApmTri and H4K5/K8ApmTri when bound to BRD4(1) as illustrated in figure 42C.

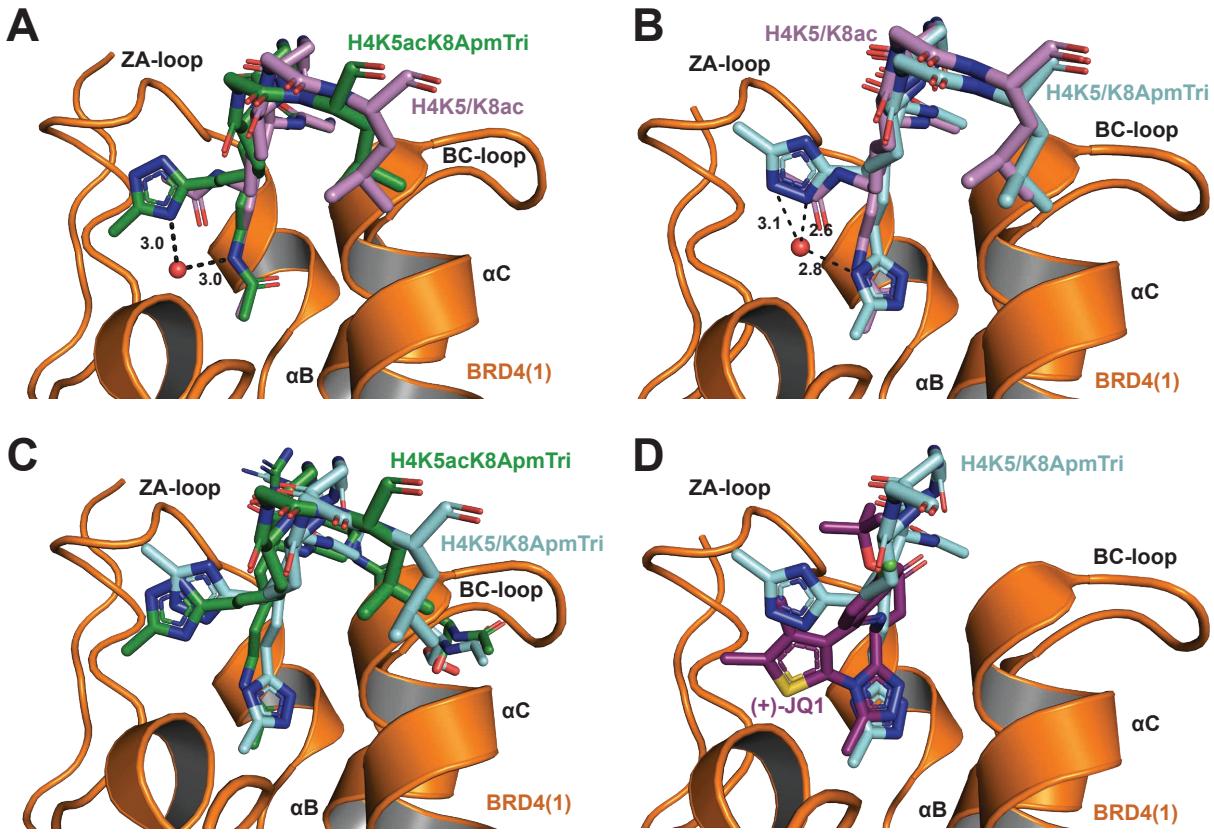


Figure 42: Structural overlay of L-ApmTri substrate peptides, H4K5/K8ac and JQ1 to the binding site of BRD4(1). **A** Structural overlay of H4K5acK8ApmTri (green) and H4K5/K8ac (purple, PDB code: 3UVW [242]) in complex with BRD4(1) (orange). **B** Structural overlay of H4K5/K8ac and H4K5/K8ApmTri (blue) in complex with BRD4(1). **C** Structural overlay of H4K5/K8ApmTri and H4K5acK8ApmTri in complex with BRD4(1). **B** Structural overlay of H4K5/K8ApmTri and JQ1 (PDB code: 3MXF [180]) in complex with BRD4(1). Water molecules are illustrated as red spheres and hydrogen bonds (distance cut-off: 3.6 Å) are shown as grey dashes and labeled in Å. For reasons of clarity, the peptide backbones were shortened.

Independently of its conformation, K8ApmTri is able to approach K5ac and K5ApmTri by the essential water-mediated hydrogen bonds (see figure 42A, B). This finding could explain the property of ApmTri to serve as mimic for K8ac in H4K5K8 substrates. Furthermore, it can be suspected, that the C-3 methyl group of K8ApmTri is not as important for BRD binding as the C-3 methyl group of K5ApmTri. The positioning of K8ApmTri also differs from the K8ac residue when aligned with a structure of BRD4(1) in complex with H4K5/K8ac. This supports the notion of a more flexible interaction in the K8 position (figure 42A, B) and could explain the finding that substitution of K8ac with L-ApmTri maintains high affinity to BRD4(1), whereas substitution of K5ac reduces BRD interaction.
A structural alignment of BRD4(1) in complex with H4K5/K8ApmTri and JQ1 revealed similar orientations of the triazole moieties in JQ1 and L-ApmTri (figure 42D) as already observed for BRD3(2) in complex with H4K20_L-ApmTri (figure 41B). The small-molecule inhibitor and the triazole amino acid are mimicking the native acetyl-lysine substrate by an analogous binding mechanism with N140. However, JQ1 is only targeting the binding cleft of BRD4(1), whereas the interaction of H4K5acK8ApmTri and H4K5/K8ApmTri with BRD4(1) is further stabilized by the water-mediated hydrogen bonding between the K5 and K8 residue.

3.1.4 Genetic encoding of L-ApmTri and AsuHd

Successful incorporation of L-ApmTri and AsuHd into the iRFP-GFP^{Y39TAG} reporter construct as well as into specific substrate proteins (Don1-4) was achieved by the wild type PylRS/tRNA^{Pyl} and the mutated PylRS^{AF}/tRNA^{Pyl} pair, respectively. Compared to the well-accepted substrate BocK, significantly higher concentrations of both ncAAs were required for efficient amber suppression (figure 31 and 34). Both amber suppression systems utilized for L-ApmTri and AsuHd incorporation were identified by screening of already established PylRS enzymes. The usage of optimized PylRS/tRNA^{Pyl} pairs for L-ApmTri and AsuHd might result in substantially increased protein expression efficiencies at lower ncAA concentrations.
Especially for AsuHd, low amino acid concentrations are desirable, due to reduced cell viability at ncAA concentrations >500 µM [235]. Besides HDAC binding, the ability of the hydroxamate moiety to chelate metal ions may also affect the function of non-HDAC metal-dependent proteins. The hydroxamate-based HDAC inhibitor SAHA induces cell growth arrest in a wide range of cells at concentrations of 2-5 µM [270]. Those effects could also explain the observed reduced expression of transfected constructs in AsuHd treated cells. Nevertheless, protein yields of Don2-4 were sufficient for co-immunoprecipitation of endogenous HDACs (figure 36). However, no differences in HDAC binding between the studied substrate sequences (minimal sequence context, α-tubulin, p53) were observed. With HDAC1, 2 and 6, the most abundant HDACs of HEK293T cell lysate were enriched at all three sites, indicating a competitive probe binding independent of the sequence context. As a result of low bait concentrations

the most abundant HDACs will likely outcompete low abundant AsuHd binders.
Compared to AsuHd, L-ApmTri did not show any toxic effects in HEK293T cells up to 20 mM, allowing cell treatment with higher concentrations of ncAA. This resulted in substantially increased expression levels of the amber construct. Furthermore, limited perturbation of ncAAs on cellular processes is highly desirable for studying BRDs in live cells. The potential of L-ApmTri for investigating BRD substrate binding in cells by using the PylRS/tRNA^{Pyl} system was demonstrated by co-precipitation of overexpressed BRD3(2) with the H4K20^{TAG}-mCerulean3 substrate construct Donor1 (figure 33), confirming functionality of genetically encoded ApmTri in this context. Furthermore, this protein-protein interaction was blocked by the small-molecule ligand JQ1, offering new prospects for characterizing BET inhibitors in cellular lysates or even living cells.

3.1.5 L-ApmTri in comparison to established acetyl-lysine mimics

So far, most of the developed non-natural acetyl-lysine mimics were designed for studying histone deacetylases [224]. Despite their important role in recruitment of PTM readers, only few acetyl-lysine mimics for bromodomains were reported. These include Nε-methanesulfonyl-lysine, methylthiocarbonyl-thialysine (MTCTK) and a set of isoxazole-containing amino acids (figure 43).

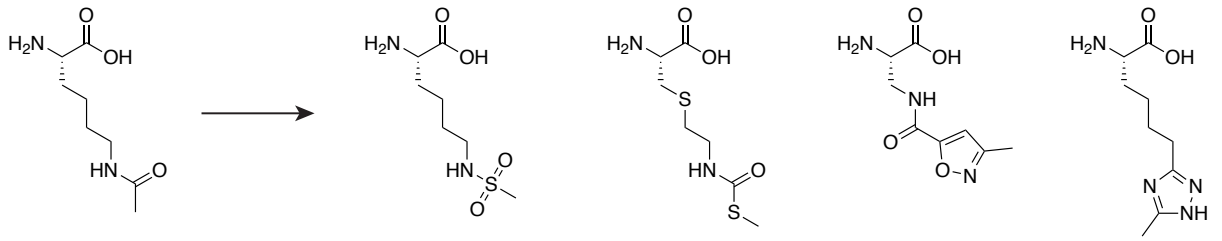


Figure 43: Stable acetyl-lysine mimics for investigating bromodomains. BRD substrate recognition of methanesulfonyl-lysine was shown for CREBBP, of MTCTK for BRDT and isoxazole-containing amino acids for several BRD subfamilies [271–273]. The new L-ApmTri acetyl-lysine mimic shows potential BET-specific recognition properties.

Nε-methanesulfonyl-lysine was reported by Jamonnak et al. in 2007 as non-hydrolyzable functional surrogate for acetyl-lysine [271]. The acetyl-lysine mimic was incorporated into a p53-derived peptide scaffold as SPPS building block, replacing K382, a known binding site of CREBBP. Interaction of the CREBBP bromodomain with peptide probes containing lysine, acetyl-lysine or methanesulfonyl-lysine were determined in an *in vitro* pull-down assay, showing similar BRD recruitment to mimic and native substrate. The resistance of methanesulfonyl-lysine to deacetylases was evaluated in deacetylase assays with human HDAC8 and SIRT1. Weak inhibitory effects on the deacetylases were observed but the methanesulfonyl moiety was not cleaved by the enzymes.
In 2010, Philip A. Cole and coworkers reported the thiocarbamate-containing amino

acid methylthiocarbonyl-thialysine (MTCTK) as stable acetyl-lysine analog for BRDs [272]. It can be incorporated into peptides or proteins by using methylthiocarbonyl-aziridine (MTCA) for the alkylation of cysteine residues. Substitutions of acetyl-lysine at the K5 or K8 position of H4 derived peptide probes with MTCTK were able to compete against the native di-acetylated substrate in BRDT(1) pull-down assays, although with approximately 2 to 4-fold lower efficiency than the native acetyl-lysine substrate. Notable is also the recognition of MTCTK by site-specific anti-acetyl-lysine antibodies on histone H3 proteins as well as the demonstrated resistance to the deacetylases HDAC8 and SIRT2 when incorporated into a p53 peptide context.
More recently, Conway and coworkers reported a set of 10 isoxazole-derived amino acids as acetyl-lysine mimics for bromodomains [273]. Following a similar approach as for the L-ApmTri development, a 3,5-dimethylisoxazole moiety was derived from a small-molecule BRD ligand and grafted onto an amino acid scaffold. The amino acids were tested for binding of BRD4(1), BAZ2A and BRD9 and each of the amino acids showed some concentration-dependent binding in an AlphaScreen assay. Furthermore, isoxazole amino acids were incorporated as substitutes of selected acetylation sites in poly-acetylated histone H4 peptides (K5, K8, K12 and K16). Binding of BRD4(1) to those peptides was tested, revealing highest BRD binding when isoxazole replaced K12. The usability of those novel mimics for semisynthetic approaches was shown by site-directed incorporation of isoxazole moieties at cysteine residues into recombinant histone H3 by alkylation [274].
When comparing the structures of the discussed acetyl-lysine substitutes (figure 43), it appears that all amino acids preserve the methyl group of acetyl-lysine. The variation of the C-3 substituent on the 1,2,4-triazole of L-ApmTri also pointed out the importance of this methyl group for BRD binding.
Compared to the reported acetyl-lysine mimics, L-ApmTri offers several advantages. The accessibility of an Fmoc building block allows comfortable incorporation of L-ApmTri into peptides by SPPS, whereas MTCTK requires an additional chemical modification step. Furthermore, the genetic encoding of L-ApmTri and the resistance towards enzymatic cleavage provides a powerful instrument to generate specifically modified proteins in bacteria as well as in mammalian cells. In addition, L-ApmTri appears by now as the best characterized acetyl-lysine mimic for BRDs. The characterization of L-ApmTri properties includes BRD binding selectivity, binding affinities for several substrate peptides, crystal structures in complex with BRDs and stability studies in cellular environment.

3.2 AsuHd-containing NCPs as probe for HDAC recruitment on a proteome-wide level

3.2.1 Generation of AsuHd-containing designer nucleosomes

The generation of semisynthetic designer nucleosomes by ligation of synthetic H3 depsipeptides to truncated nucleosomal H3 tails via sortase-mediated ligation was already established by *Diego Aparicio Pelaz* and successfully applied for the investigation of HP1 recruitment to modified H3 tails [207]. The efficient trapping of HDACs by the hydroxamate-based amino acid AsuHd was demonstrated in a number of peptide-based pull-down approaches including agarose-linked peptide probes as well as peptide arrays for screening of potential HDAC substrates [225, 275]. The combination of designer NCPs with AsuHd for establishing *in vitro* pull-down assays in a more biological context is a consequent extension of the toolbox for HDAC research in chemical biology.
Compared to previous preparations of designer NCP libraries in microtiter plates, the conditions for SML on streptavidin-coated beads were adjusted due to altered sortase kinetics in this setting. Especially SML reactions with biotin-linker peptides differed from the ligation to NCPs and exhibited much faster kinetics. This is probably caused by increased accessibility of the immobilized glycine donor peptide when compared to NCPs. After SML optimization, final ligation conversions of >90 % for peptide and NCP probes were achieved (figure 38), similar to previous SML reactions on NCPs [207]. Consistent SML and immobilization levels of peptides and NCPs allowed direct comparison of pull-down results on both types of probes.

3.2.2 HDAC recruitment on AsuHd-containing peptides and nucleosomes

HDAC recruitment by the novel AsuHd-based probes was tested in pull-down assays from HEK293T cell lysate, followed by western-blot analysis (figure 39). HDACs 1, 2, 3 and 6 were efficiently recruited to all three AsuHd sites of peptide probes, whereas NCP-probes did not recruit HDAC3. HDACs 1-3 and the respective multiprotein complexes are primarily localized in the nucleus and most of the reported functions are associated with chromatin remodeling and activity. HDAC6 is primarily localized to the cytoplasm and deacetylates non-histone proteins. All peptide probes showed similar HDAC recruitment patterns, which reflected HDAC binding patterns of peptide probes with minimal sequence context (G-AsuHd-G), indicating a low impact of peptide sequence context on HDAC binding [225]. Whole cell lysate contains nuclear and cytoplasmic proteins and HDAC6 as one of the most abundant HDACs in HEK293T cells could simply outcompete other HDACs from AsuHd probes. Interestingly, HDAC3 recruitment to peptide

probes was detected, but not to NCP probes. This indicates a more selective binding of HDACs to AsuHd in a nucleosomal than in a peptide context. This presumption is also confirmed by significantly reduced HDAC binding to the unmodified H3 tail when compared to peptide-based probes, where strong unspecific binding to probes without AsuHd was observed. A consistent interaction pattern of peptide and NCP probes was the weak recruitment of HDACs to H3K4AsuHd probes, probably caused by the close proximity of the AsuHd moiety to the histone N-terminus, reducing potential interactions with the enclosing tail region. This binding behaviour was observed in the proteomics data of the pull-down experiments as well.

After verification of HDAC recruitment to the novel AsuHd-based probes, the pull-down approach was extended to a proteome-wide level. Proteomics experiments on AsuHd peptides already demonstrated enrichment of HDACs and their corresponding complexes [227, 276]. However, the volcano blots of proteomic pull-down experiments on H3 tail peptides indicated weak enrichment of HDACs and complex components (figure 40). HDAC complexes were mostly not significantly enriched when compared to the unmodified H3 peptide. Several explanations are conceivable for this findings: Compared to previous proteomics experiments on peptide probes (immobilized on SulfoLink™ agarose beads), substantially reduced amounts of peptide probes were applied to the pull-down assay, in order to reproduce the amount of immobilized NCP material. This probably results in less interaction between HDACs and proteins that support HDAC recruitment to AsuHd probes. Along these lines, lower functionalization of the resin allows for more unspecific interaction, causing stronger binding of HDACs to the unmodified peptide probe. Increased binding to the negative probe can also be caused by charge-related unspecific interactions between proteins and the highly charged H3 peptide probes. As a result, enrichment ratios of HDACs and HDAC complexes were closer to the background level. Such charge-related effects of H3 derived peptides could be partly neutralized by addition of nucleosomal context, where histone tails are embedded into a structural scaffold. Within the NCP context, charge neutralization by interaction of histone tails and DNA appears likely, thereby reducing unspecific interactions with the highly charged H3 tail [277]. This could explain reduced unspecific binding to unmodified H3 peptide when ligated to NCPs (figure 39). Furthermore, this results in an increased relative enrichment of HDAC complexes on AsuHd-containing NCP probes when compared to the non-AsuHd control (figure 40).

Proteomic pull-downs on AsuHd probes without nucleosomal context have been reported before [227, 276]. While these studies also showed significant enrichment of HDACs and HDAC complexes, the observed $\log_2$ fold changes of HDAC enrichment were generally higher in the NCP context. This observation is in agreement with the notion, that the NCP probes reflect the physiological HDAC substrate more closely than peptide probes. In addition, HDAC complexes were almost entirely enriched. This finding further underlines the crucial role of the nucleosomal context for investigating HDAC multiprotein complexes.

Besides HDAC1 and 2 (and also HDAC6), it would be expected that HDAC3 is recruited

on NCP probes as well, due to its reported role in chromatin remodeling and deacetylation of the histone H3 tail [278]. In the proteomic analysis of pull-down experiments, HDAC3 and N-Cor/SMRT complex components were not significantly enriched on any of the three investigated H3 sites. The only exception was the lysine-specific demethylase 3A (KDM3A, also known as JHDM2A), which catalyzes demethylation of H3K9me2. However, KDM3A also exhibits a JmjC domain with a bound Fe(II) ion and a zinc finger domain, which are probably enriched directly on the hydroxamate moiety [279]. The absence of HDAC3 on NCP-based probes could be caused by several effects: One possibility is a competitive binding of HDACs and their complexes to AsuHd, where HDAC3 and N-Cor/SMRT are outcompeted by more abundant HDACs. Another explanation is that recruitment of HDAC3 and N-Cor/SMRT is assisted by further essential histone modifications absent in the NCP probes. However, such requirements are not reported for HDAC3 so far.

3.2.3 Identification of potential HDAC interaction partners

A primary objective of the established proteome-wide pull-down assay on AsuHd-containing NCPs was the identification of unknown potential HDAC interaction partners. The successful recruitment of known HDAC1 and HDAC2 complex components demonstrated the potential of NCP-based probes for this purpose. Besides well-known HDAC complexes, numerous additional proteins were found significantly enriched on the NCP probes, providing a set of potential interaction partners of HDAC1, HDAC2 and HDAC6. After filtering the dataset by function, exclusion of known HDAC binders, potential false-positive (proteins exhibiting metal-binding domains) and non-nuclear proteins, a selection of 38 new potential HDAC interaction partners remained (table 5). The primary biological functions of the protein hits and their potential link to HDAC activities is discussed in the following:

DNA repair proteins

Most of the identified protein hits are involved in DNA repair and transcriptional regulation and known to interact with one another in DNA repair complexes. The proteins LIG3, PNKP, XRCC1, PARP1 and PARP2 are components of the base excision repair (BER) complex, involved in DNA single-strand break repair [280]. The single-stranded DNA-dependent ATP-dependent helicases XRCC5 and XRCC6 (also known as Ku80 and Ku70) are forming the heterodimer protein complex Ku, involved in DNA non-homologous end joining (NHEJ) during double-strand break repair [281]. Ku is also interacting with XRCC4, which is also enriched at all investigated H3 sites. An interaction between Ku and the multifunctional enzyme WRN (helicase and exonuclease activity) and its involvement in the repair of double strand DNA breaks is also reported [282]. Another enriched DNA repair complex is the general transcription and DNA repair factor IIH (TFIIH) core complex. The components ERCC2 and ERCC3 (also known as XPB and

XPD), GTF2H3 and GTF2H4 were enriched on AsuHd NCP probes. The TFIIH complex is involved in transcription initiation by RNA polymerase II and nucleotide excision DNA repair [283, 284]. XPC is the main component of the XPC complex, involved in global genome nucleotide excision repair (GG-NER). Another significantly enriched component of the complex is CENT2 [285].
The installation of post-translational modifications on histones and DNA repair proteins as DNA-damage-induced chromatin response has been reported as key event for several DNA repair processes. The observed PTMs include poly(ADP-ribosyl)ation by PARP proteins, phosphorylation of H2AX by the serine/threonine kinase ATM (ataxia telangiectasia mutated) and acetylation of histones, responsible for chromatin restructuring and increasing accessibility for DNA repair complexes [286]. Deacetylation of histones during chromatin reconstitution after DNA repair is primarily carried out by HDACs [287]. A direct link of HDACs to DNA-damage response has been shown for double-strand break repair via NHEJ, where acetylation of lysine H3K56 and H4K16 is regulated by HDAC1 and HDAC2. However, the mechanism of NHEJ stimulation by the deacetylation events is not completely understood. Besides an indirect mechanism by local inhibition of transcription at double-strand breaks, HDAC1 and 2 may also prevent dissociation of the protein complex Ku from double-strand break termini to further mediate DNA repair [288]. The enrichment of multiple DNA repair complex components in the NCP-based pull-down approach now indicates additional direct interactions between HDACs and DNA repair proteins. The access of 5' and 3' of the NCP might be recognized as DNA damage sites and induces the association of HDACs with DNA repair complexes on the respective NCPs. However, it should be noted, that the accessibility of the 5' and 3' end of the DNA is limited by the Cy5 fluorophore and the biotin handle.

Histone modifying enzymes

Few of the identified potential HDAC interaction partners are associated with the installation or removal of histone modifications. The identified proteins comprises methyltransferases (TRMT112), HATs (GANP), kinases (BMP2K) and ubiquitinases (Cullin-4A, DTX3L). GANP stands out because this protein is only significantly enriched on the H3K4AsuHd-NCPs. In addition to its catalytic function as histone acetyl-transferase, GANP is also a component of the TREX-2 complex, involved in nuclear export of mRNAs to the cytoplasm through the nuclear pores. Another member of TREX-2 is CENT2, already found and described as part of the XPC complex [289]. TREX-2 is localized to the nuclear pore complexes (NPCs) and their constituent nucleoporins (Nups), which are also associated with regulation of gene expression besides their main function in nuclear transport. However, the regulation of the transcription machinery by TREX-2 is still poorly understood. Interactions of NPC proteins and HDACs were already described: HDAC4 interacts with Nup155 in mammalian cardiomyocytes [290] and HDAC1 binds Nup98 [291], indicating a relation between regulation of chromatin states and nuclear export mechanisms.

mRNA processing components and transcriptional regulators

Several AsuHd-NCP enriched proteins are mRNA processing factors and transcriptional regulators, including transcriptional activators, repressors and transcription factors. Maturation of mRNA occurs co-transcriptionally, thus linking transcription and RNA processing. Indirect enrichment of mRNA processing enzymes by interaction with recruited transcriptional regulators appears plausible. In contrast to the recruitment of DNA repair complexes on AsuHd-NCP probes, most of the identified mRNA processing components and transcriptional regulators were enriched independently of other known binding partners.

The mRNA-capping enzyme RNGTT is involved in the mRNA cap formation [292], whereas the pre-mRNA-splicing factors PRPF38A and CWC25 are components of the spliceosome B and C complex, respectively [293, 294]. EXOSC7 is a component of the RNA exosome core (Exo-9) complex and is involved in mRNA processing and degradation of various substrates [295]. Although the molecular mechanism remains unclear, recent studies revealed a role of HDACs in regulation of splicing-associated histone marks as well as direct interaction of HDAC1 and 2 with the serine/arginine-rich splicing factor 1 (SRSF1), which directly regulates protein splicing [296].

The interaction of HDACs with transcriptional regulators is essential for regulation of gene expression, thus enrichment of such proteins was expected. Besides known components of the HDAC1 and 2 multiprotein complexes, several additional transcription regulators were recruited to AsuHd-containing NCP probes. The homeobox protein SIX1 was strongly enriched at all three AsuHd histone sites with $\log_2$ fold changes comparable to components of the Sin3 complex, indicating a specific interaction with HDAC1 or 2 and their complexes. SIX1 is a member of the sine oculis homeobox (SIX) family, consisting of evolutionarily conserved transcription factors. SIX1 plays an important role in cell cycle regulation and overexpression in several tumor types has been reported [297]. SIX1 interacts with the dachshund homolog 1 (DACH1), a known interaction partner of HDAC1, 2 and 3 and the Sin3 complex, which is also significantly enriched on all three AsuHd sites [298, 299].

With the transcription factors MYC and AP-1 (also known as c-Jun) two prominent proteins classified as proto-oncogenic were significantly enriched on H3K9AsuHd and H3K4AsuHd. In addition to their widespreaded interactions with other transcription factors, especially Myc is associated with the regulation of euchromatin by recruitment of HATs such as GCN5 and TIP60 [300].

The proteins nucleolar transcription factor 1 (UBF-1) and the TATA box-binding protein-associated factor RNA polymerase I subunit C (TAF1C) showed both strongest enrichment on the H3K9AsuHd-containing NCP probe. UBF-1 interacts with SL1/TIF-IB, a transcription factor complex, also consisting of TAF1C [301, 302]. Direct interactions of the SL1/TIF-IB complex with HDACs are not reported so far.

Several of the enriched proteins are associated with ATP-dependent chromatin remodeling processes, such as CHD-9, HLTF and RBPJ. For RBPJ, a key effector of Notch signaling pathways, the recruitment of large co-repressor complexes containing HDAC1

and HDAC3 on the (HIV)-1 LTR promoter was reported [303]. CHD-9 is a member of the CHD chromatin remodeling proteins, which also comprises the NuRD components CHD-3 and 4. To date, the biological functions of CHD-9 is only poorly investigated and interactions with HDAC complexes have not been reported so far [304, 305].

3.3 Outlook

The crucial roles of bromodomains in regulatory processes and their involvement in the development of a multitude of diseases resulted in great aspirations in bromodomain research in the last decade which will likely continue in the future. The novel acetyl-lysine mimicking amino acid L-ApmTri provides an additional tool to study bromodomains and their interactions.

BRD binding properties to L-ApmTri were widely characterized in this work, but could be further extended in future studies. An interesting question which remains to be answered is, why triazole-based amino acids and small-molecule ligands bind only BET bromodomains. L-ApmTri could be further modified on the C-3 or N-2 position of the 1,2,4-triazole moiety and the amino acid side-chain by integration of amide bonds or aromatic systems, which was already shown to improve recognition of isoxazole-derived amino acids by distinct BRDs [273]. The determined BRD protein structures with bound L-ApmTri-containing peptides provide insight to the recognition and binding mechanism and could be utilized for direct design of novel triazole-based BRD probes.

This work provides ApmTri as building block for SPPS and genetically encoded ncAA. The latter could be used for investigating BRD-substrate interactions in living cells by FRET as well as *in vivo* activity of BET inhibitors. Nowadays, the development of BRD inhibitors, particularly of iBETs, is focussing on small-molecule ligands, which reach into the binding pocket and are able to form the anchoring hydrogen bond with the conserved asparagine residue and further interactions with the hydrophobic cavity. L-ApmTri as stable and non-hydrolyzable acetyl-lysine mimic now allows the development of a new, peptide-based class of BET inhibitors. Such inhibitors could address not only the acetyl-lysine binding pocket, but also the surface of bromodomains. However, the binding affinity of L-ApmTri is substantially lower when compared to small-molecule ligands. The development of optimized recognition sequences in combination with simultaneous targeting of bromodomain 1 and 2 of BET proteins may overcome this limitation.

The demonstrated potential of NCP-based pull-down assays in the recruitment of HDACs and HDAC complexes by AsuHd-containing NCPs provides a new *in vitro* approach for the discovery of HDAC protein-protein interactions. Using the concept of semisynthetic NCP probes with peptides derived from SPPS, various additional histone tail modifications such as methylation and phosphorylation could be incorporated in order to investigate their effects on recruitment of HDAC complexes.

An interesting observation for further investigations is the absence of HDAC3 and the

N-Cor/SMRT complex at the three investigated acetylation sites on histone H3 in NCP context. The introduction of additional PTMs on the applied histone H3 tails may affect the enrichment of HDAC3, revealing novel crosstalks of multiple histone modifications in N-Cor/SMRT recruitment. Another interesting approach would be the incorporation of the recently established amino acid 2-amino-8-[(2-aminophenyl)amino]-8-oxooctanoic acid (AsuApa), because it is specific for HDACs 1-3 with a preference for HDAC3 [276]. The application of AsuApa may reduce competitive binding effects between HDAC1, 2, 3 and 6 on AsuHd probes and probably provides an additional application for the investigation of HDAC3 recruitment to histone H3.

The selection of potential HDAC interaction partners, which were identified by enrichment in the proteome-wide pull-down assays on AsuHd-containing NCP probes was already discussed in terms of biological functions and reasonable HDAC interactions. To further exclude false positive hits and to validate interactions between HDACs and the identified protein hits, further experiments are required. For example, co-immunoprecipitations of HDACs/HDAC complexes with the respective proteins or HDAC knockdown experiments could be used to validate the interaction of the newly identified potential HDAC binders.

4 Materials and Methods

4.1 Materials

4.1.1 Suppliers

Suppliers, their headquarter locations and the abbreviations used in the following are listed in table A.4 of the appendix.

4.1.2 Chemicals

Standard Fmoc-amino acid building blocks and HBTU (2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluroniumhexafluorophosphate) were purchased from GL Biochem. Amino dicarboxylic acid derivatives and Fmoc-L-Lys(Ac)-OH were purchased from Bachem. Fmoc-Ahx-OH was purchased from Iris Biotech and TentaGel® HL RAM resins from Rapp Polymere. Unless otherwise stated, chemicals and organic solvents were purchased from Sigma-Aldrich, Merck KGaA, Carl Roth, Iris Biotech, Carbolution, J. T. Baker, VWR, Thermo Fisher Scientific, Biosolve and Th. Geyer. Plastic consumables were purchased from Sarstedt and Greiner.

4.1.3 Biological materials

4.1.3.1 Buffers and media

Unless otherwise stated, all components for buffers were purchased from Sigma-Aldrich or Carl Roth. LB and 2xYT medium for *E.coli* cultivation were purchased from Carl Roth. Dulbecco's Modified Eagle's Medium (DMEM) (high glucose), fetal calf serum (FCS) and PBS for cell culture were purchased from Sigma-Aldrich.

4.1.3.2 Kits and enzymes for molecular cloning

Table 6: Kits for molecular cloning.

Kit	Supplier
peqGOLD Plasmid Miniprep Kit I	Peqlab
GeneJET Plasmid Maxiprep Kit	Thermo Fisher Scientific
GeneJET Gel Extraction Kit	Thermo Fisher Scientific
GeneJET PCR Purifcation Kit	Thermo Fisher Scientific

Table 7: Enzymes and buffers for molecular cloning.

Equipment	Supplier
BamHI	Thermo Fisher Scientific
FastDigest BamHI	Thermo Fisher Scientific
FastDigest Bsu15I (isoschizomer of ClaI)	Thermo Fisher Scientific
FastDigest EcoRI	Thermo Fisher Scientific
FastDigest Eco32I (isoschizomer of EcoRV)	Thermo Fisher Scientific
FastDigest PstI	Thermo Fisher Scientific
FastDigest XbaI	Thermo Fisher Scientific
HindIII	Thermo Fisher Scientific
NdeI	Thermo Fisher Scientific
Pfu DNA Polymerase	Thermo Fisher Scientific
Q5® High-Fidelity DNA Polymerase	New England Biolabs
T4 DNA Ligase	Thermo Fisher Scientific
Buffer	**Supplier**
Buffer R (10x)	Thermo Fisher Scientific
dNTP Mix, 2 mM each	Thermo Fisher Scientific
FastDigest Buffer (10x)	Thermo Fisher Scientific
FastDigest Green Buffer (10x)	Thermo Fisher Scientific
Pfu Buffer with $MgSO_4$ (10x)	Thermo Fisher Scientific
Q5® Reaction Buffer (5x)	New England Biolabs
Tango Buffer (10x)	Thermo Fisher Scientific
T4 DNA Ligase Buffer (10x)	Thermo Fisher Scientific

4.1.3.3 Plasmids and Primers

Table 8: Empty expression vectors used for cloning.

Vector	Promotor	Tags	Supplier
pBI-CMV1	$P_{minCMV1}$, $P_{minCMV2}$	-	Clontech
pET28a(+)	T7	$^{N}His_6$, $^{C}His_6$	Merck

Table 9: Plasmids containing synthetic genes (purchased from GeneArt Gene Synthesis, Thermo Fisher Scientific).

Backbone	Synthetic gene	Cloning sites	Codon optimization
pMK-RQ	BAZ2B-mTagBFP-*Strep*	5'-NdeI, 3'-HindIII	*E. coli*
pMA-T	BRD3(2)-TagRFP-*Strep*	5'-NdeI, 3'-HindIII	*E. coli*
pMA-RQ	BRD4(1)-TagGFP2-*Strep*	5'-NdeI, 3'-HindIII	*E. coli*
pMA-RQ	BRD4(2)	5'-NdeI, 3'-BamHI	*E. coli*
pMA-T	CREBBP-TurboYFP-*Strep*	5'-NdeI, 3'-HindIII	*E. coli*
pMA-RQ	*Strep*-H4K20^{TAG}-mCerulean3-*His*$_6$	5'-BamHI, 3'-EcoRV	*H. sapiens*
pMA-RQ	BRD3(2)-EYFP-*3xFLAG*	5'-EcoRI, 3'-XbaI	*H. sapiens*
pMA-RQ	NES-*Spot*-mini^{TAG}	5'-BamHI, 3'-ClaI	*H. sapiens*
pMA-RQ	NES-*Spot*-p53^{TAG}-mClover3-*Myc*	5'-BamHI, 3'-EcoRV	*H. sapiens*
pMA-RQ	NES-*Spot*-αTubulin^{TAG}	5'-BamHI, 3'-ClaI	*H. sapiens*
pMA-RQ	NES-HDAC1-mRuby3-*3xFLAG*	5'-EcoRI, 3'-XbaI	*H. sapiens*

Table 10: Provided plasmids containing gene constructs. Plasmids used for amber suppression experiments were provided by *Prof. Dr. Edward A. Lemke* (Johannes Gutenberg University Mainz). Sortase A expression vector was provided by *Lena Schmohl*.

Backbone	Gene	Provider
pCMV	NES-PylRS^{WT}-hUtRNA^{Pyl}	Lemke group
pCMV	NES-PylRS^{AF}-hUtRNA^{Pyl}	Lemke group
pCI	NES-iRFP-25Helix-*FLAG*-GFP(Y39TAG)-*His*$_6$	Lemke Group
pET23b	Sortase A-*His*$_6$	Schwarzer group

Table 11: PCR and sequencing primers. Restriction sites are underlined.

Primer for PCR reactions	**Sequence (5' to 3')**
NdeI-BRD3(2)-fwd	GGCAGCCATATGGGTAAACTGAGC
BRD3(2)-BamHI-rev	GTGTGGATCCACCGCTCTGAAAAT
BamHI-TurboYFP-fwd	ATATGGATCCAGCGGTGCACTGC
TurboYFP-HindIII-rev	TGCGCAAGCTTATTTTTCAAACTGCGG
BRD3(2)-HindIII-rev	TGTGAAGCTTAAGGCATTTTGGCAAAACG
NdeI-BRD4(1)-fwd	CAGCCATATGAATCCGCCTC
BRD4(1)-HindIII-rev	ATATAAGCTTATTCTTCGGTCGGCAG
BamHI-Kozak-fwd	ATATGGATCCGCCACCATG
TEV-ClaI-rev	ATATATCGATGCTCTGGAAGTACAG
Primer for sequencing	**Sequence (5' to 3')**
CMV-Promotor-fwd	CGCAAATGGGCGGTAGGCGTG
pBI MCS1 rev	AGTCAGTGAGCGAGGAAG
pBI MCS2 rev	GTTCCGCGCACATTTCC
pCEP fwd	AGAGCTCGTTTAGTGAACCG
T7 Promotor fwd	TAATACGACTCACTATAGGG
T7 Terminator rev	GCTAGTTATTGCTCAGCGG

4.1.3.4 Bacterial strains and mammalian cell lines

Table 12: Overview of used bacterial strains and mammalian cell lines.

Bacterial strains	**Part-No.**	**Supplier**
BL21 (DE3) Competent *E. coli*	200131	Agilent
XL1-Blue Competent *E. coli*	200228	Agilent
Mammalian cell lines	**DSMZ-No.**	**Supplier**
HEK293T	ACC 635	DSMZ
Hela S3	ACC 161	DSMZ

4.1.4 Equipment

Table 13: Equipment and devices.

Equipment	Specification	Provider
Centrifuge	Heraeus Megafuge™ 16R	Thermo Fisher Scientific
	MiniSpin® plus	Eppendorf
	Sorvall™ RC6 Plus	Thermo Fisher Scientific
Electroporator	MicroPulser™	Bio-Rad
Flow cytometer	CytoFLEX S	Beckman Coulter
Freeze dryer	Alpha™ 2-4 LDplus	Martin Christ
Gel system	PerfectBlue Mini M	Peqlab
	Mini-PROTEAN® Tetra cell	Bio-Rad
	Mini Trans-Blot® Cell	Bio-Rad
Gel/Western Blot imager	ChemiDoc MP	Bio-Rad
Homogenizer	EmulsiFlex-C5	Avestin
HPLC, analytical	LC-20AD	Shimadzu
HPLC, preparative	ProStar 210	Varian
Incubator	Multitron Standard	Infors HT
	BB 150 CO_2	Thermo Fisher Scientific
LC-MS	LCMS-2020	Shimadzu
MALDI-TOF	Reflex IV	Bruker Daltonics
Microplate reader	Infinite 200 PRO	Tecan
Microscope	Examiner.Z1	Carl Zeiss
MST system	Monolith NT.115	NanoTemper
Peptide synthesizer	Syro I	Biotage
Rotary evaporator	RV 10	IKA
Shaker	ThermoMixer® Compact	Eppendorf
	ThermoMixer® Comfort	Eppendorf
	ThermoMixer® C	Eppendorf
Thermal cycler	TC-3000	Techne
UV-Vis spectrometer	GENESYS™ 10S	Thermo Fisher Scientific
	NanoDrop™ ND-1000	Peqlab

4.1.5 Software

Table 14: Software and their application.

Software	Application
ACD/NMR Processor 12.01	NMR data analysis
Adobe Illustrator CC 2021	Editing of vector graphics
Adobe Photoshop CC 2017	Editing of images
ApE - A plasmid Editor 3.0.3	DNA sequence analysis
ChemDraw Professional 19.1.0.5	Creating chemical structures
FlowJo 10.5.3	Flow cytometry data analysis
GraphPad Prism 9.0	Data statistics and diagrams
ImageJ 2.1.0/1.53c	Image processing
Image Lab 6.1	Gel and western blot analysis
LabSolutions LCMS	LC-MS data analysis
Microsoft Excel 2020	Data analysis
MO.Affinity Analysis v.2.3	MST data analysis
mMass 5.5.0	Mass spectra analysis
Perseus v1.6.15.0	Proteomics data analysis
PyMOL 2.5.1	Protein structure visualization
R 4.1.0	Proteomics data statistics
SnapGene 5.3	Plasmid editor

4.2 Methods

4.2.1 General chemical methods

4.2.1.1 Analytical HPLC

Analytical reversed-phase high-performance liquid chromatography (RP-HPLC) was performed on a Shimadzu LC10 system with a Reprospher 100 C18 column (5 µm, 100 Å, 4.6 × 250 mm, Dr. Maisch) and with eluents HPLC-A (0.1 % TFA in water) and HPLC-B (80 % ACN, 0.1 % TFA in water). Samples were eluted with an analytical gradient of 5-95 % HPLC-B over 20 min with a flow rate of 1.5 mL/min. Absorption was detected at 218 nm and 274 nm. Cyanine5 fluorescence was detected using a Jasco FP-2020 fluorescence detector with excitation at 649 nm and emission detection at 670 nm.

4.2.1.2 Preparative HPLC

Preparative purifications were carried out on a Varian ProStar 210 HPLC system equipped with a preparative Reprospher 100 C18 column (5 µm, 100 Å, 20 × 250 mm, Dr. Maisch) and a flow rate of 13 mL/min using eluents HPLC-A and HPLC-B. For semi-preparative purifications, a ReproSil 100 C18 column (5 µm, 100 Å, 10 × 250 mm, Dr. Maisch) was used with a flow rate of 4 mL/min. If not stated otherwise, samples were eluted with a gradient of 5-95 % HPLC-B over 40 min and absorption was detected at 218 nm. Collected fractions were analyzed by LC-MS and combined fractions were lyophilized.

4.2.1.3 Liquid chromatography mass spectrometry

Liquid chromatography mass spectrometry (LC-MS) measurements were performed on a Shimadzu LCMS-2020 System equipped with a Kinetex® C18 column (2.6 µm, 100 Å, 2.1 × 100 mm, Phenomenex) using the eluents LCMS-A (0.1 % formic acid (FA) in water) and LCMS-B (80 % ACN, 0.1 % FA in water). Samples were eluted with a flow rate of 0.2 mL/min using a gradient of 5-95 % LCMS-B over 12.75 min or 5-100 % LCMS-B over 17.75 min. UV-absorption was detected at 218 nm and the connected ESI-MS was operated in positive mode.

4.2.1.4 MALDI-TOF mass spectrometry

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry was performed on a Bruker Reflex IV MALDI-TOF in reflection mode. Equal volumes of sample and α-cyano-4-hydroxy-cinnamic acid (CHCA) as matrix (5 mg/mL CHCA, 0.1 % TFA, 50 % ACN in water) were spotted together on a polished steel target. Spots were measured with at least 200 shots at a laser intensity of 35-55 %. Measured mass spectra were analyzed with the mMass 5.5.0 software.

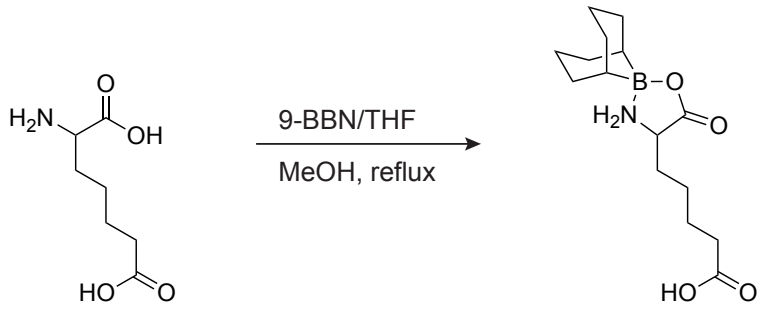
4.2.1.5 NMR spectroscopy

NMR spectra were recorded either on a Bruker Advance III HD 300 XWB or on a Bruker Advance III HD 400 spectrometer. The ^{1}H and ^{13}C chemical shifts are given in units of parts per million (ppm) relative to tetramethylsilane. ^{1}H spectra were referenced to the residual proton signal of deuterated solvents (DMSO-d_6: δ [ppm] = 2.50, D_2O: δ [ppm] = 4.80). ^{13}C spectra were referenced to the ^{13}C signal of the respective solvent (DMSO-d_6: δ [ppm] = 39.51). NMR spectra were analyzed with the ACD/NMR Processor 12.01 software.

4.2.2 Building block synthesis

4.2.2.1 Synthesis of triazole building blocks

Synthesis of 9-BBN-DL-Apm (2)



H-DL-Apm-OH **1** 9-BBN-DL-Apm **2**

DL-2-Aminoheptanedioic acid (H-DL-Apm-OH) (**1**) (700.8 mg, 4 mmol) was dissolved in anhydrous MeOH (25 mL) and heated at reflux under argon atmosphere. 9-Borabicyclo[3.3.1]nonane (9-BBN) (0.5 M in THF, 8.8 mL, 4.4 mmol, 1.1 eq) was added dropwise and the reaction mixture was stirred under reflux until the mixture became a clear solution (approximately 2 h). The solvent was evaporated under reduced pressure and the residue was lyophilized. The crude product was purified by preparative RP-HPLC (50-95 % B, 40 min). 9-BBN-DL-Apm (**2**) was obtained as a white powder (1.056 g, 3.58 mmol, 89 %).

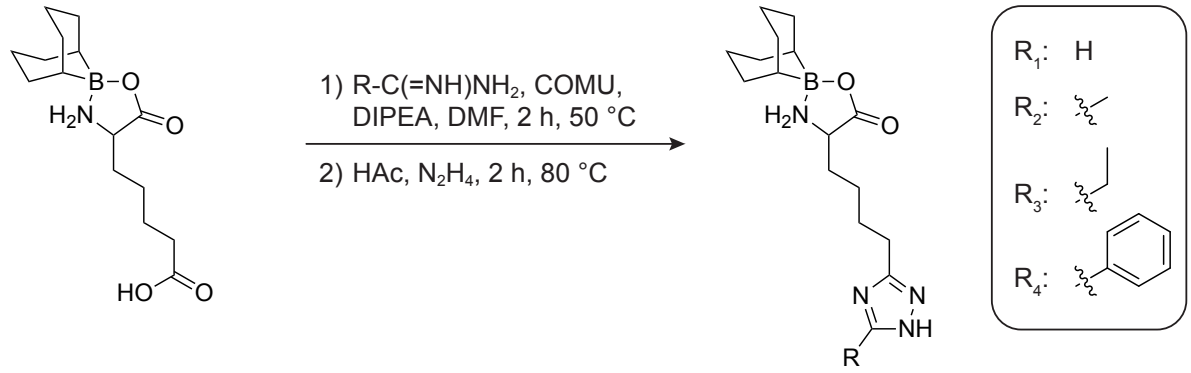
HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 12.17 min.

ESI-MS: $[M+H]^+$, m/z = 295.2 (calculated), 296.1 (found); $[2M\text{-}BBN+H]^+$, m/z = 471.3 (calculated), 471.2 (found); $[2M+H]^+$, m/z = 591.4 (calculated), 591.3 (found).

^{1}H NMR (300 MHz; DMSO-d_6): δ [ppm] = 0.46 (s, 1 H), 0.51 (s, 1 H), 1.28-1.67 (m, 12 H), 1.69-1.88 (m, 6 H), 2.22 (t, J=6.88 Hz, 2 H), 3.38-3.55 (m, 1 H), 5.67-5.86 (m, 1 H), 6.29-6.46 (m, 1 H), 11.97 (bs, 1 H).

^{13}C NMR (75 MHz; DMSO-d_6): δ [ppm] = 1.15, 22.32, 23.42, 23.94, 24.22, 24.26, 25.22, 30.18, 30.72, 31.25, 31.26, 33.55, 54.36, 173.69, 174.41.

Synthesis of 9-BBN-DL-Apm(1,2,4-triazole) derivatives



9-BBN-DL-Apm **2**

R_1: 9-BBN-DL-Apm(1,2,4-triazole) **3**
R_2: 9-BBN-DL-Apm(3-methyl-1,2,4-triazole) **4**
R_3: 9-BBN-DL-Apm(3-ethyl-1,2,4-triazole) **5**
R_4: 9-BBN-DL-Apm(3-phenyl-1,2,4-triazole) **6**

Synthesis of 9-BBN-DL-Apm(1,2,4-triazole) (3)

9-BBN-DL-Apm (**2**) (295.2 mg, 1 mmol, 1 eq) was dissolved in DMF (5 mL) and heated at 50 °C. (1-Cyano-2-ethoxy-2-oxoethyliden-aminooxy)dimethylamino-morpholino carbenium hexafluorophosphate (COMU) (471.1 mg, 1.1 mmol, 1.1 eq) and formamidine hydrochloride (161.04 mg, 2 mmol, 2 eq) were each dissolved in DMF (10 mL) separately. The COMU solution and *N,N*-Diisopropylethylamine (DIPEA) (1.02 mL, 6 mmol, 6 eq) were added to the heated amino acid solution and stirred for 5 min before the formamidine solution was added. The reaction took place for 2 h at 50 °C. In a second step, the triazole was synthesized by adding acetic acid (571.9 µL, 10 mmol, 10 eq) followed by hydrazine monohydrate ($N_2H_4 \cdot H_2O$) (72.75 µL, 1.5 mmol, 1.5 eq) and stirring of the reaction solution for additional 2 h at 80 °C. DMF was removed under reduced pressure and the residue was lyophilized to yield the crude product as a reddish oil. Purification was performed by RP-HPLC and 9-BBN-DL-Apm(1,2,4-triazole) (**3**) was obtained as a white solid (128 mg, 402 µmol, 40 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 10.00 min.

ESI-MS: $[M+H]^+$, m/z = 319.2 (calculated), 319.0 (found); $[2M+H]^+$, m/z = 637.4 (calculated), 637.4 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 0.46 (s, 1 H), 0.51 (s, 1 H), 1.33-1.65 (m, 18 H), 2.83 (t, *J*=7.58 Hz, 2 H), 3.42-3.54 (m, 1 H), 5.71-5.84 (m, 1 H), 6.36-6.49 (m, 1 H), 8.7 (s, 1 H).

^{13}C NMR (100 MHz; DMSO-d_6): δ [ppm] = 21.78, 22.33, 23.48, 23.94, 24.27, 24.98, 26.11, 26.50, 29.98, 30.73, 31.25, 31.28, 54.27, 145.36, 156.66, 173.71.

Synthesis of 9-BBN-DL-Apm(3-methyl-1,2,4-triazole) (4)

9-BBN-DL-Apm (**2**) (1.004 g, 3.4 mmol, 1 eq) was dissolved in DMF (20 mL) and heated

at 50 °C. COMU (1.602 g, 3.74 mmol, 1.1 eq) and acetamidine hydrochloride (1.929 g, 20.4 mmol, 6 eq) were each dissolved in DMF (30 mL) separately. The COMU solution and DIPEA (3.47 mL, 20.4 mmol, 6 eq) were added to the heated amino acid solution and stirred for 5 min before the acetamidine solution was added. The reaction took place for 2 h at 50 °C. The triazole was synthesized by adding acetic acid (1.944 mL, 34 mmol, 10 eq) followed by $N_2H_4 \cdot H_2O$ (247.3 µL, 5.1 mmol, 1.5 eq) and stirring of the reaction solution for additional 2 h at 80 °C. DMF was removed under reduced pressure and the residue was lyophilized to yield the crude product as a reddish oil. Purification was performed by RP-HPLC and 9-BBN-DL-Apm(3-methyl-1,2,4-triazole) (**4**) was obtained as a white solid (502 mg, 1.51 mmol, 43 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.30 min.

ESI-MS: $[M+H]^+$, m/z = 333.2 (calculated), 333.1 (found).

^{1}H NMR (300 MHz; DMSO-d$_6$): δ [ppm] = 0.46 (s, 1 H), 0.51 (s, 1 H), 1.37-1.43 (m, 18 H), 2.46 (s, 3 H), 2.80 (t, *J*=7.52 Hz, 2H), 3.43-3.55 (m, 1 H), 5.7-5.86 (m, 1 H), 6.36-6.49 (m, 1 H).

^{13}C NMR (75 MHz; DMSO-d$_6$): δ [ppm] = 10.83, 22.31, 23.50, 23.91, 24.25, 24.83, 24.89, 26.24, 29.94, 30.71, 31.24, 31.27, 54.23, 152.99, 156.46, 173.67.

Synthesis of 9-BBN-DL-Apm(3-ethyl-1,2,4-triazole) (5)

9-BBN-DL-Apm (**2**) (590.4 mg, 2 mmol, 1 eq) was dissolved in DMF (10 mL) and heated at 50 °C. COMU (942.2 mg, 2.2 mmol, 1.1 eq) and propionamidine hydrochloride (651.42 mg, 6 mmol, 3 eq) were each dissolved in DMF (15 mL) separately. The COMU solution and DIPEA (2.04 mL, 12 mmol, 6 eq) were added to the heated amino acid solution and stirred for 5 min before the propionamidine solution was added. The reaction took place for 2 h at 50 °C. The triazole was synthesized by adding acetic acid (1.143 mL, 20 mmol, 10 eq) followed by $N_2H_4 \cdot H_2O$ (145.5 µL, 3 mmol, 1.5 eq) and stirring of the reaction solution for additional 2 h at 80 °C. DMF was removed under reduced pressure and the residue was lyophilized to yield the crude product as a reddish oil. Purification was performed by RP-HPLC and 9-BBN-DL-Apm(3-ethyl-1,2,4-triazole) (**5**) was obtained as a white solid (326.3 mg, 942 µmol, 47 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.57 min.

ESI-MS: $[M+H]^+$, m/z = 347.2 (calculated), 347.1 (found).

^{1}H NMR (300 MHz; DMSO-d$_6$): δ [ppm] = 0.46 (s, 1 H), 0.51 (s, 1 H), 1.26 (t, *J*=7.64 Hz, 3 H), 1.38-1.82 (m, 18 H), 2.78-2.88 (m, 4 H), 3.45-3.54 (m, 1 H), 5.74-5.85 (m, 1 H), 6.37-6.49 (m, 1 H).

^{13}C NMR (75 MHz; DMSO-d$_6$): δ [ppm] = 11.10, 18.64, 22.32, 23.46, 23.91, 24.24, 24.80, 24.88, 26.18, 29.91, 30.69, 30.72, 31.22, 31.26, 54.21, 156.35, 157.46, 173.65.

Synthesis of 9-BBN-DL-Apm(3-phenyl-1,2,4-triazole) (6)

9-BBN-DL-Apm (**2**) (590.4 mg, 2 mmol, 1 eq) was dissolved in DMF (10 mL) and heated at 50 °C. COMU (942.2 mg, 2.2 mmol, 1.1 eq) and benzamidine hydrochloride hydrate (626.44 mg, 4 mmol, 2 eq) were each dissolved in DMF (15 mL) separately. The COMU solution and DIPEA (2.04 mL, 12 mmol, 6 eq) were added to the heated amino acid solution and stirred for 5 min before the benzamidine solution was added. The reaction took place for 2 h at 50 °C. The triazole was synthesized by adding acetic acid (1.143 mL, 20 mmol, 10 eq) followed by $N_2H_4 \cdot H_2O$ (145.5 µL, 3 mmol, 1.5 eq) and stirring of the reaction solution for additional 2 h at 80 °C. DMF was removed under reduced pressure and the residue was lyophilized to yield the crude product as a reddish oil. Purification was performed by RP-HPLC and 9-BBN-DL-Apm(3-phenyl-1,2,4-triazole) (**6**) was obtained as a white solid (446 mg, 1.13 mmol, 57 %).

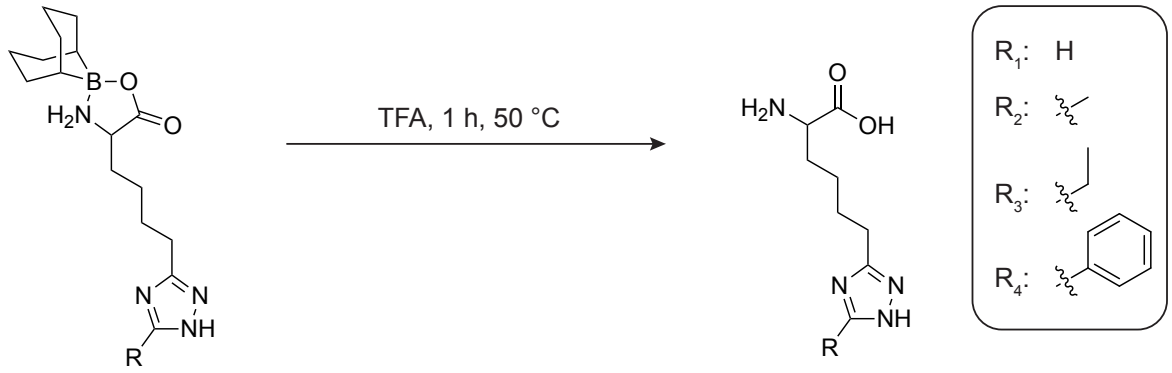
HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 12.13 min.

ESI-MS: $[M+H]^+$, m/z = 395.2 (calculated), 395.1 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 0.47 (s, 1 H), 0.52 (s, 1 H), 1.36-1.91 (m, 18 H), 2.79 (t, *J*=7.58 Hz, 2 H), 3.45-3.55 (m, 2 H), 5.73-5.83 (m, 1 H), 6.36-6.48 (m, 1 H), 7.41-7.51 (m, 3 H), 7.97-8.01 (m, 2 H).

^{13}C NMR (100 MHz; DMSO-d_6): δ [ppm] = 22.34, 23.47, 23.92, 24.26, 25.16, 25.85, 27.03, 30.13, 30.72, 31.23, 31.26, 54.33, 125.90, 128.79, 129.41, 129.65, 158.22, 158.58, 173.71.

Synthesis of H-DL-Apm(1,2,4-triazole)-OH derivatives



9-BBN-DL-Apm(triazole)
R_1: **3** R_3: **5**
R_2: **4** R_4: **6**

R_1: H-DL-Apm(1,2,4-triazole)-OH	**7**
R_2: H-DL-Apm(3-methyl-1,2,4-triazole)-OH	**8**
R_3: H-DL-Apm(3-ethyl-1,2,4-triazole)-OH	**9**
R_4: H-DL-Apm(3-phenyl-1,2,4-triazole)-OH	**10**

Synthesis of H-DL-Apm(1,2,4-triazole)-OH (7)

9-BBN-DL-Apm(1,2,4-triazole)-OH (**3**) (318.2 mg, 1 mmol) was deprotected with TFA (20 mL) at 50 °C for 1 h under stirring before TFA was removed under reduced pressure. The brown residue was dissolved in water (3 mL) and washed with ethyl acetate

(3 × 10 mL). The aqueous phase was lyophilized and H-DL-Apm(1,2,4-triazole)-OH · TFA (**7**) was obtained as a white crystalline solid (162.5 mg, 820 µmol, 82 %).

ESI-MS: $[M+H]^+$, *m/z* = 199.1 (calculated), 199.2 (found).

^{1}H NMR (400 MHz; D_2O): δ [ppm] = 1.31-1.52 (m, 2 H), 1.73-1.95 (m, 4 H), 2.96 (t, *J*=7.64 Hz, 2 H), 3.84 (t, *J*=6.24 Hz, 1 H), 8.75 (s, 1 H).

^{13}C NMR (100 MHz; D_2O): δ [ppm] = 23.49, 23.77, 25.57, 29.39, 53.41, 142.85, 155.72, 173.09.

Synthesis of H-DL-Apm(3-methyl-1,2,4-triazole)-OH (8)

9-BBN-DL-Apm(3-methyl-1,2,4-triazole)-OH (**4**) (810 mg, 2.44 mmol) was deprotected with TFA (20 mL) at 50 °C for 1 h under stirring before TFA was removed under reduced pressure. The brown residue was dissolved in water (10 mL) and washed with ethyl acetate (4 × 10 mL). The aqueous phase was lyophilized and H-DL-Apm(3-methyl-1,2,4-triazole)-OH · TFA (**8**) was obtained as a white crystalline solid (770 mg, 2.36 mmol, 97 %).

ESI-MS: $[M+H]^+$, m/z = 213.1 (calculated), 213.1 (found).

Synthesis of H-DL-Apm(3-ethyl-1,2,4-triazole)-OH (9)

9-BBN-DL-Apm(3-ethyl-1,2,4-triazole)-OH (**5**) (346.3 mg, 1 mmol) was deprotected with TFA (20 mL) at 50 °C for 1 h under stirring before TFA was removed under reduced pressure. The brown residue was dissolved in water (3 mL) and washed with ethyl acetate (3 × 10 mL). The aqueous phase was lyophilized and H-DL-Apm(3-ethyl-1,2,4-triazole)-OH · TFA (**9**) was obtained as a white crystalline solid (190 mg, 840 µmol, 84 %).

ESI-MS: $[M+H]^+$, *m/z* = 227.1 (calculated), 227.0 (found).

^{1}H NMR (400 MHz; D_2O): δ [ppm] = 1.30 (t, *J*=7.64 Hz, 3 H), 1.34 - 1.55 (m, 2 H), 1.72 - 1.83 (m, 2 H), 1.84-1.97 (m, 2 H), 2.88 - 2.97 (m, 4 H), 3.83 (t, *J*=6.17 Hz, 1 H).

^{13}C NMR (100 MHz; D_2O): δ [ppm] = 9.85, 18.02, 23.48, 23.93, 25.52, 29.50, 53.67, 155.64, 157.11, 173.45.

Synthesis of H-DL-Apm(3-phenyl-1,2,4-triazole)-OH (10)

9-BBN-DL-Apm(3-phenyl-1,2,4-triazole)-OH (**6**) (993.1 mg, 2 mmol) was deprotected with TFA (20 mL) at 50 °C for 1 h under stirring before TFA was removed under reduced pressure. The brown residue was dissolved in water (10 mL) and washed with ethyl acetate (3 × 10 mL). The aqueous phase was lyophilized and H-DL-Apm(3-phenyl-1,2,4-triazole)-OH · TFA (**10**) was obtained as a white crystalline solid (310 mg, 1.13 mmol, 57 %).

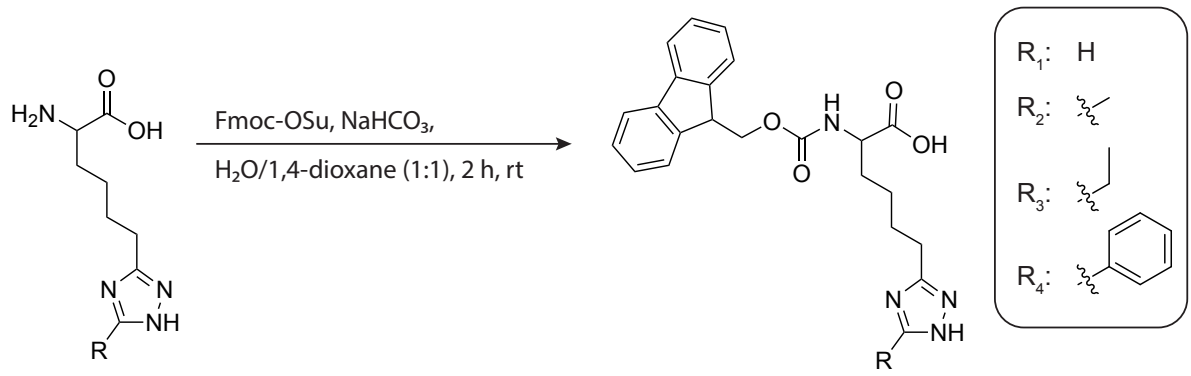
HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 6.00 min.

ESI-MS: $[M+H]^+$, *m/z* = 275.1 (calculated), 275.2 (found).

^{1}H NMR (400 MHz; D_2O): δ [ppm] = 1.34-1.56 (m, 2 H), 1.76 (quin, *J*=7.64 Hz, 2 H), 2.80-2.91 (m, 2 H), 3.88 (t, *J*=6.24 Hz, 1 H), 7.41-7.55 (m, 3 H), 7.68-7.75 (m, 2 H).

^{13}C NMR (100 MHz; D_2O): δ [ppm] = 23.60, 24.32, 25.88, 29.45, 53.41, 124.04, 126.48, 129.31, 131.87, 155.20, 157.14, 173.05.

Synthesis of Fmoc-DL-Apm(1,2,4-triazole)-OH derivatives



H-DL-Apm(triazole)-OH
R_1: **7** R_3: **9**
R_2: **8** R_4: **10**

R_1: Fmoc-DL-Apm(1,2,4-triazole)-OH **11**
R_2: Fmoc-DL-Apm(3-methyl-1,2,4-triazole)-OH **12**
R_3: Fmoc-DL-Apm(3-ethyl-1,2,4-triazole)-OH **13**
R_4: Fmoc-DL-Apm(3-phenyl-1,2,4-triazole)-OH **14**

Synthesis of Fmoc-DL-Apm(1,2,4-triazole)-OH (11)

H-DL-Apm(1,2,4-triazole)-OH · TFA (**7**) (162.25 mg, 820 µmol, 1 eq) was dissolved in water and the pH value was adjusted to 8-9 with $NaHCO_3$ (1 M in water) receiving a final volume of 20 mL aqueous solution. *N*-(9-Fluorenylmethoxycarbonyloxy)succinimid (Fmoc-OSu) (304.2 mg, 902 µmol, 1.1 eq) was dissolved in 1,4-dioxane (20 mL) and added to the amino acid solution. The reaction mixture was stirred for 2 h at room temperature before the solvent was reduced under vacuum and the residue lyophilized. The crude product was purified by RP-HPLC (40-70 % B, 40 min). Fmoc-DL-Apm(1,2,4-triazole)-OH (**11**) was obtained as a white solid (191 mg, 454 µmol, 55 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 10.37 min.

ESI-MS: $[M+H]^+$, *m/z* = 421.2 (calculated), 421.1 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 1.30-1.44 (m, 2 H), 1.56-1.81 (m, 4 H), 2.78 (t, *J*=7.58 Hz, 2 H), 3.90 - 3.97 (m, 1 H) 4.20 - 4.25 (m, 1 H), 4.26-4.33 (m, 2H), 7.30-7.35 (m, 2 H), 7.41 (t, *J*=7.40 Hz, 2 H), 7.64 (d, *J*=7.95 Hz, 1 H), 7.69-7.77 (m, 2 H), 7.89 (d, *J*=7.46 Hz, 2 H), 8.54 (s, 1 H).

^{13}C NMR (101 MHz; DMSO-d_6): δ [ppm] = 25.08, 25.22, 26.59, 30.35, 46.69, 53.69, 65.62, 120.13, 125.30, 127.10, 127.67, 140.75, 143.83, 143.88, 146.01, 156.21, 173.94.

Synthesis of Fmoc-DL-Apm(3-methyl-1,2,4-triazole)-OH (12)

H-DL-Apm(3-methyl-1,2,4-triazole)-OH · TFA (**8**) (501.7 mg, 1.51 mmol, 1 eq) was dissolved in water and the pH value was adjusted to 8-9 with $NaHCO_3$ (1 M in water) receiving a final volume of 20 mL aqueous solution. Fmoc-OSu (560 mg, 1.66 mmol, 1.1 eq) was dissolved in 1,4-dioxane (20 mL) and added to the aqueous solution. The reaction mixture was stirred for 2 h at room temperature before the solvent was reduced under vacuum and the residue lyophilized. The crude product was purified by RP-HPLC (40-70 %, 40 min). Fmoc-DL-Apm(3-methyl-1,2,4-triazole)-OH (**12**) was obtained as a white solid (399 mg, 918 µmol, 61 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.67 min.

ESI-MS: $[M+H]^+$, m/z = 435.2 (calculated), 435.1 (found).

^{1}H NMR (300 MHz; DMSO-d_6): δ [ppm] = 1.30-1.47 (m, 2 H) 1.56-1.81 (m, 4 H) 2.41 (s, 3 H), 2.75 (t, *J*=7.52 Hz, 2 H), 3.87-3.99 (m, 1 H), 4.17-4.25 (m, 1 H), 4.26-4.34 (m, 2 H), 7.28-7.37 (m, 2 H), 7.42 (t, *J*=7.11 Hz, 2 H), 7.64 (d, *J*=8.07 Hz, 1 H), 7.72 (d, *J*=7.34 Hz, 2 H), 7.89 (d, *J*=7.52 Hz, 2 H).

^{13}C NMR (75 MHz; DMSO-d_6): δ [ppm] = 11.12, 25.02, 25.09, 26.32, 30.32, 46.68, 53.66, 65.60, 120.12, 125.29, 127.07, 127.65, 140.74, 143.81, 143.86, 153.42, 156.19, 173.91.

Synthesis of Fmoc-DL-Apm(3-ethyl-1,2,4-triazole)-OH (13)

H-DL-Apm(3-ethyl-1,2,4-triazole)-OH · TFA (**9**) (190 mg, 840 µmol, 1 eq) was dissolved in water and the pH value was adjusted to 8-9 with $NaHCO_3$ (1 M in water) receiving a final volume of 20 mL aqueous solution. Fmoc-OSu (560 mg, 1.66 mmol, 1.1 eq) was dissolved in 1,4-dioxane (20 mL) and added to the aqueous solution. The reaction mixture was stirred for 2 h at room temperature before the solvent was reduced under vacuum and the residue lyophilized. The crude product was purified by RP-HPLC (40-70 % B, 40 min). Fmoc-DL-Apm(3-ethyl-1,2,4-triazole)-OH (**13**) was obtained as a white solid (239 mg, 533 µmol, 63 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.96 min.

ESI-MS: $[M+H]^+$, m/z = 449.2 (calculated), 449.4 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 1.24 (t, *J*=7.64 Hz, 3 H), 1.33-1.45 (m, 2 H), 1.58-1.80 (m, 4 H), 2.72 - 2.85 (m, 4 H), 3.89 - 3.99 (m, 1 H), 4.18 - 4.25 (m, 1 H), 4.26 - 4.33 (m, 2 H), 7.28 - 7.37 (m, 2 H), 7.42 (t, *J*=7.34 Hz, 2 H), 7.64 (d, *J*=8.19 Hz, 1 H), 7.69-7.75 (m, 2 H), 7.89 (d, *J*=7.58 Hz, 2 H).

^{13}C NMR (100 MHz; DMSO-d_6): δ [ppm] = 11.25, 18.85, 25.03, 26.25, 30.29, 46.67, 53.63, 65.60, 120.12, 125.28, 127.07, 127.65, 140.73, 140.74, 143.80, 143.86, 156.19, 156.72, 173.91.

Synthesis of Fmoc-DL-Apm(3-phenyl-1,2,4-triazole)-OH (14)

H-DL-Apm(3-phenyl-1,2,4-triazole)-OH · TFA (**10**) (310 mg, 1.13 mmol, 1 eq) was dissolved in water and the pH value was adjusted to 8-9 with $NaHCO_3$ solution (1 M in water) receiving a final volume of 20 mL aqueous solution. Fmoc-OSu (560 mg, 1.66 mmol, 1.1 eq) was dissolved in 1,4-dioxane (20 mL) and added to the aqueous solution. The reaction mixture was stirred for 2 h at room temperature before the solvent was reduced under vacuum and the residue lyophilized. The crude product was purified by RP-HPLC (50-80 %, 40 min). Fmoc-DL-Apm(3-phenyl-1,2,4-triazole)-OH (**14**) was obtained as a white solid (360 mg, 725 µmol, 64 %).

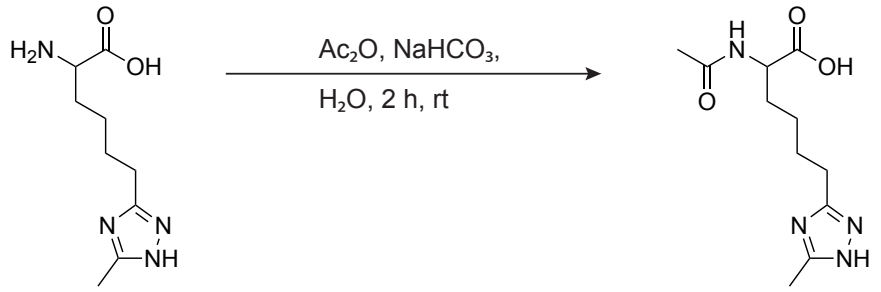
HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 12.38 min.

ESI-MS: $[M+H]^+$, m/z = 497.2 (calculated), 497.1 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 1.35-1.51 (m, 2 H), 1.61-1.82 (m, 4 H), 2.79 (t, *J*=7.58 Hz, 2 H), 3.91 - 4.00 (m, 1 H), 4.19 - 4.24 (m, 1 H), 4.25 - 4.32 (m, 2 H), 7.29 - 7.34 (m, 2 H), 7.37 - 7.52 (m, 5 H), 7.65 (d, *J*=8.07 Hz, 1 H), 7.69 - 7.74 (m, 2 H), 7.88 (d, *J*=7.58 Hz, 2 H), 7.97-8.02 (m, 2 H).

^{13}C NMR (101 MHz; DMSO-d_6): δ [ppm] = 25.21, 25.72, 26.85, 30.43, 46.68, 53.72, 65.61, 120.12, 125.31, 125.98, 127.09, 127.65, 128.86, 129.60, 140.74, 140.76, 143.82, 143.88, 156.21, 158.24, 158.61, 173.98.

Synthesis of Ac-DL-Apm(3-methyl-1,2,4-triazole)-OH (15)



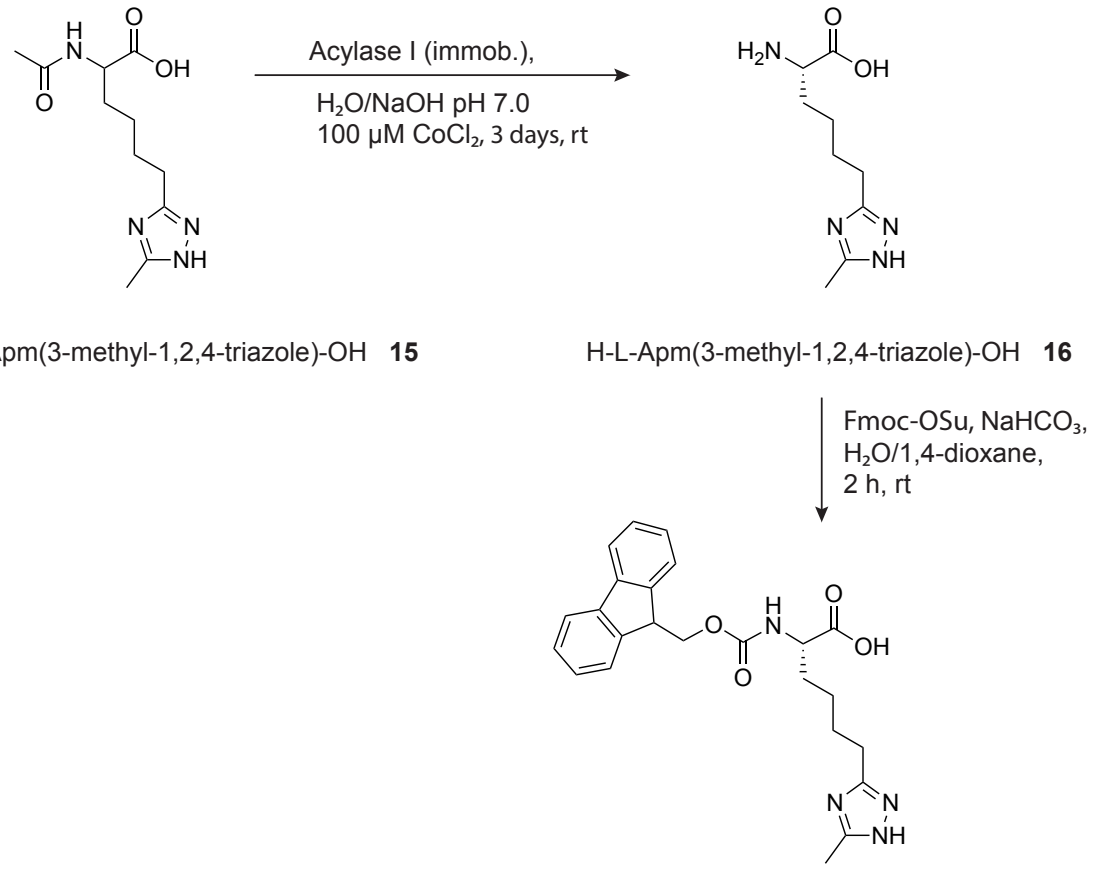
H-DL-Apm(3-methyl-1,2,4-triazole)-OH **8** Ac-DL-Apm(3-methyl-1,2,4-triazole)-OH **15**

H-DL-Apm(3-methyl-1,2,4-triazole)-OH · TFA (**8**) (454.2 mg, 1.392 mmol, 1 eq) was dissolved in water (10 mL) and the pH value was adjusted to 8-9 with $NaHCO_3$ (1 M in water). Acetic anhydride (263.2 µL, 2.784 mmol, 2 eq) was added dropwise to the solution and the mixture was stirred for 2 h at room temperature. The solution was lyophilized and the crude product was purified by RP-HPLC. Ac-DL-Apm(3-methyl-1,2,4-triazole)-OH (**15**) was obtained as colorless solid (326 mg, 1.282 mmol, 92 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 4.14 min.

ESI-MS: $[M+H]^+$, m/z = 255.1 (calculated), 255.2 (found).

Synthesis of Fmoc-L-Apm(3-methyl-1,2,4-triazole)-OH (17)



Ac-DL-Apm(3-methyl-1,2,4-triazole)-OH **15**

H-L-Apm(3-methyl-1,2,4-triazole)-OH **16**

Fmoc-L-Apm(3-methyl-1,2,4-triazole)-OH **17**

Ac-DL-Apm(3-methyl-1,2,4-triazole)-OH (**15**) (326 mg, 1.282 mmol) was dissolved in water (10 mL) and the pH value was adjusted to pH 7.0 by adding NaOH. $CoCl_2$ was added to receive a concentration of 100 µM in a final volume of 1.5 mL. The reaction mixture was incubated with acylase I (from Aspergillus sp., immobilized on Eupergit® C, Sigma Alrich) (20 U, 300 mg) for 3 days at room temperature. The solution containing deacetylated H-L-Apm(3-methyl-1,2,4-triazole)-OH (**16**) was directly used for Fmoc-protection without further purification.

The pH value of the reaction solution was adjusted to 8-9 with $NaHCO_3$ solution (1 M in water) receiving a final volume of 20 mL aqueous solution. Fmoc-OSu (216.21 mg, 0.641 mmol, 1 eq with respect to 50 % of (**15**)) was dissolved in dioxane (20 mL) and added to the aqueous solution. The reaction took place for 2 h at room temperature. The solvent was evaporated under vacuum and the residue was lyophilized. Purification by RP-HPLC (40-70 % B, 40 min) yielded Fmoc-L-Apm(3-methyl-1,2,4-triazole)-OH (Fmoc-L-ApmTri-OH) (**17**) as a white powder (196 mg, 451 µmol, 70.4 % with respect to 50 % of (**15**)) and a non-racemic mixture of Ac-D-Apm(3-methyl-1,2,4-triazole)-OH and Ac-L-Apm(3-methyl-1,2,4-triazole)-OH (**15***) as colorless solid (122 mg, 480 µmol).

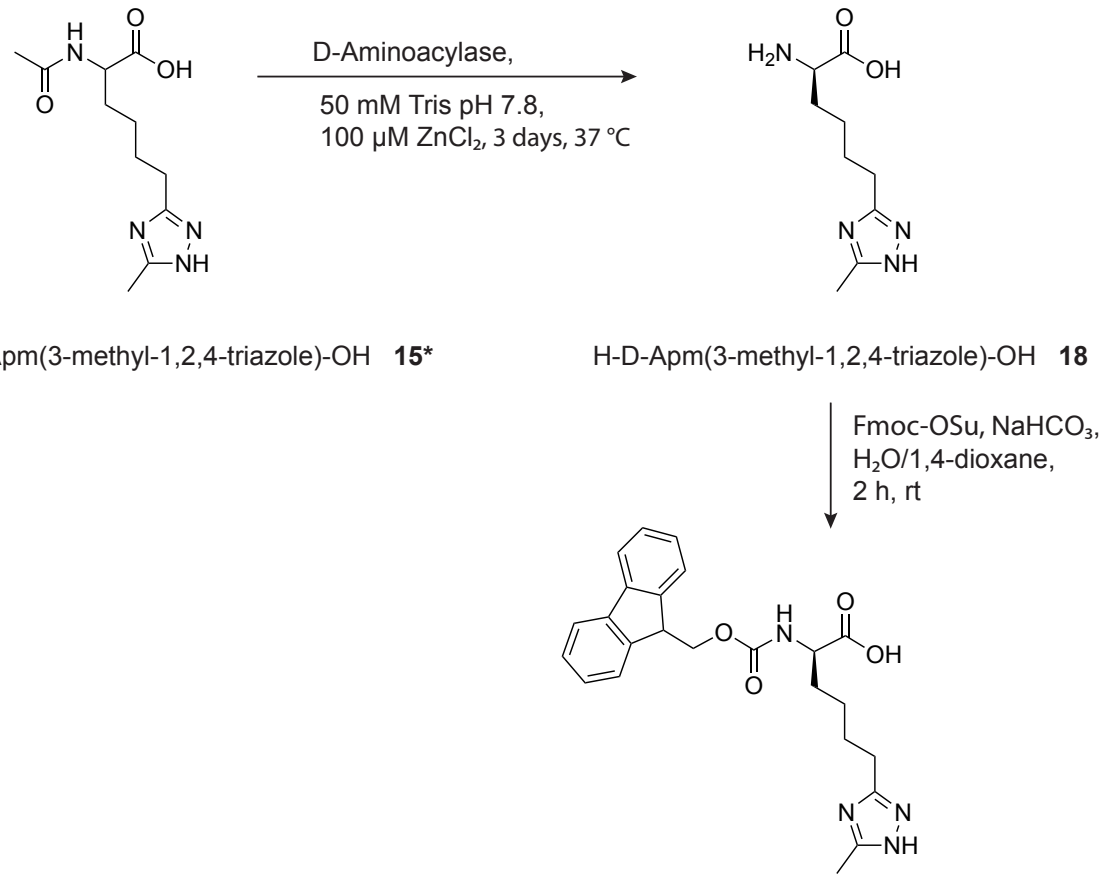
HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.52 min.

ESI-MS: $[M+H]^+$, m/z = 435.2 (calculated), 435.1 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 1.30-1.46 (m, 2 H), 1.56-1.80 (m, 4 H), 2.43 (s, 3 H), 2.76 (t, J=7.52 Hz, 2 H), 3.84-4.00 (m, 1 H,) 4.18-4.25 (m, 1 H), 4.26-4.33 (m, 2 H), 7.25-7.38 (m, 2 H), 7.42 (t, J=7.20 Hz, 2 H), 7.64 (d, J=8.07 Hz, 1 H), 7.72 (d, J=7.43 Hz, 2 H), 7.89 (d, J=7.52 Hz, 2 H).

^{13}C NMR (101 MHz; DMSO-d_6): δ [ppm] = 10.98, 24.94, 24.99, 26.21, 30.29, 46.67, 53.63, 65.60, 120.13, 125.29, 127.08, 127.66, 140.74, 143.81, 143.86, 153.20, 156.19, 156.63, 173.91.

Synthesis of Fmoc-D-Apm(3-methyl-1,2,4-triazole)-OH (19)



Ac-DL-Apm(3-methyl-1,2,4-triazole)-OH **15***

H-D-Apm(3-methyl-1,2,4-triazole)-OH **18**

Fmoc-D-Apm(3-methyl-1,2,4-triazole)-OH **19**

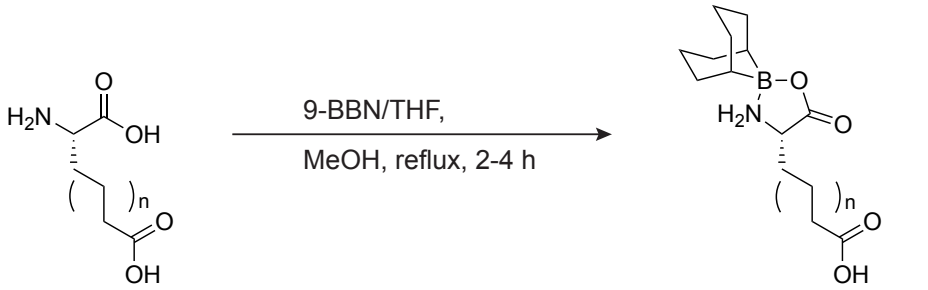
The mixture containing Ac-D-Apm(3-methyl-1,2,4-triazole)-OH (**15***) (122 mg, 480 µmol, 1 eq) was dissolved in D-acylase buffer (50 mM Tris-HCl, 100 µM $ZnCl_2$, pH 7.6) and the pH value was adjusted to pH 7.6 by adding NaOH. D-Aminoacylase (recombinant from *E. coli*, Sigma-Aldrich) (3140 U, 10 mg) was dissolved in acylase buffer and added to the amino acid solution receiving a final volume of 3 mL. After incubation of the reaction mixture for 3 days at 37 °C the enzyme was precipitated by acidifying with HCl and incubation for 1 h at 95 °C. Precipitate was removed by centrifugation (14,100 g,

10 min) and supernatant was neutralized with NaOH to pH 7.0 and filled up to a volume of 20 mL with $NaHCO_3$ solution (1 M in water). Fmoc-OSu (216.21 mg, 0.641 mmol, 1 eq) was dissolved in 1,4-dioxane (20 mL) and added to the aqueous solution. Reaction took place for 2 h at room temperature. The solution was concentrated under reduced pressure and the residue was lyophilized. Purification by RP-HPLC (40-70 % B, 40 min) yielded Fmoc-D-Apm(3-methyl-1,2,4-triazole)-OH (Fmoc-D-ApmTri-OH) (**19**) as a white powder (135 mg, 311 µmol, 64.8 % with respect to (**15***)).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.80 min.

ESI-MS: $[M+H]^+$, m/z = 435.2 (calculated), 435.1 (found).

Synthesis of 9-BBN-L-Aad (22) and 9-BBN-L-Asu (23)



H-L-Aad-OH (n=1) **20**
H-L-Asu-OH (n=3) **21**

9-BBN-L-Aad (n=1) **22**
9-BBN-L-Asu (n=3) **23**

Synthesis of 9-BBN-L-Aad (22)

L-2-Aminoadipic acid (H-L-Aad-OH) (**20**) (966.96 mg, 6 mmol, 1 eq) was dissolved in anhydrous MeOH (30 mL) and heated at reflux under argon atmosphere. 9-BBN (0.5 M in THF, 18 mL, 9 mmol, 1.5 eq) was added dropwise and the reaction mixture was stirred under reflux until the mixture became a clear solution (approximately 4 h). The solvent was evaporated under reduced pressure and the residue lyophilized. The crude product was purified by preparative RP-HPLC (50-95 % B, 40 min). 9-BBN-L-Aad (**22**) was obtained as a white powder (2.514 g, 5.385 mmol, 90 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 10.25 min.

ESI-MS: $[M+H]^+$, m/z = 282.2 (calculated), 282.1 (found); $[2M\text{-}9BBN+H]^+$, m/z = 443.2 (calculated), 443.1 (found); $[2M+H]^+$, m/z = 563.3 (calculated), 563.1 (found); $[2M+Na]^+$, m/z = 585.3 (calculated), 585.1 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 0.47 (s, 1 H), 0.51 (s, 1 H), 1.32-1.86 (m, 16 H), 2.16-2.28 (m, 2 H), 3.44-3.55 (m, 1 H), 5.74-5.87 (m, 1 H), 6.36-6.50 (m, 1 H), 11.8 (bs, 1 H).

^{13}C NMR (101 MHz; DMSO-d_6): δ [ppm] = 1.18, 21.28, 22.32, 23.46, 23.96, 24.28, 29.87, 30.72, 31.26, 31.28, 33.31, 54.21, 173.64, 174.37.

Synthesis of 9-BBN-L-Asu (23)

L-2-Aminosuberic acid (H-L-Asu-OH) (**21**) (1.514 g, 8 mmol, 1 eq) was dissolved in anhydrous MeOH (30 mL) and heated at reflux under argon atmosphere. 9-BBN (0.5 M in THF, 17.6 mL, 8.8 mmol, 1.1 eq) was added dropwise and the reaction mixture was stirred under reflux until the mixture became a clear solution (approximately 2 h). The solvent was evaporated under reduced pressure and the residue lyophilized. The crude product was purified by preparative RP-HPLC (50-95 % B, 40 min). 9-BBN-L-Asu (**23**) was obtained as a white powder (2.282 g, 7.38 mmol, 92 %).

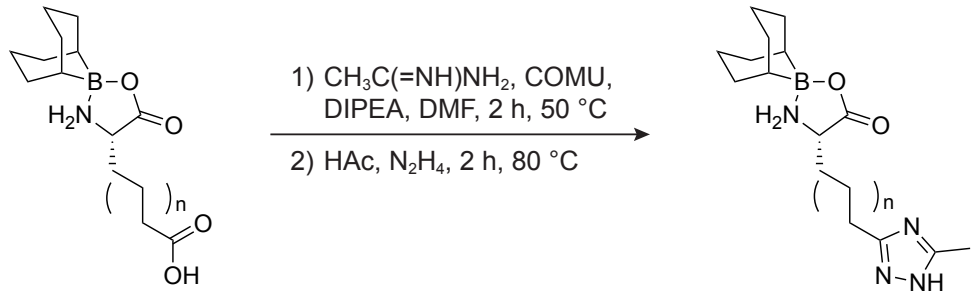
HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 11.31 min.

ESI-MS: $[M+H]^+$, m/z = 282.2 (calculated), 310.1 (found); $[2M\text{-}9BBN+H]^+$, m/z = 499.3 (calculated), 499.1 (found); $[2M+H]^+$, m/z = 619.4 (calculated), 619.2 (found); $[2M+Na]^+$, m/z = 641.4 (calculated), 641.2 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 0.46 (s, 1 H), 0.51 (s, 1 H), 1.23-1.82 (m, 20 H), 2.21 (t, J=7.4 Hz, 2 H), 3.42-3.52 (m, 1 H), 5.71-5.82 (m, 1 H), 6.33-6.44 (m, 1 H), 11.96 (bs, 1 H).

^{13}C NMR (101 MHz; DMSO-d_6): δ [ppm] = 1.18, 22.35, 23.49, 23.97, 24.30, 25.28, 28.24, 30.29, 30.72, 30.75, 31.26, 31.29, 33.56, 54.47, 173.76, 174.53.

Synthesis of 9-BBN-AadTri (24) and 9-BBN-AsuTri (25)



9-BBN-L-Aad (n=1) **22**
9-BBN-L-Asu (n=3) **23**

9-BBN-L-Aad(3-methyl-1,2,4-triazole) (n=1) **24**
9-BBN-L-Asu(3-methyl-1,2,4-triazole) (n=3) **25**

Synthesis of 9-BBN-L-Aad(3-methyl-1,2,4-triazole) (24)

9-BBN-L-Aad (**22**) (562.32 mg, 2 mmol, 1 eq) was dissolved in DMF (10 mL) and heated at 50 °C. COMU (942.2 mg, 2.2 mmol, 1.1 eq) and acetamidine hydrochloride (1.134 g, 12 mmol, 6 eq) were each dissolved in DMF (20 mL) separately. The COMU solution and DIPEA (2.04 mL, 12 mmol, 6 eq) were added to the heated amino acid solution and stirred for 5 min, before the acetamidine solution was added. The reaction took place for 2 h at 50 °C. The triazole was synthesized by adding acetic acid (1.143 mL, 20 mmol, 10 eq) followed by $N_2H_4 \cdot H_2O$ (145.5 µL, 3 mmol, 1.5 eq) and stirring of the reaction solution for additional 2 h at 80 °C. DMF was removed under reduced pressure and the

residue was lyophilized to yield the crude product as a reddish oil. Purification was performed by RP-HPLC and 9-BBN-L-Aad(3-methyl-1,2,4-triazole) (9-BBN-L-AadTri) (**24**) was obtained as a white solid (372 mg, 1.169 mmol, 58 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.06 min.

ESI-MS: $[M+H]^+$, m/z = 319.2 (calculated), 319.2 (found).

^{1}H NMR (300 MHz; DMSO-d_6): δ [ppm] = 0.47 (s, 1 H), 0.51 (s, 1 H), 1.30 - 1.95 (m, 16 H), 2.45 (s, 3 H), 2.47-2.86 (m, 2 H), 3.48-3.61 (m, 1 H), 5.80-5.92 (m, 1 H), 6.41-6.53 (m, 1 H).

^{13}C NMR (75 MHz; DMSO-d_6): δ [ppm] = 1.15, 10.96, 22.27, 23.39, 23.91, 24.24, 24.82, 29.63, 30.72, 31.22, 31.25, 54.00, 153.20, 156.65, 173.53.

Synthesis of 9-BBN-L-Asu(3-methyl-1,2,4-triazole) (25)

9-BBN-L-Asu (**23**) (618.42 mg, 2 mmol, 1 eq) was dissolved in DMF (10 mL) and heated at 50 °C. COMU (942.2 mg, 2.2 mmol, 1.1 eq) and acetamidine hydrochloride (1.134 g, 12 mmol, 6 eq) were each dissolved in DMF (20 mL) separately. The COMU solution and DIPEA (2.04 mL, 12 mmol, 6 eq) were added to the heated amino acid solution and stirred for 5 min, before the acetamidine solution was added. The reaction took place for 2 h at 50 °C. The triazole was synthesized by adding acetic acid (1.143 mL, 20 mmol, 10 eq) followed by $N_2H_4 \cdot H_2O$ (145.5 µL, 3 mmol, 1.5 eq) and stirring of the reaction solution for additional 2 h at 80 °C. DMF was removed under reduced pressure and the residue was lyophilized to yield the crude product as a reddish oil. Purification was performed by RP-HPLC and 9-BBN-L-Asu(3-methyl-1,2,4-triazole) (9-BBN-L-AsuTri) (**25**) was obtained as a white solid (462 mg, 1.323 mmol, 66 %).

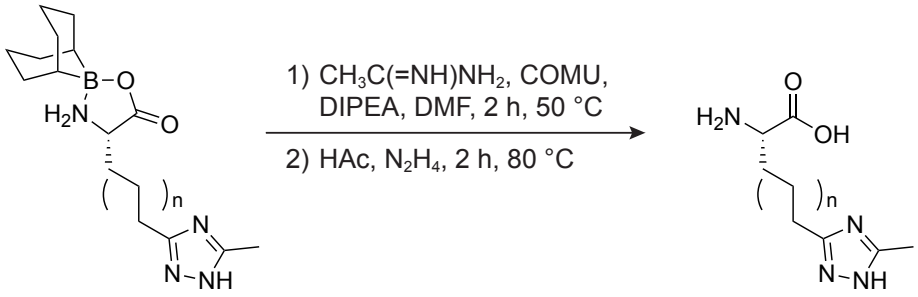
HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.27 min.

ESI-MS: $[M+H]^+$, m/z = 347.3 (calculated), 347.3 (found).

^{1}H NMR (300 MHz; DMSO-d_6): δ [ppm] = 0.46 (s, 1 H), 0.51 (s, 1 H), 1.26-1.84 (m, 20 H), 2.47 (s, 3 H), 2.80 (t, J=7.57 Hz, 2 H), 3.42-3.54 (m, 1 H), 5.70-5.84 (m, 1 H), 6.35-6.51 (m, 1 H), 14.23 (bs, 1 H).

^{13}C NMR (75 MHz; DMSO-d_6): δ [ppm] = 10.76, 22.36, 23.49, 23.92, 24.26, 24.81, 25.04, 26.34, 27.98, 30.21, 30.70, 31.25, 31.27, 54.39, 152.84, 156.30, 173.72.

Synthesis of H-L-AadTri-OH (26) and H-L-AsuTri-OH (27)



9-BBN-L-Aad(3-methyl-1,2,4-triazole) (n=1) **24**
9-BBN-L-Asu(3-methyl-1,2,4-triazole) (n=3) **25**

H-L-Aad(3-methyl-1,2,4-triazole)-OH (n=1) **26**
H-L-Asu(3-methyl-1,2,4-triazole)-OH (n=3) **27**

Synthesis of H-L-Aad(3-methyl-1,2,4-triazole)-OH (26)

9-BBN-L-Aad(3-methyl-1,2,4-triazole) (**24**) (318.23 mg, 1 mmol) was deprotected with TFA (20 mL) at 50 °C for 1 h under stirring before TFA was removed under reduced pressure. The brown residue was dissolved in water (10 mL) and washed with ethyl acetate (4 × 10 mL). The aqueous phase was lyophilized and H-L-Aad(3-methyl-1,2,4-triazole)-OH · TFA (L-AadTri) (**26**) was obtained as a white crystalline solid (295 mg, 945 µmol, 94 %).

ESI-MS: $[M+H]^+$, *m/z* = 199.2 (calculated), 199.1 (found).

^{1}H NMR (400 MHz; D_2O): δ [ppm] = 1.80-2.00 (m, 4 H), 2.58 (s, 3 H), 2.95 (t, *J*=7.09 Hz, 2 H), 3.96 (t, *J*=5.87 Hz, 1 H).

^{13}C NMR (101 MHz; D_2O): δ [ppm] = 9.38 (s), 21.72 (s), 23.88 (s), 28.96 (s), 52.85 (s), 152.52 (s), 154.96 (s), 172.33 (s).

Synthesis of H-L-Asu(3-methyl-1,2,4-triazole)-OH (27)

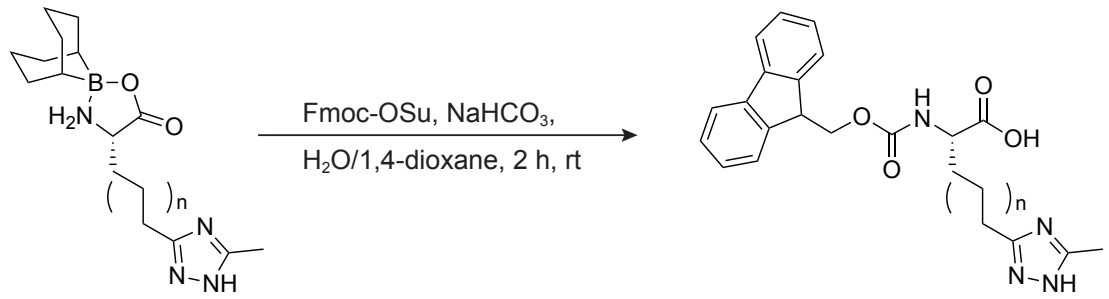
H-L-Asu(3-methyl-1,2,4-triazole)-OH (**25**) (349.3 mg, 1 mmol) was deprotected with TFA (20 mL) at 50 °C for 1 h under stirring before TFA was removed under reduced pressure. The brown residue was dissolved in water (10 mL) and washed with ethyl acetate (4 × 10 mL). The aqueous phase was lyophilized and H-L-Asu(3-methyl-1,2,4-triazole)-OH · TFA (L-AsuTri) (**27**) was obtained as a white crystalline solid (325 mg, 949 µmol, 95 %).

ESI-MS: $[M+H]^+$, *m/z* = 227.1 (calculated), 227.2 (found).

^{1}H NMR (400 MHz; D_2O): δ [ppm] = 1.31-1.47 (m, 4 H), 1.74 (quin, *J*=7.37 Hz, 2 H), 1.79 - 1.94 (m, 2 H), 2.55 (s, 3 H), 2.90 (t, *J*=7.58 Hz, 2 H), 3.91 (t, *J*=6.24 Hz, 1 H).

^{13}C NMR (101 MHz; D_2O): δ [ppm] = 9.50, 23.58, 23.99, 25.50, 27.34, 29.59, 53.31, 152.47, 155.75, 172.92.

Synthesis of Fmoc-L-AadTri-OH (28) and Fmoc-L-AsuTri-OH (29)



H-L-Aad(3-methyl-1,2,4-triazole)-OH (n=1) **26**
H-L-Asu(3-methyl-1,2,4-triazole)-OH (n=3) **27**

Fmoc-L-Aad(3-methyl-1,2,4-triazole)-OH (n=1) **28**
Fmoc-L-Asu(3-methyl-1,2,4-triazole)-OH (n=3) **29**

Synthesis of Fmoc-L-Aad(3-methyl-1,2,4-triazole)-OH (28)

H-L-Aad(3-methyl-1,2,4-triazole)-OH · TFA (**26**) (205.2 mg, 657 µmol, 1 eq) was dissolved in H_2O and the pH value was adjusted to 8-9 with $NaHCO_3$ (1 M in H_2O) receiving a final volume of 10 mL aqueous solution. Fmoc-OSu (243.8 mg, 722.7 µmol, 1.1 eq) was dissolved in 1,4-dioxane (10 mL) and added to the aqueous solution. The reaction mixture was stirred for 2 h at room temperature before the solvent was reduced under vacuum and the residue lyophilized. The crude product was purified by RP-HPLC (40-70 % B, 40 min). Fmoc-L-Aad(3-methyl-1,2,4-triazole)-OH (Fmoc-L-AadTri) (**28**) was obtained as a white solid (116 mg, 277 µmol, 42 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.35 min.

ESI-MS: $[M+H]^+$, *m/z* = 421.2 (calculated), 421.4 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 1.59-1.85 (m, 4 H), 2.43 (s, 3 H), 2.71-2.80 (m, 2 H), 3.92-4.02 (m, 1 H), 4.18-4.26 (m, 1 H), 4.26-4.34 (m, 2 H), 7.28-7.36 (m, 2 H), 7.41 (t, *J*=7.34 Hz, 2 H), 7.67 (d, *J*=8.07 Hz, 1 H), 7.72 (d, *J*=7.46 Hz, 2 H), 7.89 (d, *J*=7.58 Hz, 2 H).

^{13}C NMR (101 MHz; DMSO-d_6): δ [ppm] = 11.02, 23.56, 24.81, 30.00, 46.66, 53.47, 65.63, 120.12, 125.28, 127.08, 127.65, 140.73, 143.80, 143.84, 153.27, 156.19, 156.64, 173.71.

Synthesis of Fmoc-L-Asu(3-methyl-1,2,4-triazole)-OH (29)

H-L-Asu(3-methyl-1,2,4-triazole)-OH · TFA (**27**) (239.6 mg, 700 µmol, 1 eq) was dissolved in H_2O and the pH value was adjusted to 8-9 with $NaHCO_3$ (1 M in H_2O) receiving a final volume of 10 mL aqueous solution. Fmoc-OSu (259.7 mg, 770 µmol, 1.1 eq) was dissolved in 1,4-dioxane (10 mL) and added to the aqueous solution. The reaction mixture was stirred for 2 h at room temperature before the solvent was reduced under vacuum and the residue lyophilized. The crude product was purified by RP-HPLC (40-70 % B, 40 min). Fmoc-L-Asu(3-methyl-1,2,4-triazole)-OH (Fmoc-L-AsuTri) (**29**) was obtained as a white solid (157 mg, 350 µmol, 50 %).

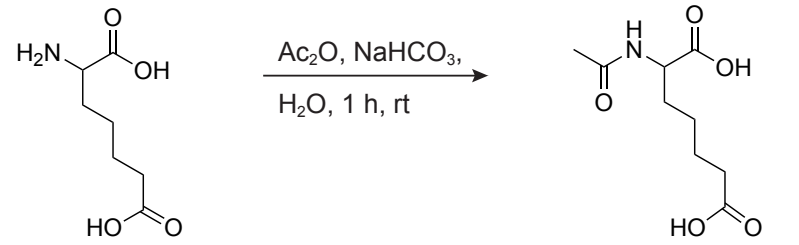
HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.78 min.

ESI-MS: $[M+H]^+$, m/z = 449.2 (calculated), 449.1 (found).

^{1}H NMR (400 MHz; DMSO-d_6): δ [ppm] = 1.24-1.41 (m, 4 H), 1.55-1.75 (m, 4 H), 2.45 (s, 3 H), 2.77 (t, J=7.58 Hz, 2 H), 3.89-3.97 (m, 1 H), 4.19-4.25 (m, 1 H), 4.26-4.32 (m, 2 H), 7.32 (t, J=7.46 Hz, 2 H), 7.41 (t, J=7.34 Hz, 2 H), 7.63 (d, J=8.07 Hz, 1 H), 7.70-7.75 (m, 2 H), 7.89 (d, J=7.46 Hz, 2 H).

^{13}C NMR (101 MHz; DMSO-d_6): δ [ppm] = 10.91, 24.90, 25.11, 26.42, 27.79, 30.53, 46.68, 53.71, 65.58, 120.12, 125.28, 127.07, 127.65, 140.73, 140.74, 143.80, 143.87, 153.07, 156.19, 156.52, 173.97.

Synthesis of Ac-DL-Apm-OH (30)



H-DL-Apm-OH **1** Ac-DL-Apm-OH **30**

H-DL-Apm-OH (**1**) (2.102 g, 12 mmol, 1 eq) was dissolved in water (20 mL) and the pH value was adjusted to 8-9 with $NaHCO_3$ (1 M in water). Acetic anhydride (263.2 µL, 2.784 mmol, 2 eq) was added dropwise to the solution and the mixture was stirred for 1 h at room temperature. The solution was lyophilized and the crude product was purified by RP-HPLC. Ac-DL-Apm-OH (**30**) was obtained as colorless amorphous solid (2.465 g, 11.35 mmol, 92 %).

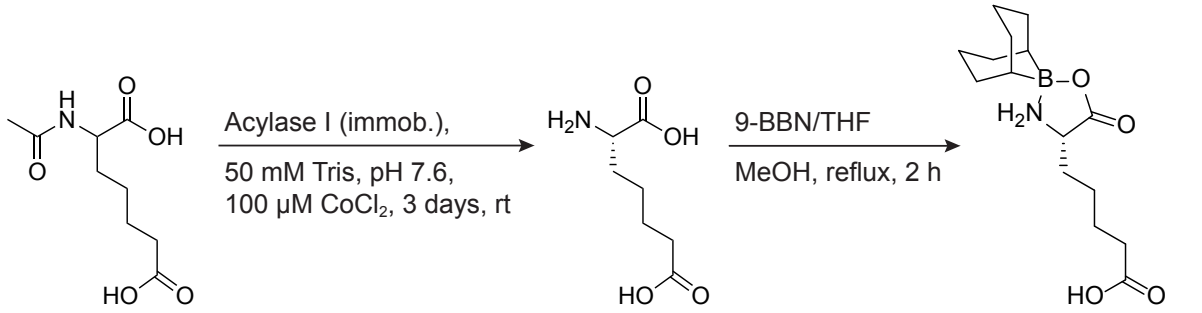
HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 5.02 min.

ESI-MS: $[M+H]^+$, m/z = 218.1 (calculated), 218.3 (found); $[M+Na]^+$, m/z = 240.1 (calculated), 240.2 (found); $[2M+H]^+$, m/z = 435.2 (calculated), 435.2 (found); $[2M+Na]^+$, m/z = 457.2 (calculated), 457.2 (found).

^{1}H NMR (300 MHz; D_2O): δ [ppm] = 1.29-1.42 (m, 2 H), 1.49-1.63 (m, 2 H), 1.64-1.90 (m, 2 H), 1.97 (s, 3 H), 2.33 (t, J=7.34 Hz, 2 H), 4.22-4.30 (m, 1 H).

^{13}C NMR (75 MHz; D_2O): δ [ppm] = 21.49, 23.63, 24.37, 30.02, 33.40, 52.66, 174.19, 175.95, 178.63.

Synthesis of 9-BBN-L-Apm-OH (32)



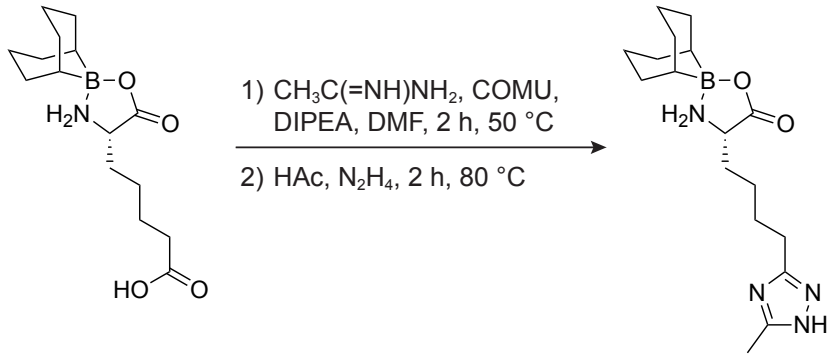
Ac-DL-Apm-OH **30** H-L-Apm-OH **31** 9-BBN-L-Apm-OH **32**

Ac-DL-Apm-OH (**30**) (1.738 g, 8 mmol, 1 eq) was dissolved in acylase I buffer (50 mM Tris-HCl, 100 µM $CoCl_2$, pH 7.6) and the pH value was adjusted to pH 7.6 by adding NaOH. The reaction mixture was incubated with acylase I (from Aspergillus sp., immobilized on Eupergit® C, Sigma Alrich) (20 U, 300 mg) for 3 days at room temperature. The solution containing H-L-Apm-OH (**31**) was lyophilized and the receiving colorless oil was directly used for 9-BBN-protection without further purification.
The crude product of (**31**) was dissolved in anhydrous MeOH (30 mL) and heated at reflux under argon atmosphere. 9-BBN (0.5 M in THF, 8 mL, 4 mmol, 1 eq with respect to 50 % of **30**) was added dropwise. Reaction mixture was stirred under reflux for 2 h. The solvent was removed under reduced pressure and the residue lyophilized. The crude product was purified by preparative RP-HPLC (50-95 % B, 40 min). 9-BBN-L-Apm (**32**) was obtained as a white powder (911 µg, 3.086 mmol, 77 % with respect to 50 % of **30**).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 12.12 min.

ESI-MS: $[M+H]^+$, m/z = 296.2 (calculated), 296.1 (found); $[2M\text{-}9BBN+H]^+$, m/z = 471.3 (calculated), 471.3 (found); $[2M+H]^+$, m/z = 591.4 (calculated), 591.5 (found); $[2M+Na]^+$, m/z = 613.4 (calculated), 613.2 (found).

Synthesis of 9-BBN-L-Apm(3-methyl-1,2,4-triazole) (33)



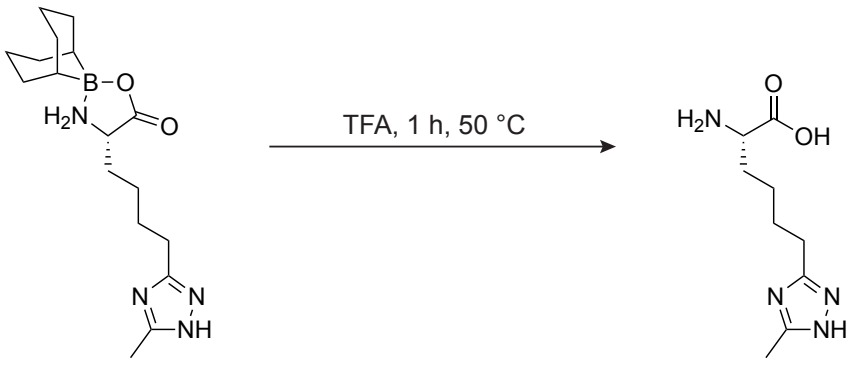
9-BBN-L-Apm-OH **32** 9-BBN-L-Apm(3-methyl-1,2,4-triazole) **33**

9-BBN-L-Apm (**32**) (885.6 mg, 3 mmol, 1 eq) was dissolved in DMF (10 mL) and heated at 50 °C. COMU (1.413 g, 3.3 mmol, 1.1 eq) and acetamidine hydrochloride (1.702 g, 18 mmol, 6 eq) were each dissolved in DMF (20 mL) separately. The COMU solution and DIPEA (3.061 mL, 18 mmol, 6 eq) were added to the heated amino acid solution and stirred for 5 min, before the acetamidine solution was added. The reaction took place for 2 h at 50 °C. The triazole was synthesized by adding acetic acid (1.716 mL, 30 mmol, 10 eq) followed by $N_2H_4 \cdot H_2O$ (218.24 µL, 4.5 mmol, 1.5 eq) and stirring of the reaction solution for additional 2 h at 80 °C. DMF was removed under reduced pressure and the residue was lyophilized to yield the crude product as a reddish oil. Purification was performed by RP-HPLC and 9-BBN-L-Apm(3-methyl-1,2,4-triazole) (9-BBN-L-ApmTri) (**33**) was obtained as a white solid (565 mg, 1.7 mmol, 57 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 9.03 min.

ESI-MS: $[M+H]^+$, m/z = 333.2 (calculated), 333.2 (found).

Synthesis of H-L-Apm(3-methyl-1,2,4-triazole)-OH (16)



9-BBN-L-Apm(3-methyl-1,2,4-triazole)-OH **33** H-L-Apm(3-methyl-1,2,4-triazole)-OH **16**

9-BBN-L-Apm(3-methyl-1,2,4-triazole) (**33**) (318.23 mg, 1.7 mmol) was deprotected with TFA (20 mL) at 50 °C for 1 h while stirring before TFA was removed under reduced pressure. The resulting brown residue was dissolved in water (10 mL) and washed with ethyl acetate (4 × 10 mL). The aqueous phase was lyophilized and H-L-Apm(3-methyl-1,2,4-triazole)-OH · TFA (**16**) was obtained as a white crystalline solid (516 mg, 1.581 mmol, 93 %).

ESI-MS: $[M+H]^+$, m/z = 213.1 (calculated), 213.1 (found).

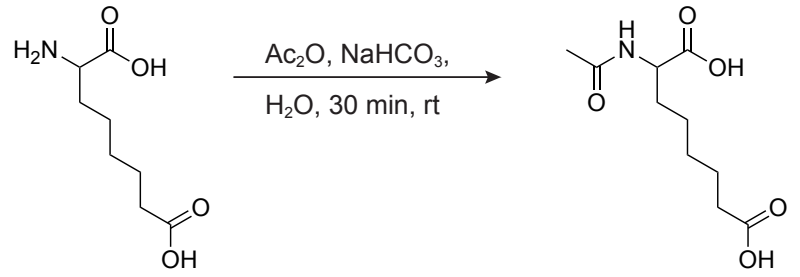
^{1}H NMR (300 MHz; D_2O): δ [ppm] = 1.34-1.54 (m, 2 H), 1.76 (s, 2 H), 1.84-2.02 (m, 2 H), 2.54 (s, 3 H), 2.86-2.96 (m, 2 H), 3.96 (t, J=6.24 Hz, 1 H).

^{13}C NMR (75 MHz; D_2O): δ [ppm] = 7.20, 21.18, 21.61, 23.11, 26.95, 50.57, 150.22, 153.20, 170.01.

4.2.2.2 Synthesis of hydroxamate building blocks

The amino acid H-Asu(NH-OH)-OH (**37**) and the building block Fmoc-L-Asu(NH-OTrt)-OH (**40**) were synthesized by *Annika Kühn* [235] according to the protocol of *Julian Seidel* [236, 306].

Synthesis of Ac-DL-Asu-OH (35)



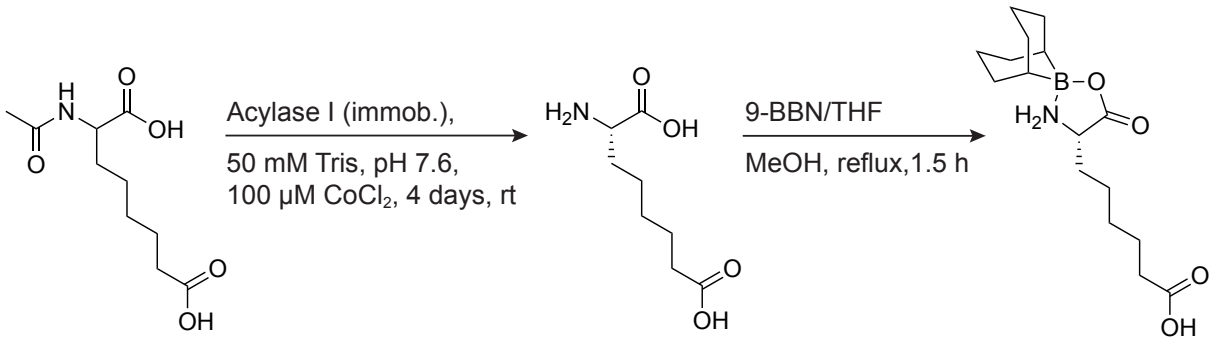
H-DL-Asu-OH **34** Ac-DL-Asu-OH **35**

H-DL-Asu-OH (**34**) (3.78 g, 20 mmol, 1 eq) was dissolved in 40 mL of a $NaHCO_3$ solution (1 M in water). Acetic anhydride (3.78 mL, 40 mmol, 2 eq) was added dropwise to the aqueous solution and the mixture was stirred for 30 min at room temperature. The solution was lyophilized and the crude product was purified by RP-HPLC. Product-containing fractions were combined and neutralized by adding 5 mL of a NaOH solution (6 M in water). After lyophilization, the product Ac-DL-Asu-OH (**35**) was obtained as colorless oil.

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 5.94 min.

ESI-MS: $[M+H]^+$, m/z = 232.1 (calculated), 232.2 (found); $[M+Na]^+$, m/z = 254.1 (calculated), 254.1 (found).

Synthesis of 9-BBN-L-Asu (23)



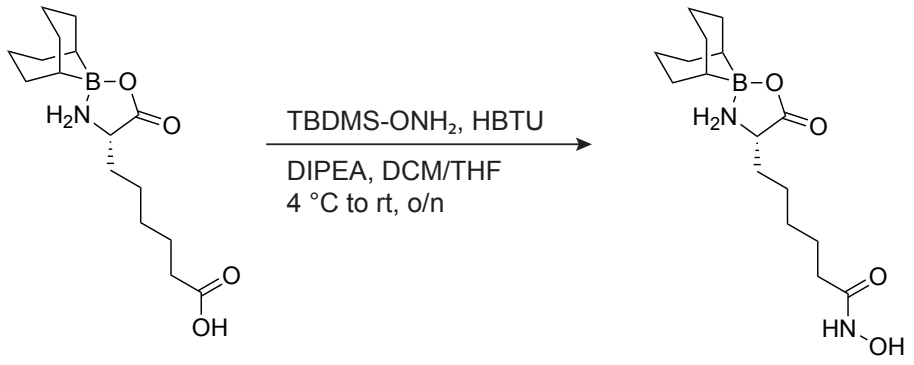
Ac-DL-Asu-OH **35** H-L-Asu-OH **21** 9-BBN-L-Asu **23**

Ac-DL-Asu-OH (**35**) was dissolved in acylase I buffer (50 mM Tris-HCl, 100 µM $CoCl_2$, pH 7.6) and the pH value was adjusted to pH 7.6 by adding NaOH. The reaction mixture was incubated with acylase I (from Aspergillus sp., immobilized on Eupergit® C, Sigma Aldrich) (25 U, 500 mg) for 4 days at room temperature. The solution containing H-L-Asu-OH (**21**) was lyophilized and the receiving colorless oil was directly used for 9-BBN-protection without further purification.
The crude product of (**21**) was dissolved in anhydrous MeOH (60 mL) and heated at reflux under argon atmosphere. 9-BBN (0.5 M in THF, 20 mL, 10 mmol, 1 eq with respect to 50 % of **34**) was added dropwise and the reaction mixture was stirred under reflux for 1.5 h. The solvent was removed under reduced pressure and the crude product was purified by preparative RP-HPLC (50-80 % B, 40 min). 9-BBN-L-Asu (**23**) was obtained as a white crystalline solid. (2.63 g, 8.51 mmol, 85 % with respect to 50 % of **34**).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 11.30 min.

ESI-MS: $[M+H]^+$, m/z = 310.2 (calculated), 310.0 (found); $[2M\text{-}9BBN+H]^+$, m/z = 499.3 (calculated), 499.1 (found); $[2M+H]^+$, m/z = 619.4 (calculated), 619.2 (found).

Synthesis of 9-BBN-L-Asu(NH-OH) (36)



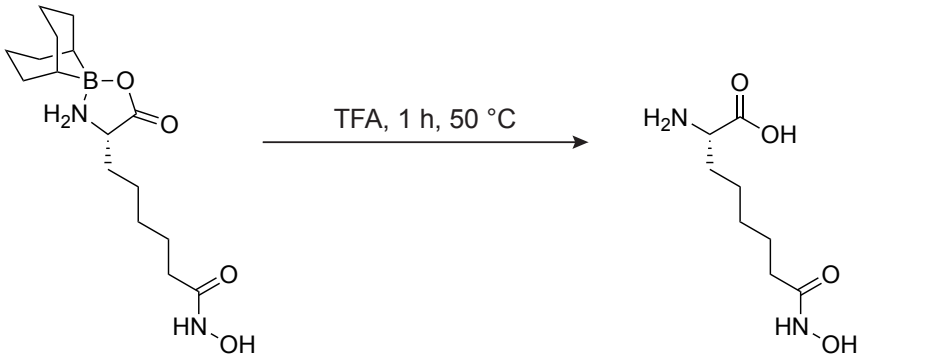
9-BBN-L-Asu **23** 9-BBN-L-Asu(NH-OH) **36**

9-BBN-L-Asu (**23**) (1.26 g, 4.08 mmol, 1 eq) was dissolved in 41 mL of anhydrous DCM/THF (1:1) under argon atmosphere and the solution was cooled in an ice bath. HBTU (1.86 g, 4.90 mmol, 1.2 eq) was added and the solution was stirred for 15 min. O-(tert-butyldimethylsilyl)hydroxylamine (TBDMS-ONH_2) (0.78 g, 5.30 mmol, 1.3 eq) and DIPEA (1.42 mL, 8.16 mmol, 2 eq) were added and the ice bath was removed. The reaction mixture was stirred over night at room temperature. After adding 20 mL of water, the reaction mixture was extracted with DCM (2 × 20 mL). The organic phases were combined and the solvent removed under reduced pressure. The crude product was purified by preparative HPLC (50-80 % B, 40 min). 9-BBN-L-Asu(NH-OH) (**36**) was obtained as a white powder (1.109 g, 3.42 mmol, 84 %).

HPLC (5-95 % B in 12.75 min, 0.2 mL/min): $t_{ret.}$ = 11.29 min.

ESI-MS: $[M+H]^+$, m/z = 325.2 (calculated), 324.9 (found); $[2M+H]^+$, m/z = 649.4 (calculated), 649.2 (found).

Synthesis of H-L-Asu(NH-OH)-OH (37)



9-BBN-L-Asu(NH-OH) **36** H-L-Asu(NH-OH)-OH **37**

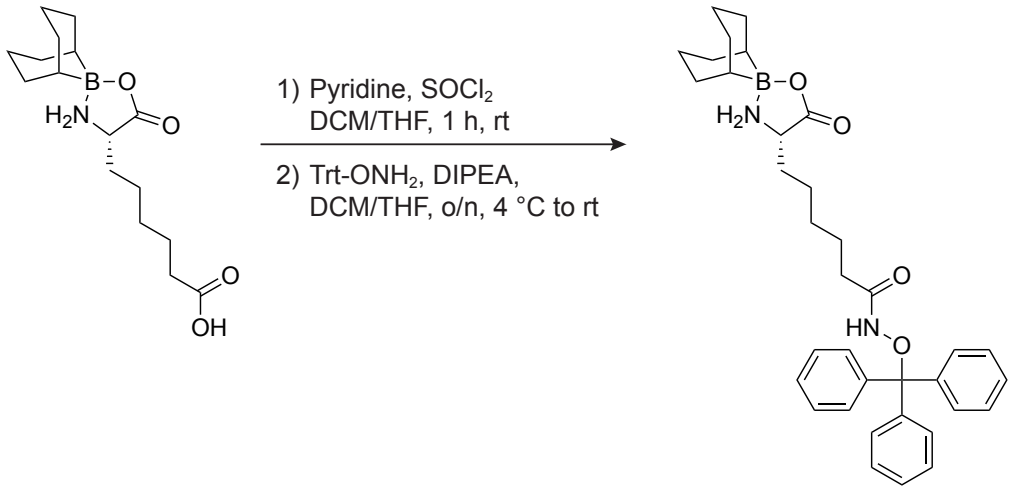
9-BBN-L-Asu(NH-OH) (**36**) (1.109 g, 3.42 mmol, 1 eq) was deprotected with TFA (40 mL) at 50 °C for 1 h while stirring before TFA was removed under reduced pressure. The brown residue was dissolved in water (40 mL) and washed with ethyl acetate (2 × 20 mL). The aqueous phase was lyophilized and H-L-Asu(NH-OH)-OH · TFA (L-AsuHd) (**37**) was obtained as a grey powder (973 mg, 3.06 mmol, 89 %).

ESI-MS: $[M+H]^+$, m/z = 205.1 (calculated), 205.2 (found).

^{1}H NMR (300 MHz; D_2O): δ [ppm] = 1.25-1.43 (m, 4 H), 1.48-1.60 (m, 2 H), 1.77-1.94 (m, 2 H), 2.11 (t, J=7.34 Hz, 2 H), 3.96 (t, J=6.24 Hz, 1 H).

^{13}C NMR (75 MHz; D_2O): δ [ppm] = 23.7, 24.42, 27.46, 29.53, 32.04, 53.02, 172.54, 173.22.

Synthesis of 9-BBN-L-Asu(NH-OTrt) (38)



9-BBN-L-Asu **23** 9-BBN-L-Asu(NH-OTrt) **38**

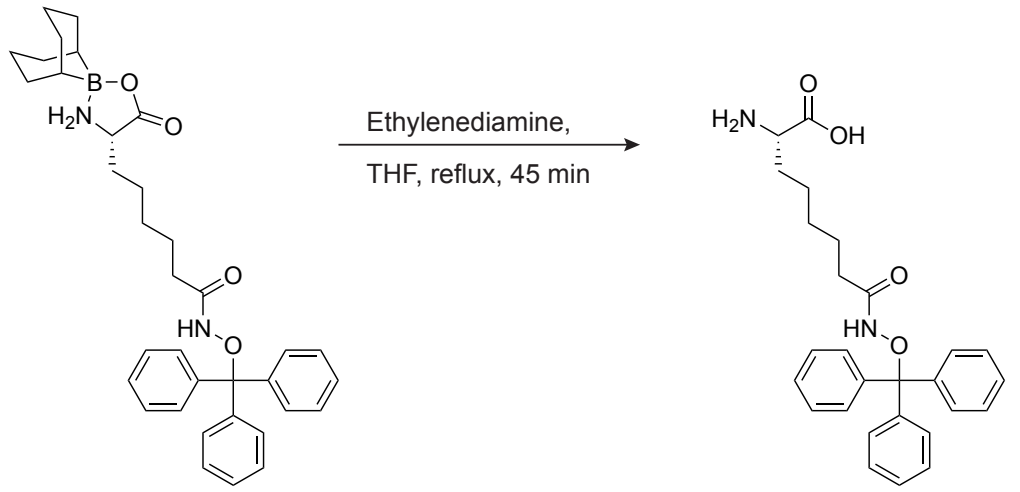
9-BBN-L-Asu (**23**) (2.51 g, 4.43 mmol, 1 eq) was dissolved in 90 mL of anhydrous DCM/THF (1:1) under argon atmosphere. Pyridine (0.36 mL, 4.43 mmol, 1 eq) and thionyl chloride

(0.32 mL, 4.43 mmol, 1 eq) were added and the solution was stirred for 1 h at room temperature. After cooling the reaction mixture on ice, a solution of O-tritylhydroxylamine (Trt-ONH_2) (1.22 g, 4.43 mmol, 1 eq) and DIPEA (2.31 mL, 13.29 mmol, 3 eq) in 22 mL of anhydrous DCM/THF (1:1) was added dropwise. The ice bath was removed and the reaction mixture was stirred over night at room temperature. The solvent was removed under reduced pressure and the crude product of 9-BBN-L-Asu(NH-OTrt) (**38**) was used in the following reaction without further purification.

HPLC (40-99 % B in 12.00 min, 0.2 mL/min): $t_{ret.}$ = 14.90 min.

ESI-MS: $[M+Na]^+$, m/z = 589.3 (calculated), 589.1 (found).

Synthesis of H-L-Asu(NH-OTrt)-OH (39)



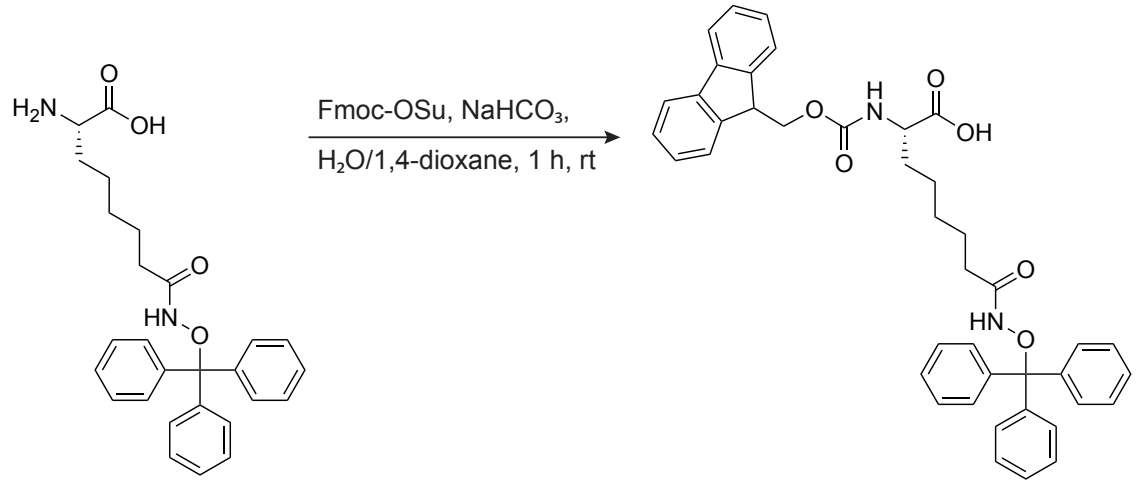
9-BBN-L-Asu(NH-OTrt) **38** H-L-Asu(NH-OTrt)-OH **39**

The crude product of 9-BBN-L-Asu(NH-OTrt) (**38**) (4.43 mmol, 1 eq) was dissolved in THF (45 mL) and ethylenediamine (1.48 mL, 19.80 mmol, 5 eq) was added. The reaction mixture was heated to reflux for 45 min. The precipitated product was collected by centrifugation (4700 × g, 20 min), suspended in H_2O (5 mL) and lyophilized. After dissolving of the residue in HPLC solvent, the solution was acidified with HCl (1 M in H_2O) to pH 5 and purified by preparative RP-HPLC (40-95 % B, 40 min). The product-containing fractions were combined, neutralized with NaOH (6 M in H_2O) and lyophilized over night. An impure product H-L-Asu(NH-OTrt)-OH (**39**) was obtained as a white powder (2.271 g, 5.09 mmol) and completely used in the following reaction.

HPLC (40-99 % B in 12.00 min, 0.2 mL/min): $t_{ret.}$ = 9.66 min.

ESI-MS: $[M+Na]^+$, m/z = 569.2 (calculated), 469.1 (found); $[M+H]^+$, m/z = 893.4 (calculated), 893.5 (found).

Synthesis of Fmoc-L-Asu(NH-OTrt)-OH (40)



H-L-Asu(NH-OTrt)-OH **39**

Fmoc-L-Asu(NH-OTrt)-OH **40**

H-L-Asu(NH-OTrt)-OH (**39**) (4.43 mmol, 1 eq) and $NaHCO_3$ (1.12 g, 13.29 mmol, 3 eq) were dissolved in 45 mL of 1,4-dioxane/H_2O (1:1). Fmoc-OSu (0.60 g, 1.77 mmol, 0.4 eq) was dissolved in 5 mL of 1,4-dioxane and added to the amino acid solution. The reaction was monitored by LC-MS analysis and additional Fmoc-OSu was added in small portions (0.15 g, 0.44 mmol, 0.1 eq). After full conversion, the reaction mixture was extracted with ethyl acetate (3 $\times$ 50 mL) and the combined organic phases were dried over Na_2SO_4. The solvent was removed under reduced pressure and the product Fmoc-L-Asu(NH-OTrt)-OH (**40**) was obtained as a white powder (1.70 g, 2.54 mmol, 57 % with respect to **23**).

HPLC (40-99 % B in 12 min, 0.2 mL/min): $t_{ret.}$ = 14.50 min.

ESI-MS: $[M+Na]^+$, m/z = 691.2 (calculated), 691.1 (found).

4.2.3 Solid-phase peptide synthesis

Peptides were synthesized via solid-phase peptide synthesis (SPPS) [307] by using the Fmoc strategy. TentaGel® HL RAM resins (loading capacity: 0.32-0.38 mmol/g) were used as solid support. Peptide synthesis was performed in scales of 25 µmol or 50 µmol by using the following standard amino acid building blocks: Fmoc-L-Ala-OH, Fmoc-L-Arg(Pbf)-OH, Fmoc-L-Asn(Trt)-OH, Fmoc-L-Asp(OtBu)-OH, Fmoc-L-Cys(Trt)-OH, Fmoc-L-Gln(Trt)-OH, Fmoc-L-Glu(OtBu)-OH, Fmoc-Gly-OH, Fmoc-L-His(Trt)-OH, Fmoc-L-Ile-OH, Fmoc-L-Leu-OH, Fmoc-L-Lys(Boc)-OH, Fmoc-L-Phe-OH, Fmoc-L-Pro-OH, Fmoc-L-Ser(tBu)-OH, Fmoc-L-Thr(OtBu)-OH, Fmoc-L-Trp(Boc)-OH, Fmoc-L-Tyr(tBu)-OH, Fmoc-L-Val-OH.

4.2.3.1 Automated peptide synthesis

Automated SPPS was performed on a MultiSynTech Syro I peptide synthesizer (Biotage) using the following double coupling method: The first coupling of each amino acid (3 eq relative to the resin) was performed under activation with *N, N'*-diisopropylcarbodiimid (DIC) (3 eq) and ethyl cyanohydroxyiminoacetate (Oxyma Pure) (3 eq) in DMF for 40 min. Each coupling reaction was repeated with the amino acid (3 eq) activated with HBTU (2.7 eq) or 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) (2.7 eq) and *N*-methylmorpholine (NMM) (41 mM) as base in DMF for 30 min. Fmoc deprotection was accomplished by treating the resin twice with 40 % piperidine in DMF for 3 min each. Coupling reaction and Fmoc deprotection were followed by washing the resin with DMF (6 $\times$ 200 µL).

4.2.3.2 Manual peptide synthesis

Manual coupling of amino acids

Commercially available Fmoc-L-Lys(ivDde)-OH (43.1 mg, 75 µmol, 3 eq), Fmoc-L-Lys(Mtt)-OH (46.9 mg, 75 µmol, 3 eq) and Fmoc-Lys(biotin)-OH (44.6 mg, 75 µmol, 3 eq) (all Iris Biotech) were coupled manually with HATU (25.7 mg, 67.5 µmol, 2.7 eq) in 1.5 mL of 400 mM NMM in DMF for 1 h at room temperature under shaking. Afterwards, the resin was washed with DMF (4 $\times$ 2 min) and DCM (4 $\times$ 1 min).
Synthesized triazole-containing amino acid building blocks **(11-14, 17, 19, 28, 29)** and Fmoc-L-Asu(NH-OTrt)-OH (**40**) (66.9 mg, 100 µmol, 2 eq) were coupled with 2-fold molar excess of the amino acid and equimolar amounts of [ethyl cyano(hydroxyimino)acetato-O^2]tri-1-pyrrolidinylphosphonium hexafluorophosphate (PyOxim) in 1.5 mL of 400 mM NMM in DMF for 1 h at room temperature under shaking. Subsequently, the resin was washed with DMF (4 $\times$ 2 min) and DCM (4 $\times$ 1 min).
Coupling of the pseudo prolines Fmoc-Ala-Thr(psiMe,Mepro)-OH (AT) (33.9 mg, 75 µmol, 3 eq) Fmoc-Gln(Trt)-Thr(psiMe,Mepro)-OH (QT) (56.4 mg, 75 µmol, 3 eq) and Fmoc-Ser(tBu)-Thr(psiMe,Mepro)-OH (ST) (39.3 mg, 75 µmol, 3 eq) (all Iris Biotech) was carried out manually with HATU (25.7 mg, 67.5 µmol, 2.7 eq) in 1 mL of 400 mM NMM in DMF for 1 h at room temperature under shaking. Afterwards, the resin was washed with DMF (4 $\times$ 2 min) and DCM (4 $\times$ 1 min).
The Fmoc group was removed manually with 20 % piperidine in DMF (3 $\times$ 5 min). Afterwards, the resin was washed with DMF (3 $\times$ 2 min) and DCM (3 $\times$ 2 min).

Synthesis of Nε-acyl lysine modified peptides

Lysine Nεacyl modifications were installed by incorporation of Fmoc-Lys(Mtt)-OH at position H4K20. After N-terminal acetylation of the peptide lysine residue 20 was orthogonally deprotected by flow-washing the resin with 1 % TFA in DCM.
Nεlysine formylation was achieved by using an acetic formic mixed anhydride [308]. Therefore, formic acid (160 µL, 400 µmol, 1 eq) was added dropwise to ice-cold acetic

anhydride (160 µL, 212 µmol, 0.53 eq) and the solution was stirred on ice for 10 min. Afterwards the mixture was heated up to 50 °C for 5 min under shaking before it was cooled down on ice. The mixed anhydride and anhydrous DCM (2 mL) were added to the resin under argon atmosphere and the resin was shaken for 2 h. Afterwards, the resin was washed with DMF (3 × 2 min) and DCM (3 × 2 min).
Propionylation was obtained by incubation of the resin with a solution of 6 % lutidin and 5 % propionic anhydride in DMF (2 mL) for 20 min twice. Afterwards, the resin was washed with DMF (3 × 2 min) and DCM (3 × 2 min).
Benzoic acid (15.27 mg, 125 µmol, 5 eq) was coupled with PyOxim (54.31 mg, 125 µmol, 5 eq) in NMM (400 mM in DMF) for 1 h at room temperature. After coupling, the resin was washed with DMF (4 × 2 min) and DCM (4 × 2 min).

Installation of the Boc protection group on probe H4K20Boc

The H4K20 probe was synthesized with incorporated Lys(ivDde) at position K16 and cleaved off the resin (see 4.2.3.3). After purification via preparative HPLC, the purified peptide (9 mg, 4.46 µmol, 1 eq) was dissolved in 100 µL DMF. Boc_2O (1.07 mg, 4.9 µmol, 1.1 eq) and DIPEA (26.76 µmol, 4.55 µL, 6 eq) were added and the solution was incubated for 1 h under shaking at room temperature. The 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)isovaleryl (ivDde) protecting group was removed by adding 400 µL of a solution of $NH_2OH \cdot HCl$ (10 mg, 144 µmol) and imidazole (7.4 mg, 108 µmol) and incubating overnight under shaking at room temperature [241]. The reaction solution was diluted with HPLC-A (4 mL) and peptide was purified by semi-preparative HPLC (1.3 mg, 0.68 µmol, 14.4 %).

Ester bond formation for depsipeptide synthesis

The formation of an ester bond between threonine and glycine was achieved by the Steglich reaction [309]. TentaGel HL RAM resin was deprotected with 20 % piperidine in DMF (3 × 1 mL) for 5 min and washed with DMF (3 × 2 min) and DCM (3 × 2 min). Glycolic acid (10 eq) was coupled with HATU (9 eq) in NMM (400 mM in DMF) for 15 min twice. In a round bottom flask equipped with a $CaCl_2$ drying tube, Fmoc-Thr(tBu)-OH (20 eq) was dissolved in anhydrous DCM and cooled down to 0 °C, before DIC (10 eq) was added. After 20 min under stirring at room temperature, the solvent was removed under reduced pressure. The residue was solved in DMF and half of the solution was added to the resin with catalytic amounts of 4-dimethylaminopyridine (DMAP). After 30 min the reaction was repeated. Afterwards the resin was washed with DMF (3 × 2 min) and DCM (3 × 2 min).

Synthesis of the biotin-linker peptide

The peptide Fmoc-GR-Ahx-K-Ahx-RRK(biotin)-NH_2 was cleaved off the resin (see 4.2.3.3) and purified by preparative HPLC. The purified peptide (5.0 mg, 3.39 µmol, 1 eq) and amine-reactive Cyanine5 NHS ester (2.49 mg, 3.73 µmol, 1.1 eq) (Abcam) were dissolved

in 50 µL DMF separately. After mixing both solutions, 400 µL of a $NaHCO_3$ solution (1 M in water) were added and the reaction took place for 1 h under shaking at room temperature. The reaction mixture was lyophilized and the residue dissolved in DMF (80 µL). Fmoc deprotection was performed by adding 20 µL piperidine (to a final concentration of 20 %) and the reaction was stopped after 10 s by adding TFA (100 µL). After lyophilization, the peptide was purified by semi-preparative HPLC (3.1 mg, 2.48 µmol, 73.2 %).

N-terminal acetylation of peptides

Peptides for pull-down assays were acetylated on the N-terminus with a solution of 5 % acetic anhydride and 6 % lutidine in DMF for 20 min at room temperature under shaking. The acetylation was repeated before the resin was washed with DMF and DCM.

4.2.3.3 Peptide cleavage

Deprotection of amino acid side chains and cleavage from the resin was achieved by treating the resin with 7 mL cleavage cocktail (90 % TFA, 5 % water, 5 % TIPS, 67 mg/mL phenol) for 3 h under shaking at room temperature. The incubation was repeated twice with 1.5 mL cleavage cocktail for 30 min. The resin was washed with DCM (2 $\times$ 5 mL) and the combined cleavage solutions were concentrated under reduced pressure to a final volume of 1-2 mL. Peptides were precipitated in 40 mL pre-cooled diethyl ether (-80 °C) and the suspension was centrifuged (4000 $\times$ g, -10 °C, 10 min). The pellet was washed with 20 mL pre-cooled diethyl ether (-80 °C) and the centrifugation was repeated. After removing the supernatant, the pellet was dissolved in water/ACN and lyophilized over night. All synthesized peptides were purified by preparative HPLC (4.2.1.2).

4.2.3.4 Validation and immobilization of peptide probes

Purified peptides were dissolved in 30 % ACN in water to yield 10 mM or 100 mM stock solutions. The stock solutions were quantified by analytical HPLC where equal amounts of each peptide were injected and the peak area A_p for each peptide was determined by integration of the absorption peak at 218 nm. To calculate the used stock solution volume V_{exact} for immobilization, the volume V was adjusted according to

$$V_{exact} = V \cdot f_p \quad \text{with} \quad f_p = \frac{\frac{1}{n}\sum_{p=1}^{n} A_p}{A_p}$$

where the correction factors f_p were calculated for each peptide p to adjust the amounts of n peptides based on the peak areas A_p. The volumes V_{exact} were used for the dilution of the peptide stock solutions to a concentration of 1 mM in coupling buffer (50 mM Tris-HCl, 5 mM EDTA-Na, pH 8.5). The peptide probes were coupled covalently to SulfoLink™ Coupling Resin (Thermo Scientific). The principle is shown in figure figure 44.

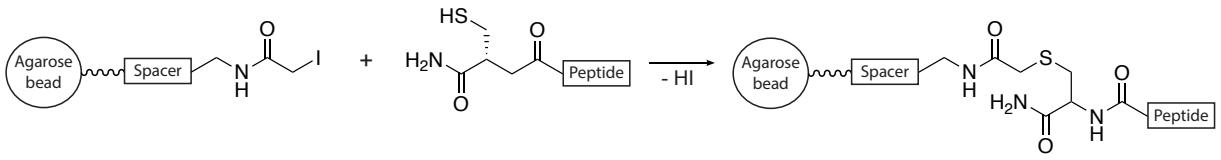


Figure 44: Immobilization of peptide probes on SulfoLink™ agarose matrix.

For each peptide probe 300 µL SulfoLink™ suspension (supplied as a 50 % slurry) were used and the beads were equilibrated (5 × 800 µL) with coupling buffer. The resin was incubated for 15 min with 300 µL of the prepared 1 mM peptide solutions while shaking and for additional 45 min without shaking at room temperature. After washing (3 × 500 µL) with coupling buffer, the matrix was blocked by incubation with blocking buffer (50 mM β-mercaptoethanol in coupling buffer) for 15 min while shaking and for additional 45 min without shaking at room temperature. Afterwards, the resin was washed with 1 M NaCl (6 ×1 mL), water (2 ×1 mL) and 50 % ACN in water (4 ×1 mL). The final resin was stored in 50 % ACN in water as 20 µL aliquots (50 % slurry) at -20 °C.

4.2.4 Biochemical methods

4.2.4.1 General cultivation of bacteria

If not stated otherwise, *E. coli* cells were cultivated in LB medium supplemented with the corresponding antibiotics (ampicillin: 100 µg/mL, kanamycin: 30 µg/mL) under shaking (170 rpm) at 37 °C. Newly generated *E. coli* strains were stored as glycerol stocks at -80 °C. Glycerol stocks were prepared by adding sterile glycerol (200 µL) to 800 µL of an overnight bacteria culture (20 % glycerol, v/v).

4.2.4.2 General methods for molecular cloning

Preparation of electrocompetent bacteria

Electrocompetent cells (*E. coli* BL21 (DE3) and *E. coli* XL1-Blue) for transformations were prepared by inoculation of 700 mL 2xYT without antibiotics with 7 mL overnight culture. Cells were grown until an OD_{600} of 0.5-0.6 and then cooled down for 30 min on ice. The medium was removed by centrifugation (2000 × g, 10 min, 4 °C) and cells were resuspended in 500 ml cold sterile water. After repeating the centrifugation, the pellet was resuspended in 250 mL water, followed by centrifugation (2000 × g, 10 min, 4 °C) and resuspension of the cells in 20 % glycerol in water (v/v). The suspension was centrifuged (5000 × g, 30 min, 4 °C) and the pellet was resuspended in 2 mL of 20 % glycerol in water (v/v). Aliquots of 50 µL were frozen in liquid nitrogen and stored at -80 °C.

Transformation of electrocompetent bacteria

An aliquot (50 µL) of electrocompetent bacteria (*E. coli* BL21 (DE3) or *E. coli* XL1-Blue)

was thawed on ice and mixed with 1 ng of plasmid DNA or 5 µL of ligation solution. The mixture was transferred to an electroporation cuvette (2 mm, 450 µL) (Carl Roth) and the cells were pulsed (U = 2.5 kV, 5 ms) using a MicroPulser™ electroporator (Bio-Rad). Pre-warmed LB medium (1 mL) was added immediately and the bacteria were incubated for 30-60 min at 37 °C while shaking (400 rpm). Cells were plated on LB agar plates containing the corresponding antibiotics and incubated over night at 37 °C.

Preparation of plasmid DNA

Single colonies were picked from LB agar plates, transferred into 5 mL LB medium with the corresponding antibiotics and incubated over night. For plasmid isolation, the peqGOLD Plasmid Miniprep Kit I (Peqlab) was used according to the manufacturer's protocol. The DNA concentration was determined with a NanoDrop ND-1000 (Peqlab) at 280 nm.
Plasmids for transfection of mammalian cells were preparated in larger scale by inoculation of 2 mL LB medium supplemented with the appropriate antibiotic with bacteria from glycerol stocks. After incubation for 6-8 h, 250 mL antibiotic-containing LB medium were inoculated with 250 µL (1:1000) of the grown culture and incubated over night. Plasmid DNA was isolated using the GeneJET Plasmid Maxiprep Kit (Thermo Fisher Scientific) according to the manufacturer's protocol and DNA concentration was determined with a NanoDrop ND-1000 (Peqlab) at 260 nm. The obtained plasmid DNA solutions were aliquoted and stored at -20 °C.

Agarose gel electrophoresis

Separation of restriction digest and PCR products was carried out by agarose gel electrophoresis using 0.8-1.2 % agarose gels in TAE buffer (40 mM Tris-HCl, 20 mM acetic acid, 1 mM EDTA, pH 8.5) and Midori Green Advance (Biozym) for DNA detection. DNA samples were prepared by mixing with 6x DNA Loading Dye (Thermo Fisher Scientific) and loaded together with 10 µL GeneRuler 1 kB DNA Ladder (Thermo Fisher Scientific) as size marker on the gel. Electrophoresis was accomplished at 80 mA for 60-70 min. Agarose gel images were taken on a ChemiDoc MP imaging system (Bio-Rad). DNA bands were cut out of the gel and purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's protocol.

Restriction enzyme digestion

Purified plasmids and PCR products were digested either with standard or FastDigest (FD) restriction enzymes (all Thermo Fisher Scientific, see table 7). Double digests with standard enzymes were performed with 1 µg DNA and 0.5-1 µL of each enzyme in compatible buffer in a total volume of 20 µL. After incubation for 2 h at 37 °C under shaking (600 rpm) the digestion mixtures were purified by agarose gel electrophoresis.
Double digests with FD restriction enzymes were performed with 1 µg DNA and 1 µL of each FD enzyme in FastDigest Green Buffer (1x) in a total volume of 20 µL. The reactions

were incubated for 15 min at 37 °C under shaking (600 rpm) and the DNA fragments purified by agarose gel electrophoresis.

Ligation of DNA fragments

Ligations of digested and purified vectors with inserts were performed with 2 µL of the vector and 10 µL of the insert (1:5 ratio) in T4 DNA Ligase buffer using T4 DNA Ligase (1 U) (Thermo Fisher Scientific) in a total volume of 20 µL. The ligation reactions were incubated for 30 min at room temperature, followed by inactivation of the DNA ligase at 65 °C for 10 min.

DNA sequencing

The sequence of cloned plasmids was verified by DNA sequencing. Therefore, 5 µL of a plasmid solution (80-100 ng/µL) were mixed with 5 µL of the corresponding sequencing primer (5 µM) (see table 11) and the sequencing was conducted by GATC Biotech using LIGHTRUN™ Sanger sequencing. Sequencing results were analyzed with the ApE software.

4.2.4.3 Polymerase chain reaction

Polymerase chain reaction (PCR) was used for amplification of DNA sequences. PCRs were performed with *Pfu* DNA Polymerase (Thermo Fisher Scientific) in a final volume of 50 µL or with Q5® DNA Polymerase (NEB) in a final volume of 25 µL. The pipetting schemes are shown in table 15. PCR components were mixed on ice in PCR tubes and reactions were conducted in a Techne TC-3000 thermocycler. The general thermocycling conditions for PCRs using *Pfu* and Q5® DNA Polymerase are shown in table 16 and table 17, respectively.

Table 15: Pipetting schemes for PCRs with *Pfu* or Q5® DNA Polymerase.

Component	Final concentrations	
	Pfu	**Q5**
Reaction buffer (10x *Pfu* or 5x Q5)	1x	1x
Template DNA	50 ng	1-100 ng
10 µM Primer forward	0.2 µM	0.5 µM
10 µM Primer reverse	0.2 µM	0.5 µM
2 mM dNTPs	200 µM	200 µM
Polymerase (*Pfu* or Q5)	0.05 U/µL	0.02 U/µL

Table 16: Thermocycling conditions for PCRs using *Pfu* DNA Polymerase.

Step	Temperature	Time
Initial denaturation	95 °C	3 min
30 cycles	95 °C	30 s
	T_a	30 s
	72 °C	t_e
Final extension	72 °C	15 min
Hold	4 °C	∞

Table 17: Thermocycling conditions for PCRs using Q5® DNA Polymerase.

Step	Temperature	Time
Initial denaturation	98 °C	30 s
35 cycles	98 °C	10 s
	T_a	20 s
	72 °C	t_e
Final extension	72 °C	2 min
Hold	4 °C	∞

Annealing temperatures (T_a) were adapted for each primer pair and, depending on length of the target DNA sequence, the elongation time (t_e) was adjusted. The specific conditions for the different PCRs are listed in table 18. PCR products were purified via agarose gel electrophoresis (4.2.4.2).

4.2.4.4 Cloning of bacterial expression vectors

The cloned vectors for bacterial protein expressions are listed in table 19. All genes for recombinant expression of bromodomain constructs were purchased as synthetic, codon optimized (*E. coli*) genes from GeneArt Gene Synthesis (Thermo Fisher Scientific). The plasmids containing the synthetic genes are listed in table 9. The constructs BAZ2B-mTagBFP-*Strep*, BRD4(1)-TagGFP2-*Strep* (cloned by *Maren Scheffler* [239]) and CREBBP-TurboYFP-*Strep* were directly inserted into the pET28(+) vector using the NdeI/HindIII restriction sites. This added a N-terminal *His_6*-tag and a thrombin cleavage site (Thr) resulting in the final expression constructs.
The BRD3(2) gene was amplified by PCR using the primers NdeI-BRD3(2)-fwd/BRD3(2)-BamHI-rev and the TurboYFP gene was amplified using the primers BamHI-TurboYFP-

Table 18: Conditions for cloning PCRs. Primers are listed in table 11.

Target	Primer	Polymerase	T_a	t_e
BRD3(2)	NdeI-BRD3(2)-fwd, BRD3(2)-BamHI-rev	*Pfu*	65 °C	1 min
BRD3(2)	NdeI-BRD3(2)-fwd, BRD3(2)-HindIII-rev	*Pfu*	65 °C	1 min
BRD4(1)	NdeI-BRD4(1)-fwd/BRD4(1)-HindIII-rev	Q5	67 °C	20 s
TurboYFP	BamHI-TurboYFP-fwd, TurboYFP-HindIII-rev	*Pfu*	65 °C	1 min
$mini^{TAG}$, αTub^{TAG}	BamHI-Kozak-fwd, TEV-ClaI-rev	Q5	64 °C	20 s

fwd/TurboYFP-HindIII-rev (primers listed in table 11). A 3-fragment ligation into the pET28(+) vector led to the BRD3(2)-TurboYFP expression construct with an additional BamHI restriction site between BRD and fluorescent label.
The BRD4(2) gene was cloned into the pET28-BRD3(2)-TurboYFP plasmid using the restriction sites NdeI/BamHI resulting in the final BRD4(2)-TurboYFP expression construct.
The fluorescent label from BRD3(2) was removed by amplifying the gene by PCR using the primer pair NdeI-BRD3(2)-fwd/BRD3(2)-HindIII-rev and insertion of the fragment into the pET28a(+) vector using the restriction sites NdeI/HindIII. The fluorescent label from BRD4(1) was removed analogous using the primer pair NdeI-BRD4(1)-fwd/BRD4(1)-HindIII-rev for PCR amplification. Plasmid maps of the final expression vectors are shown in figure A.2.

Table 19: Overview of the cloned bacterial expression vectors.

Backbone	Protein construct	Resistance	Size
pET28a(+)	His_6-Thr-BAZ2B-TEV-mTagBFP-*Strep*	Kan	6416 bp
pET28a(+)	His_6-Thr-BRD3(2)	Kan	5645 bp
pET28a(+)	His_6-Thr-BRD3(2)-TEV-TurboYFP-*Strep*	Kan	6488 bp
pET28a(+)	His_6-Thr-BRD4(1)	Kan	5687 bp
pET28a(+)	His_6-Thr-BRD4(1)-TEV-TagGFP2-*Strep*	Kan	6467 bp
pET28a(+)	His_6-Thr-BRD4(2)-TEV-TurboYFP-*Strep*	Kan	6488 bp
pET28a(+)	His_6-Thr-CREBBP-TEV-TurboYFP-*Strep*	Kan	6455 bp

4.2.4.5 Cloning of mammalian expression vectors

Codon optimized (*H. sapiens*) synthetic genes for mammalian protein expression were received from GeneArt Gene Synthesis (Thermo Fisher Scientific). The genes coding

for the different donor (Don) and acceptor (Acc) protein constructs were cloned into the bidirectional pBI-CMV1 vector (Clontech) by using BamHI/EcoRV (MCS 1) restriction sites for donor and EcoRI/XbaI (MCS 2) for acceptor constructs resulting in plasmids for single (Don, Acc) and combined (FRET) protein expression (table 20). The donor plasmids were cloned first and in a second step, acceptor genes were ligated into MCS 2. The plasmids Don2-4, Acc2 and FRET2-4 were cloned by *Annika Kühn* [235]. Therefore, the p53^{TAG}-mClover3 gene was directly ligated into MCS 1, resulting in the Don2 plasmid with an additional ClaI restriction site between binding sequence and fluorescent label. The synthetic genes coding for the p53^{TAG} and mini^{TAG} binding sequences were amplified using the primer pair BamHI-Kozak-fwd/TEV-ClaI-rev (table 11) before they were ligated into the Don2 plasmid using the BamHI/ClaI restriction sites (Don3, Don4).

Table 20: Overview of cloned mammalian expression vectors.

Name	Backbone	MCS 1	MCS 2
Don1	pBI-CMV1	H4K20^{TAG}-mCerulean3	-
Acc1	pBI-CMV1	-	BRD3(2)-EYFP
FRET1	pBI-CMV1	H4K20^{TAG}-mCerulean3	BRD3(2)-EYFP
Don2	pBI-CMV1	p53^{TAG}-mClover3	-
Don3	pBI-CMV1	mini^{TAG}-mClover3	-
Don4	pBI-CMV1	αTub^{TAG}-mClover3	-
Acc2	pBI-CMV1	-	HDAC1-mRuby3
FRET2	pBI-CMV1	p53^{TAG}-mClover3	HDAC1-mRuby3
FRET3	pBI-CMV1	mini^{TAG}-mClover3	HDAC1-mRuby3
FRET4	pBI-CMV1	αTub^{TAG}-mClover3	HDAC1-mRuby3

4.2.4.6 Bacterial protein expression and purification

Pre-cultures of 20 mL LB medium with kanamycin (30 µg/mL) were inoculated with cultures of *E. coli* BL-21 (DE3) (Agilent) containing the various expression vectors and incubated over night. 700 mL LB medium supplemented with kanamycin (30 µg/mL) were inoculated with 7 mL pre-culture (1:100 dilution) and incubated until the OD_{600} reached 0.5-0.7. The protein expression was induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. BRD constructs were expressed for 18-20 h at 25 °C and Sortase A for 16 h at 20 °C.
Cells were harvested by centrifugation (5000 × g, 10 min, 4 °C,) and resuspended on ice in lysis buffer (20 mM Tris-HCl, 0.1 % Triton X-100 (v/v), pH 6.5). Bacteria were lysed with an EmulsiFlex-C5 homogenizer (Avestin) and the lysates were centrifuged (15,000 × g, 30 min, 4 °C).

Ni-NTA purification

Cleared supernatant was supplemented with imidazole (30 mM) and applied to Ni-NTA Superflow resin (Qiagen). The suspension was incubated for 1 h at 4 °C on an overhead shaker. Beads were transferred into a gravity flow column, washed with washing buffer (20 mM Tris-HCl, 500 mM NaCl, 30 mM imidazole, pH 7.2) (25 bed volumes (BV)) and proteins were eluted in fractions (10 $\times$ 1 BV) with elution buffer (20 mM Tris-HCl, 500 mM NaCl, 300 mM imidazole, pH 7.2). Fractions were analyzed by coomassie-stained SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (4.2.4.12).

Strep-Tactin purification

After Ni-NTA purification the BRD3(2)-TurboYFP protein was further purified via the C-terminal *Strep*-tag using Strep-Tactin XT Superflow high capacity resin (IBA). A purification column was prepared by equilibrating Strep-Tactin XT beads with buffer W (100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0) (2 $\times$ 5 BV). The protein solution was applied to the column and the beads were washed (5 $\times$ 5 BV). Protein was eluted in fractions (5 $\times$ 0.5 BV) with buffer BXT (100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 50 mM biotin, pH 8.0). Fractions were analyzed by coomassie-stained SDS-PAGE (4.2.4.12).

Thrombin protease cleavage

The *His*$_6$-tag of BRDs used for crystallization (BRD3(2) and BRD4(1)) was removed by thrombin (GE Healthcare) digestion over night at room temperature. The reaction was stopped with 1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF) (Carl Roth) and purified by gel filtration (Superdex 75, 16/60, GE Healthcare) using crystallization buffer (10 mM HEPES, 150 mM NaCl, 1 mM DTT, pH 7.5). Fractions were analyzed by SDS-PAGE and pure fractions were pooled and concentrated (Amicon Ultra-15, PLBC Ultracel-PL Membran, 3 kDa, Merck).

Dialysis and protein storage

Pure fractions were combined and dialyzed (Spectra/Por™ Membrane, 6-8 kDa cutoff, Spectrum Laboratories) against the corresponding buffer (see table 21) over night at 4 °C. Protein concentrations were determined using a NanoDrop ND-1000 (Peqlab) at 280 nm (protein-specific values are listed in figure A.2) and protein stocks were aliquoted and stored at -20 °C.

4.2.4.7 Protein crystallization and structure determination

Crystallization condition screenings, X-ray diffraction data collection and solving of structures were carried out by *Cengiz Koç*, *Natascha Bartlick* and *Michael Braun* (group of Prof. Dr. Thilo Stehle, University of Tübingen). Final crystallization conditions are shown in table 22. Data collection was performed on a beamline X06DA (PXIII) at the Swiss Light Source (Villigen, Schwitzerland) at a wavelength of 1.0 Å and processed with XDS

Table 21: Dialysis buffers for recombinant proteins.

Protein	Dialysis buffer
BRDs for pull-downs	20 mM HEPES, 100 mM KCl, 20 % gylcerol, pH 7.9
BRDs for crystallization	20 mM Tris-HCl, 100 mM NaCl, 0.3 mM $CaCl_2$, pH 8.5
BRDs for MST	PBST
Sortase A	50 mM Tris-HCl, 150 mM NaCl, 5 mM $CaCl_2$, pH 7.5

[310]. The resulting datasets were analysed for radiation damage by XDSSTAT [311]. The structures were solved by molecular replacement with Phaser [312] in CCP4 using the Apo structures of BRD3(2) and BRD4(1) (PDB-ID: 2OO1/2OSS, [242]), respectively. Initial refinement involved simulated annealing as implemented in PHENIX [313] to reduce model bias. Further refinement was carried out by manual model building in Coot [314] alternated with restraint refinement using phenix.refine [313]. Data collection and refinement statistics are listed in table A.3 of the appendix. Illustrations of crystal structures were created with PyMOL 2.5.1.

Table 22: Crystallization conditions for BRD3(2) and BRD4(1) with L-ApmTri peptides.

	BRD3(2)	**BRD4(1)**	**BRD4(1)**
Ligand	H4K20_L-ApmTri	H4K5acK8ApmTri	H4K5/K8ApmTri
Condition	0.2 M NaF 0.1 M BTProp pH 7.5 10 % Ethylenglycol 20 % PEG 3350	0.2 M $NaOAc \cdot 3\,H_2O$ 0.1 M Tris-HCl pH 8.5 30 % (w/v) PEG 4000	0.2 M $NaOAc \cdot 3\,H_2O$ 0.1 M Tris-HCl pH 8.5 30 % (w/v) PEG 4000
Method	4 °C, 24 well, hanging drop	20 °C, 96 well, sitting drop	20 °C, 96 well, sitting drop
Prot. conc.	10 mg/mL	10 mg/mL	1 mg/mL
Prot:pep ratio	1:2.2	1:2	1:20

4.2.4.8 General cultivation of mammalian cell lines

HEK293T and HeLa-S3 cells were cultured adherent in Dulbecco's Modified Eagle's Medium (DMEM) (high glucose) supplemented with 10 % fetal calf serum (FCS) and 1 % penicillin-streptomycin (Sigma-Aldrich) at 37 °C in a humidified atmosphere with 5 % CO_2. Subcultivation was performed every 3-4 days by treating the cells with 0.05 % trypsin containing 0.2 g/L EDTA (Sigma-Aldrich). Afterwards cells were reseeded in fresh medium.

4.2.4.9 Mammalian cell transfection

HEK293T cells were seeded 24 h or 48 h prior to transfection in 24-well, 12-well or 6-well plates (all Sarstedt). Cells were transiently transfected at 60-80 % confluency using the transfection reagent jetPRIME™ (Polyplus Transfection) according to the manufacturer's protocol. The suggested amount of plasmid DNA was diluted in jetPRIME™ buffer (table 23) and vortexed for 10 s. The jetPRIME™ reagent was added and the solution was vortexed for 10 s. The transfection mixture was incubated for 10 min at room temperature before it was added dropwise to the cells. After 4 h, medium was replaced with fresh pre-warmed growth medium.

Table 23: Reagents for transfection of HEK293T cells using jetPRIME™ transfection reagent.

Reagent	24-well plate	12-well plate	6-well plate
jetPRIME™ buffer [uL/well]	50	75	200
jetPRIME™ reagent [uL/well]	1.0	1.6	4.0
Plasmid DNA [ug/well]	0.5	0.8	2.0
Volume of growth medium [mL/well]	0.5	1.0	2.0

4.2.4.10 Protein expression using amber suppression in HEK293T cells

Stock solutions of the non-canonical amino acids (ncAA) Nε-Boc-L-Lysine (BocK), L-ApmTri (**16**), L-AsuHd (**37**) and of L-Lysine were prepared in sterile solvents (table 24), aliquoted and stored at -20 °C. The ncAAs were added to transfected HEK293T cells in different concentrations after replacing the transfection medium. The BocK stock solution was diluted 1:4 with sterile 1 M HEPES buffer (pH not adjusted), ApmTri and AsuHd stock solutions were diluted 1:4 with sterile ncAA dilution buffer (0.2 M NaOH in 1 M HEPES buffer, pH not adjusted). Diluted amino acid solutions were added dropwise to the growth medium and cells were analyzed 20 h or 40 h after treatment.

Table 24: Stock solutions of ncAAs for amber suppression experiments.

NcAA	Stock conc.	Solvent
Nε-Boc-L-Lysine (BocK)	100 mM	15 % DMSO, 0.2 M NaOH/water
L-ApmTri (**16** via scheme 3)	1 M	water
L-AsuHd (**37** via scheme 4)	1 M	water
L-Lysine	1 M	water

4.2.4.11 Pull-downs with BRDs and immobilized peptide probes

Pull-down experiments with BRDs were performed on resin-bound peptide probes (see 4.2.3.4). An aliquot (10 µL beads) of each probe was equilibrated (3 × 200 µL) with PD buffer (20 mM HEPES, 100 mM KCl, 20 % glycerol, pH 7.9) by shaking at 700 rpm for 2 min at room temperature. Between each step the beads were centrifuged (1700 × g, 1 min, rt) and the supernatant discarded. Purified bromodomain constructs were diluted to the desired concentration with PD buffer and supplemented with bovine serum albumin (BSA) (concentrations are listed in table 25). The beads were incubated with 200 µL of the prepared protein solution by shaking at 700 rpm for 1 h at room temperature. Protein solution was removed by centrifugation (1700 × g, 1 min, rt) and the beads were washed (4 × 200 µL) with PD washing buffer (20 mM HEPES, 300 mM KCl, 20 % glycerol, pH 7.9) and PD buffer (2 × 200 µL) by shaking at 700 rpm for 2 min at room temperature. Pull-downs with BRD4(1) were only washed with PD buffer (4 × 200 µL) by shaking at 700 rpm for 30 s at room temperature. Between washing steps, beads were centrifuged (1700 × g, 1 min) and the supernatant discarded. Peptide-bound proteins were eluted by incubation with 30 µL of 3x SDS sample buffer (5x SDS sample buffer: 250 mM Tris, 10 % SDS, 30 % glycerol, 0.5 M DTT, 0.02 % bromophenol blue) for 10 min at 95 °C while shaking at 700 rpm, followed by centrifugation (1700 × g, 1 min). Appropriate amounts of the samples and an input sample of 2 µg protein were loaded onto polyacrylamide gels (12 %), separated by SDS-PAGE and stained with Coomassie brilliant blue G 250 staining solution. Each pull-down was performed at least in triplicate and intensities were quantified with the Image Lab software (Bio-Rad).

Table 25: Conditions for pull-down experiments with BRDs and immobilized peptide probes.

Protein	Concentration	BSA	Washing
BAZ2B	7.6 µM	10 mg/mL	4 × 2 min PD washing buffer, 2 × 2 min PD buffer
BRD3(2)	10.5 µM		
BRD4(2)	10.5 µM		
CREBBP	13.8 µM		
BRD4(1)	6.8 µM	0.1 mg/mL	4 × 30 s PD buffer

Inhibitor titration assay

For inhibitor titration experiments, pull-downs were performed as described above with a BRD protein concentration of 10.5 µM for BRD3(2) and BAZ2B. Stock solutions of the bromodomain inhibitors (+)-JQ1 or GSK2801 (both Sigma-Aldrich) were prepared (1 M in DMSO) and the inhibitors were added to the incubation solutions with final concentrations between 0 and 20 µM.

4.2.4.12 SDS-polyacrylamide gel electrophoresis

Protein samples were supplemented with SDS sample buffer and heated for 10 min at 95 °C. Protein samples were loaded onto polyacrylamide gels (table 26) and a PageRuler Prestained Protein Ladder (Thermo Fisher Scientific) was used as maker. Separation was performed at 120 V for 70 min in SDS-PAGE running buffer (buffer compositions are shown in table 27. Cy5 detection was carried out directly on a ChemiDoc MP imaging system (Bio-Rad). Protein staining was performed by boiling and incubating the gels for 5 min in water under shaking. After repeating this step, Coomassie Brilliant Blue staining solution (80 mg Coomassie Brilliant Blue G250 (Sigma Aldrich), 3 mL of 37 % HCl in 1 L water) was added and heated up. Gels were incubated for 15-30 min under shaking, before the solution was removed and destaining was performed in water over night. Destained gels were imaged with the ChemiDoc MP system.

Table 26: Polyacrylamide gels used for protein gel electrophoresis.

Gel type	Supplier
12 % polyacrylamide gel	self-casted (table 27)
Any kD™ Mini-PROTEAN® TGX™ precast gradient gel	Bio-Rad
4-20 % Mini-PROTEAN® TGX™ precast gradient gel	Bio-Rad

Table 27: Buffer compositions for SDS-PAGE.

Gel type	Composition
Polyacrylamide stacking gel	250 mM Tris-HCl pH 6.8, 4 % ROTIPHORESE® Gel 30 (37.5:1), 0.1 % SDS, 0.04 % APS, 0.01 % TEMED
Polyacrylamide separating gel	375 mM Tris-HCl pH 8.8, 12 % ROTIPHORESE® Gel 30 (37.5:1), 0.1 % SDS, 0.04 % APS, 0.01 % TEMED
SDS-PAGE running buffer	25 mM Tris-HCl pH 8.8, 192 mM glycine, 0.1 % SDS
5x SDS sample buffer	250 mM Tris, 10 % SDS, 30 % glycerol, 0.5 M DTT, 0.02 % bromophenol blue

4.2.4.13 Native-polyacrylamide gel electrophoresis

Samples for native PAGE were supplemented with native sample buffer (2x native sample buffer: 250 mM Tris, 30 % glycerol, 0.02 % bromophenol blue) and loaded onto precasted 4-20 % gradient gels (table 26). Separation was performed at 140 V for 60 min in native PAGE running buffer (25 mM Tris-HCl pH 8.8, 192 mM glycine). Cy5 detection was carried out directly on a ChemiDoc MP imaging system (Bio-Rad).

4.2.4.14 Western blotting

SDS-PAGE gels were equilibrated in blotting buffer (25 mM Tris-HCl, 192 mM glycine, 0.05 % SDS, 10 % methanol) for 30 min under shaking at room temperature. PVDF membranes (Amersham Hybond LFP 0.2 µm, GE Healthcare) were activated by incubation in methanol (15 s), followed by water (2 min) and blotting buffer (30 min). Western blotting of protein samples was performed with a Mini Trans-Blot® Cell wet tank system (Bio-Rad) with a constant current of 25 mA over night at 4 °C. After protein transfer membranes were blocked with blocking buffer (5 % low fat powdered milk in phosphate buffered saline (PBS) (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4; for incubation with Penta-His antibody: 5 % BSA in PBS) for 30 min under shaking at room temperature and washed with PBST (0.1 % Tween 20 (v/v) in PBS) (2 $\times$10 s). Primary antibody solutions were prepared with primary incubation buffer (5 % BSA in PBST, 0.05 % NaN_3) using dilutions listed in table 28. Membranes were incubated with primary antibody solutions for 1-2 h at room temperature or at 4 °C over night under shaking and were washed afterwards with PBST (3 $\times$10 s, 3 $\times$ 5 min). HRP conjugated secondary antibody solutions were prepared in secondary incubation buffer (5 % low fat powdered milk in PBST) and IRDye® conjugated secondary antibody were prepared in IR-secondary incubation buffer (5 % low fat powdered milk in PBST with 0.01 % SDS). Membranes were incubated with secondary antibody solutions for 1 h under shaking at room temperature. Membranes treated with HRP conjugated secondary antibodies were washed with PBST (3 $\times$10 s, 3 $\times$ 5 min) and chemiluminescene detection was performed with Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific). Membranes with fluorescently labeled antibodies were washed with PBST (3 $\times$ 10 s, 3 $\times$ 5 min) and PBS (2 $\times$10 s). Chemiluminescence and fluorescence were detected using a ChemiDoc MP imaging system (Bio-Rad).

Table 28: Antibodies used for immunoblotting experiments.

Primary antibodies	Host/Isotype	Dilution	Product Nr.	Supplier
Cyclophilin B	rabbit-poly-IgG	1:500	ab16045	Abcam
HDAC1	mouse-mono-IgG1	1:500	sc-81598	Santa Cruz
HDAC2	goat-poly-IgG	1:500	sc-6296	Santa Cruz
HDAC3	rabbit-poly-IgG	1:500	sc-11417	Santa Cruz
HDAC4	mouse-mono-IgG2b	1:500	sc-365093	Santa Cruz
HDAC6	rabbit-poly-IgG	1:500	sc-11420	Santa Cruz
HDAC8	rabbit-mono-IgG	1:200	MA5-32422	Thermo Fisher Scientific
Histone H3	rabbit-poly-IgG	1:1000	ab1791	Abcam
OctA-Probe FLAG	rabbit-poly-IgG	1:1000	sc-807	Santa Cruz

Penta-His	mouse-mono-IgG1	1:1000	34660	Qiagen
Spot-Label ATTO488	alpaca-mono-V_HH	1:2000	eba488-50	ChromoTek
Secondary antibodies	**Host/Isotype**	**Dilution**	**Product Nr.**	**Supplier**
mouse anti-rabbit	IgG-HRP conjugate	1:5000	sc-2357	Santa Cruz
recombinant anti-mouse	IgGκ BP-HRP conjugate	1:5000	sc-516102	Santa Cruz
donkey anti-goat	IgG-HRP conjugate	1:5000	sc-2020	Santa Cruz
donkey anti-rabbit	IgG-IRDye® 800 CW conjugate	1:5000	926-32213	LI-COR
donkey anti-mouse	IgG-IRDye® 800 CW conjugate	1:5000	926-32212	LI-COR

4.2.4.15 Mircoscale Thermophoresis

Peptide stock solutions (100 mM) were prepared in water and stored at -20 °C. Concentrations of peptide stock solutions were validated by acid hydrolysis, followed by chromatographic quantification of the amino acids (Frank Gutjahr Chromatographie). The determined concentrations were used for the dilution of the peptide stock solutions with PBST. 16-step serial dilutions of the peptides were prepared in PBST buffer (diluted with water, depending on the maximum ligand concentration).
Stock solutions of the recombinant proteins *His*$_6$-BRD3(2) and *His*$_6$-BRD4(1) were centrifuged (14,000 $\times$ g, 10 min, 4 °C) prior to use. Proteins were labeled with the Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation (NanoTemper) according to the manufacturer's protocol.
After adding the labeled BRDs to a final concentration of 50 nM, the solutions were incubated for at least 30 min in the dark at room temperature, before they were soaked into Monolith NT.115 Premium Capillaries (NanoTemper). MST was measured on a Monolith NT.115 device (NanoTemper) at 60 % LED (red) and 40 % MST power for BRD4(1) or 60 % LED (red) and 60 % MST power for BRD3(2) at 25 °C. Data were acquired with the MO.Control software v.1.6 and analyzed with the MO.Affinity Analysis v.2.3 software (both NanoTemper). Each K_D determination was performed with at least 5 independently prepared serial dilution replicates (n $\geq$ 5). The MST on-time with the highest signal-to-noise ratio was used for K_D determination and outliers were excluded (each single data point n $\geq$ 3). Data were fitted using a K_D fit model describing a molecular interaction with a 1:1 stoichiometry according to the law of mass action:

$$f(c) = Unbound + (Bound - Unbound) \cdot \frac{c + c_{target} + K_D - \sqrt{(c + c_{target} + K_D)^2 - 4cc_{target}}}{2c_{target}}$$

where the $f(c)$ is the fraction bound at a given ligand concentration c; $Unbound$ is the F_{norm} signal of the target; $Bound$ is the F_{norm} signal of the complex; K_D is the binding affinity and c_{target} is the final concentration of target in the assay [245]. Final figures were created with GraphPad Prism 9.0.0.

4.2.4.16 MALDI-MS based deacetylation assay

Deacetylation reactions for analysis with MALDI-TOF mass spectrometry were performed in a volume of 20 µL [246]. The peptide substrate was solved in HDAC buffer (10 mM HEPES, 100 mM NaCl, 8 mM KCl, pH 8.0) at a concentration of 400 µM. 15 µg HeLa whole cell extract were diluted with HDAC buffer to a volume of 10 µL and added to the peptide solution. The reaction mixture was incubated at 25 °C for 24 h. 2 µL samples were taken after 0 h and 4 h and quenched with 8 µL of stopping solution (6.25 µM trichostatin A in 0.1 % TFA in water). Afterwards, 10 µL of the quenched assay solution were diluted with 90 µL of 0.1 % TFA in water. Samples were measured on a Bruker Reflex IV MALDI-TOF mass spectrometer (4.2.1.4).

4.2.4.17 Preparation of cell extracts

HEK293T whole cell extracts for WB analysis

HEK293T cells were grown for 1-2 days after transfection in 12-well or 6-well-plates (given volumes were used for 6-well plates; for 12-well plates, half of the volumes were used). The growth medium was removed and the cells were washed with ice-cold PBS (1 mL/well). Cells were washed off the plate, resuspended with ice-cold PBS (600 µL/well) and transferred into pre-cooled reaction tubes. The suspension was centrifuged (500 × g, 3 min, 4 °C), the pellet was washed twice with ice-cold PBS (600 µL) and centrifuged again (2000 × g, 5 min, 4 °C). After removing the supernatant, cells were suspended in lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 0.5 mM EDTA, 0.5 % NP-40 (v/v), pH 7.5), supplemented with 1 mM Na_3VO_4, 2 mM NaF and 1:100 Protease Inhibitor Cocktail (Sigma-Aldrich, P8340). Cells were incubated on ice for 30 min while vortexing every 10 min. Cell lysates were cleared by centrifugation (14,100 × g, 1 h, 4 °C) and stored at -20 °C or at -80 °C.

HEK293T whole cell extracts for pull-down experiments

HEK293T cells were grown for three days in 6-well-plates. The growth medium was removed and the cells were washed with ice-cold PBS (1 mL/well). Cells were washed off the plate and resuspended with ice-cold PBS (1 mL/well) and centrifuged (2000 × g, 5 min, 4 °C). The pellet was washed with ice-cold PBS (0.5 mL/well) and centrifuged again (2000 × g, 5 min, 4 °C). After removing the supernatant, cells were suspended in CHAPS lysis buffer (20 mM HEPES, 300 mM NaCl, 2 mM $MgCl_2$, 0.4 % CHAPS (w/v), pH 7.9), supplemented with 1 mM Na_3VO_4, 2 mM NaF and 1:100 Protease Inhibitor Cocktail (Sigma-Aldrich, P8340). Cells were incubated on ice for 30 min while vortexing every

10 min. The cell lysate was cleared by centrifugation (15,000 × g, 1 h, 4 °C), aliquoted and stored at -80 °C.

HeLa whole cell extract

HeLa whole cell extract (WCE) used for the MALDI-MS based deacetylation assay was prepared following a protocol adapted from Dignam et al. [315]. Buffers are listed in table 29. Buffer A_D, B_D and C_D were supplemented with 0.5 mM AEBSF, 0.5 mM DTT, 1 mM Na_3VO_4 and 2 mM NaF and buffer D_D was supplemented with 0.25 mM AEBSF prior to use.

Table 29: Buffer compositions for preparation of WCE according to the Dignam protocol.

Buffer	Composition
A_D	10 mM HEPES, 10 mM KCl, 1.5 mM $MgCl_2$, pH 7.9 (4 °C)
B_D	300 mM HEPES, 1.4 M KCl, 30 mM $MgCl_2$, pH 7.9 (4 °C)
C_D	20 mM HEPES, 420 mM NaCl, 1.5 mM $MgCl_2$, 25 % glycerol, pH 7.9 (4 °C)
D_D	20 mM HEPES, 100 mM KCl, 20 % glycerol, pH 7.9 (4 °C)

HeLa cells were grown for four days in 165 cm^2 cell culture dishes (Sarstedt). After removing medium, cells were washed with ice-cold PBS (10 mL), harvested in additional 10 mL PBS with a cell scraper and centrifuged (1000 × g, 10 min, 4 °C). Supernatant was removed and an approximate pellet volume V_P was determined. The pellet was resuspended in ice-cold PBS (5 × V_P) and centrifuged (1000 × g, 10 min, 4 °C). Cells were suspended in buffer A_D (2 × V_P), homogenized using a glass homogenizer and centrifuged (1000 × g, 10 min, 4 °C). The supernatant was transferred to a fresh tube and the pellet was centrifuged again (25,000 × g, 20 min, 4 °C). The supernatant was combined with the previous one and the pellet was suspended in buffer C_D (1 mL per 1 × 10^8 cells). After additional homogenization, the suspension was stirred for 30 min on ice followed by centrifugation (25,000 × g, 20 min, 4 °C) to yield the nuclear extract. The volume V_{CE} of the combined supernatants was determined, buffer B_D (0.11 × V_{CE}) added and centrifuged (100,000 × g, 1 h, 4 °C) to yield the cytosolic extract. Whole cell extract was obtained by combining the nuclear and cytosolic extracts and WCE was dialyzed (Spectra/Por™ Membrane, 3 kDa cutoff, Spectrum Laboratories) against buffer D_D over night at 4 °C. The WCE was cleared by centrifugation (25,000 × g, 25 min, 4 °C), aliquoted and stored at -80 °C.

4.2.4.18 Determination of protein concentrations in cell lysates

Protein quantifications of cell lysates were performed via bicinchoninic acid (BCA) assay using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Cell extract and protein standard (BSA, 0-2 mg/mL, Thermo Fisher Scientific) dilutions were prepared in water as triplicates. In 96-well plates (CellStar®, Greiner), 200 µL of BCA solution (mixture of

reagent A and B 50:1 (v/v)) and 20 µL of sample or protein standard dilutions were mixed and incubated at 37 °C for 30 min. Readout was performed on an Infinite 200 Pro microplate reader (Tecan) at 562 nm.

4.2.4.19 Co-precipitation of *Strep*-tag labeled proteins from mammalian cell lysates.

Affinity purification of Strep-Tag labeled proteins

HEK293T cell lysates were diluted 1:1 with dilution buffer (10 mM Tris, 150 mM NaCl, 0.5 mM EDTA, pH 7.5), supplemented with 1 mM Na_3VO_4, 2 mM NaF and 1:100 Protease Inhibitor Cocktail (Sigma-Aldrich, P8340). Donor proteins were purified via N-terminal *Strep*-tag by incubation of a Strep-Tactin® XT coated 96-well plate (IBA) with 200 µL of diluted cell lysate over night at 4 °C. The incubation with (+)-JQ1 (Sigma-Aldrich) was performed by adding 1 µM (+)-JQ1 (1 M in DMSO) to the cell lysate prior to incubation. Wells were washed with PBS (5 × 200 µL) under shaking at 400 rpm for 30 s at room temperature. Proteins were eluted by adding 50 µL of pre-heated (95 °C) 2x SDS sample buffer to each well and incubation for 10 min at 70 °C while shaking at 400 rpm. Samples were analyzed by SDS-PAGE and western blotting. Applied sample amounts were adjusted by western blot intensities for Donor proteins.

Immunoprecipitation of *Spot*-tag fusion proteins

Spot-tag fusion proteins were purified by immunoprecipitation (IP) using Spot-Trap agarose beads (ChromoTek). Therefore, 200 µL HEK293T cell lysate were diluted with 300 µL dilution buffer (10 mM Tris, 150 mM NaCl, 0.5 mM EDTA, pH 7.5) supplemented with 1 mM Na_3VO_4, 2 mM NaF and 1:100 Protease Inhibitor Cocktail (Sigma-Aldrich, P8340). In a microcentrifuge filter unit (Ultrafree-MC, HV 0.45 µm, Merck), 25 µL of the Spot-Trap agarose bead slurry were mixed with 500 µL ice-cold dilution buffer and centrifuged (2500 × g, 2 min, 4 °C). The diluted cell lysate was added and the beads were incubated for 1 h on an overhead shaker at 4 °C. After incubation, lysate was removed by centrifugation (2500 × g, 2 min, 4 °C) and the beads were washed (3 × 500 µL) with Spot-Trap washing buffer (10 mM Tris, 150 mM NaCl, 0.5 mM EDTA, 0.2 % Triton X-100, 0.2 % deoxycholate, pH 7.5) for 5 min while end-over-end tumbling at 4 °C. Washing buffer was removed by centrifugation (2500 × g, 2 min, 4 °C). Protein elution was performed by suspending the beads in 100 µL 2x SDS sample buffer and incubation for 10 min at 95 °C. After centrifugation (2500 × g, 2 min, 4 °C), 20 µL of the eluate were used for western blot analysis.

4.2.4.20 Cell microscopy

For cell microscopy, coverslips were coated with Poly-D-Lysine in 24-well plates. Therefore, 300 µL of a 20 µg/mL Poly-D-Lysine solution (Thermo Fisher Scientific) were added to each well and incubated over night at room temperature. The Poly-D-Lysine solution

was removed, coverslips were washed (3 × 500 µL) with sterile water and dried for 2 h. HEK293T cells were seeded and transfected as described (4.2.4.9).
Living cells were analyzed on coverslips in PBS buffer with a microscope system consisting of an upright Examiner.Z1 (Carl Zeiss), a Yokogawa CSU-X1 spinning disk confocal scanner (Yokogawa Denki), three diode lasers (445 nm, 488 nm and 561 nm) and two water immersion objectives (W N-ACHROMAT 10/0.3, W Plan-APOCHROMAT 40/1.0 DIC VIS-IR, all Carl Zeiss). Fluorescence images were taken with a CCD camera (SPOT Pursuit, SPOT Imaging) using the excitation lasers and filter sets listed in table 30. Images were further processed using the ImageJ v2.1.0/1.53c software package.

Table 30: Laser and filter settings for the different fluorophores in fluorescent microscopy.

Fluorophore	Excitation	Emission
mCerulean3	445 nm	470/24 nm
EGFP	488 nm	525/50 nm
mClover3	488 nm	525/50 nm
EYFP	488 nm	525/50 nm
mRuby3	561 nm	ZET/405/488/561/635M
iRFP713	561 nm	ZET/405/488/561/635M

4.2.4.21 Flow Cytometry

HEK293T cell samples for flow cytometry analysis were prepared 20 h after cell transfection by treating the cells with 0.05 % trypsin-EDTA solution (100 µL/well). Dissociated cells were diluted with fresh medium (400 µL), centrifuged (500 × g, 5 min, rt) and the supernatant discarded. Cells were washed with PBS (2 × 500 µL) and resuspended in 100 µL PBS. Flow cytometric measurements were performed on a CytoFLEX S system (Beckman Coulter) equipped with 488 nm and 561 nm lasers using CytoFLEX Sheath Fluid (Beckman Coulter) as system fluid. The different filter settings for each fluorophore are shown in table 31. Each sample was analyzed with a minimum of 10,000 events in the population of living cells defined by forward and side scatter (FSC and SSC) gating. Flow cytometric data were analyzed with FlowJo v10.5.3. Full gating strategies are illustrated in the appendix, figure A.6. Reciprocal compensation of FITC-A and PE-A channel was carried out with measurements of single fluorophore expressing cells resulting in the compensated channels Comp-FITC-A and Comp-PE-A.

4.2.4.22 Immobilization of peptides and NCPs on streptavidin-coated beads

A stock solution of the biotin-linker peptide was prepared (100 mM in water) and nucleosome core particle (NCP) solution was provided by *Zhadyra Yerkesh* (King Abdullah University of Science and Technology, Saudi Arabia). The Cyanine5 (Cy5) fluorescence

Table 31: Filter settings for flow cytometric cell analysis.

Fluorophore	Channel	Excitation	Emission
GFP	FITC	488 nm	525/40 nm
iRFP713	PC5.5	561 nm	690/50 nm
mClover3	FITC	488 nm	525/40 nm
mRuby3	PE	561 nm	585/42 nm

intensity of the NCP stock solution was measured prior to immobilization ($I_{Cy5}(before)$). Therefore, 50 µL were transferred into a 96-well plate (black, flat bottom, Greiner) and the fluorescence was measured (Ex./Em.: 649/680 nm) using a Infinite 200 PRO microplate reader (Tecan). Beads of 600 µL MagneSphere® Paramagnetic Particle slurry (magnetic beads) (Promega) were equilibrated (2 × 500 µL) with NCP buffer (10 mM Tris-HCl, 25 mM NaCl, 1 mM EDTA, 2 mM DTT, pH 7.5) for 2 min under shaking (1200 rpm) and incubated with 100 µL NCP solution (2.2 mg/mL) over night on an overhead shaker at 4 °C. Afterwards, the NCP solution was removed and the Cy5 intensity was measured again ($I_{Cy5}(after)$). The difference in fluorescence intensity was calculated with

$$\Delta I_{Cy5} = I_{Cy5}(before) - I_{Cy5}(after)$$

and the biotin-linker peptide was diluted with NCP buffer to the ΔI_{Cy5} intensity. Equilibrated MagneSphere® Paramagnetic Particles were incubated with 100 µL of the adjusted peptide solution for 2 h under shaking (1200 rpm) at room temperature. The solution was removed and Cy5 fluorescence intensity was measured again for confirmation of complete binding. After incubation with NCP or peptide solution, beads were washed (4 × 500 µL) with NCP buffer for 2 min under shaking (1200 rpm) at room temperature. Beads were stored in 200 µL NCP buffer at 4 °C.

4.2.4.23 Sortase-mediated ligation

After immobilization of NCPs and biotin-linker peptide, sortase-mediated ligations (SML) were performed with the synthesized H3 depsipeptides (H3 unmod., H3K27AsuHd, H3K9AsuHd, H3K4AsuHd). Magnetic beads were washed (2 × 200 µL) with SrtA buffer (10 mM Tris, 25 mM NaCl, 33 mM $CaCl_2$, pH 7.5) and two different ligation protocols were used. The ligation on biotin-linker peptide was performed with 200 µM H3 depsipeptide and 100 µM Sortase A in SrtA buffer. After 10 min under shaking (1200 rpm) at 37 °C, additional 200 µM H3 depsipeptide were added and reaction was stopped after 25 min by removing the reaction mixture. Ligations on NCPs were carried out with 12 µM depsipeptide and 3 µM Sortase A for 2 h under shaking (1200 rpm) at 37 °C. After 75 min, additional 12 µM depsipeptide were added. Following the ligation reaction, beads were washed (6 × 500 µL) with NCP buffer for 2 min under shaking (1200 rpm) at room temperature and stored in 200 µL NCP buffer at 4 °C.

4.2.4.24 NCP release from solid support

The NCP integrity was confirmed after SrtA ligation by releasing the NCPs from the magnetic beads. Therefore, 10 µL of the bead slurry were washed (2 × 50 µL) with FastDigest Buffer (Thermo Fisher Scientific) and incubated with 1 µL of FastDigest EcoRI in 20 µL FastDigest Buffer for 1 h under shaking (1200 rpm) at 37 °C. The NCPs were analyzed by native PAGE gel electrophoresis (4.2.4.13).

4.2.4.25 Pull-downs with probes immobilized on streptavidin-coated beads

For pull-down experiments on NCPs and peptides immobilized on magnetic beads, 20 µL slurry of each probe were used after SML. Beads were equilibrated (3 × 200 µL) with PD buffer (20 mM HEPES, 100 mM KCl, 20 % glycerol, pH 7.9) by shaking at 1200 rpm for 2 min at room temperature. HEK293T whole cell lysate (4.2.4.17) was diluted to a protein concentration of 2 mg/mL with PD buffer and beads were incubated with 200 µL of the lysate solution for 1 h under shaking (1200 rpm) at room temperature. Protein solution was removed and the beads were washed (6 × 200 µL) with PD washing buffer (20 mM HEPES, 300 mM KCl, 20 % glycerol, pH 7.9) by shaking at 1200 rpm for 2 min at room temperature. Proteins were eluted by incubation with 25 µL of 3x SDS sample buffer (5x SDS sample buffer: 250 mM Tris, 10 % SDS, 30 % glycerol, 0.5 M DTT, 0.02 % bromophenol blue) for 10 min at 95 °C. An input sample was prepared with 30 µg in 1x SDS sample buffer and heating for 10 min at 95 °C. Samples were analyzed by SDS-PAGE and western blotting (4.2.4.14).

4.2.4.26 Preparation of samples for proteomic experiments

Filter aided sample preparation (FASP)

Pull-down was performed as described (4.2.4.25) but elution was performed with 20 µL of 3x SDS sample buffer without bromophenol blue (150 mM Tris, 6 % SDS, 18 % glycerol, 0.3 M DTT). As input sample, 20 µg HEK293T whole cell extract was mixed with 5x SDS sample buffer without bromophenol blue and incubated at 95 °C for 10 min. All samples were processed by filter aided sample preparation (FASP).
Each sample was diluted with 450 µL U_A buffer (8 M in 0.1 M Tris-HCl, pH 8.5), transferred into a Microcon®-30 centrifugal filter device (Ultracel PL-30, Merck) and centrifuged (14,000 × g, 12 min). Samples were further diluted (3 × 450 µL) with U_A buffer, followed by centrifugation (14,000 × g, 12 min). 100 µL of 55 mM chloroacetamide (CAA) in U_A buffer were added and the mixtures were incubated for 20 min in the dark. Solution was removed by centrifugation (14,000 × g, 10 min), followed by washing (3 × 100 µL) with U_A buffer and centrifugation (14,000 × g, 8 min). For proteolytic digestion, stock solutions of Lysyl endopeptidase (LysC) (FUJIFILM Wako Chemicals) in 50 mM ammonium bicarbonate (ABC) buffer (0.5 µg/µL) and trypsin (MS approved, SERVA) in 1 mM HCl (1 µg/µL) were prepared. Filter units were transferred into new collection tubes and 1 µg LysC in 40 µL U_A buffer was added to each sample and incubated for 1 h at room

temperature. After adding additional 1 µg LysC, protein digestion was performed over night under shaking (400 rpm) at room temperature. Afterwards, the solution was diluted with 300 µL ABC buffer, 0.2 µg trypsin were added and the mixture was incubated for 3.5 h at 37 °C. Additional 0.2 µg trypsin were added and the incubation step was repeated. The filter units were centrifuged (14,000 × g, 10 min), 50 µL ABC buffer were added and centrifugation repeated. After adding TFA to a final concentration of 1 %, the eluates were further processed using StageTips.

Sample processing using StageTips

Digested peptide solutions were desalted and concentrated using stop-and-go-extraction tips (StageTips) [266]. Therefore, three layers of two different Empore™ solid phase extraction disks 47 mm (3M) were placed in 200 µL pipette tips. For input samples, SDB-XC (Polystyrenedivinylbenzene, product number: 2240) and for pull-down probes, C18 (Bonded Silica, product number: 2215) material was used. Depending on the material, different protocols were executed.
SDB-StageTips were equilibrated with 100 µL ACN, followed by 100 µL buffer C (30 % methanol, 1 % TFA in water) and 100 µL buffer D (0.2 % TFA in water). After addition of the solutions, tips were centrifuged (1300 × g, 2 min) using a tip-to-reaction tube adapter. Input samples were loaded on the tips by centrifugation at 500 × g for 3 min and 600 × g for 1 min and washed with 100 µL buffer D. After centrifugation (1300 × g, 2 min), samples were eluted in three steps. First elution was performed with 20 µL buffer 1 (100 mM NH_4HCO_2, 40 % ACN, 0.5 % FA in water), second with 20 µL buffer 2 (150 mM NH_4HCO_2, 60 % ACN, 0.5 % FA in water) and third with 20 µL of buffer 3 (5 % ammonia, 80 % ACN in water). Each step was centrifuged (500 × g, 5 min) into separate reaction tubes. The obtained fractions were diluted with water (20 µL) and lyophilized.
C18-StageTips were equilibrated with 60 µL MeOH, followed by 60 µL buffer B (80 % ACN, 0.5 % FA in water) and 60 µL buffer A (0.5 % FA in water). After addition of the solutions, tips were centrifuged (1300 × g, 2 min) using a tip-to-reaction tube adapter. Pull-Down samples were loaded on the tips by centrifugation at 600 × g for 3 min and 800 × g for 1 min and washed (2 × 30 µL) with buffer A and centrifuged (1300 × g, 2 min). Samples were eluted with 2 × 20 µL of buffer B and centrifugation (500 × g, 5 min).

4.2.4.27 Nano-LC-MS/MS measurements

Peptides from the digested pull-down eluates were separated on an EASY-nLC™ 1200 UHPLC system (Thermo Fisher Scientific) equipped with 250 mm frit-less silica emitters (New Objective, 0.75 µm inner diameter), packed in-house with reversed-phase ReproSil-Pur C18 AQ 1.9 µm resin (Dr. Maisch). The column temperature was maintained at 50 °C and coupled to an Orbitrap Exploris™ 480 mass spectrometer (Thermo Fisher Scientific). Peptides were eluted over 115 min applying a segmented linear gradient of 0-98 % LCMS-B at a flow rate of 300 nL/min.

Mass spectra were acquired in data-dependent acquisition mode. MS^1 scans were acquired at an Orbitrap resolution of 120,000 with a scan range of 380-1500 m/z, a maximum injection time of 100 ms and a normalized AGC target of 300 %. For fragmentation only precursors with charge states 2-6 were considered. Up to 20 dependent scans were taken. For dynamic exclusion, the exclusion duration was set to 40 s with a mass tolerance of +/- 10 ppm. The isolation window was set to 1.6 m/z with no offset. A normalized collision energy of 30 was used. MS^2 scans were taken at an Orbitrap resolution of 15,000 with a fixed first mass of 120 m/z. Maximum injection time was 22 ms and the normalized AGC Target 50 %.

4.2.4.28 LC-MS/MS data processing and quantification

Processing of raw data was performed using the MaxQuant software v1.6.17.0 [316] with default settings. MS/MS spectra were assigned to a FASTA file containing UniProt entries for *Homo Sapiens* (downloaded 2019-09-30). During search, sequences of 248 common contaminant proteins as well as decoy sequences were automatically added. Trypsin specificity was required and a maximum of two missed cleavages was allowed. Carbamidomethylation (CAM) of cysteine residues was set as fixed, oxidation of methionine and protein N-terminal acetylation as variable modifications. A false discovery rate of 1 % for peptide spectrum matches and proteins was applied. The following changes were made: Match between runs, LFQ and iBAQ options were enabled. Separate parameter groups were set for input and pull-down samples and the option for separate LFQ in parameter groups was enabled. Raw data processing was carried out by *Jürgen Eirich* (group of *Prof. Dr. Iris Finkemeier*, University of Münster).
Downstream data processing was carried out using the Perseus software platform version 1.6.15.0 [317]. Protein hits only identified by modified peptides, reversed sequences and potential contaminants were excluded and the LFQ intensities were $\log_2$ transformed. Hits were further filtered for detection in at least two replicates and missing LFQ values were imputed from normal distribution separately for each column (width: 0.4, down shift: 1.6). Significance of protein enrichment on AsuHd-containing H3 probes (peptides and NCPs) against unmodified H3 was determined by using the LIMMA package in R 4.1.0 [318]. Final volcano plots were generated with GraphPad Prism 9.0 by plotting the $\log_2$ fold change against the P values derived from LIMMA analysis.

References

1. Ezkurdia, I. *et al.* Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet* **23,** 5866–5878 (2014).
2. Piovesan, A. *et al.* On the length, weight and GC content of the human genome. *BMC Research Notes* **12,** 106 (2019).
3. Malik, H. S. & Henikoff, S. Phylogenomics of the nucleosome. *Nat Struct Biol* **10,** 882–891 (2003).
4. Henneman, B., van Emmerik, C., van Ingen, H. & Dame, R. T. Structure and function of archaeal histones. *PLoS Genet* **14,** e1007582 (2018).
5. Sandman, K., Pereira, S. L. & Reeve, J. N. Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cellular and Molecular Life Sciences CMLS* **54,** 1350–1364 (1998).
6. Kasinsky, H. E., Lewis, J. D., Dacks, J. B. & Ausió, J. Origin of H1 linker histones. *The FASEB Journal* **15,** 34–42 (2001).
7. Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184,** 868–871 (1974).
8. 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).
9. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences* (2015).
10. Zhu, P. & Li, G. Structural insights of nucleosome and the 30-nm chromatin fiber. *Current Opinion in Structural Biology* **36,** 106–115 (2016).
11. Hizume, K., Araki, S., Yoshikawa, K. & Takeyasu, K. Topoisomerase II, scaffold component, promotes chromatin compaction in vitro in a linker-histone H1-dependent manner. *Nucleic Acids Res* **35,** 2787–2799 (2007).
12. Saitoh, N., Goldberg, I. G., Wood, E. R. & Earnshaw, W. C. ScII: an abundant chromosome scaffold protein is a member of a family of putative ATPases with an unusual predicted tertiary structure. *J Cell Biol* **127,** 303–318 (1994).
13. Hennig, W. Heterochromatin. *Chromosoma* **108,** 1–9 (1999).
14. Liu, B., Yip, R. K. & Zhou, Z. Chromatin remodeling, DNA damage repair and aging. *Curr Genomics* **13,** 533–547 (2012).
15. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421,** 448–453 (2003).
16. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* **319,** 1097–1113 (2002).
17. Zheng, C. & Hayes, J. J. Structures and interactions of the core histone tail domains. *Biopolymers* **68,** 539–546 (2003).

18. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).
19. Rothbart, S. B. & Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta* **1839,** 627–643 (2014).
20. Rice, J. C. *et al.* Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Molecular Cell* **12,** 1591–1598 (2003).
21. Blanc, R. S. & Richard, S. Arginine Methylation: The Coming of Age. *Molecular Cell* **65,** 8–24 (2017).
22. Arnaudo, A. M. & Garcia, B. A. Proteomic characterization of novel histone post-translational modifications. *Epigenetics Chromatin* **6,** 24 (2013).
23. Rousseaux, S. & Khochbin, S. Histone acylation beyond acetylation: Terra Incognita in chromatin biology. *Cell J* **17,** 1–6 (2015).
24. Sabari, B. R., Zhang, D., Allis, C. D. & Zhao, Y. Metabolic regulation of gene expression through histone acylations. *Nat Rev Mol Cell Biol* **18,** 90–101 (2017).
25. Huang, H. *et al.* Lysine benzoylation is a histone mark regulated by SIRT2. *Nature Communications* **9,** 3374 (2018).
26. Xu, Y.-M., Du, J.-Y. & Lau, A. T. Y. Posttranslational modifications of human histone H3: An update. *PROTEOMICS* **14,** 2047–2060 (2014).
27. Sabari, B. R., Zhang, D., Allis, C. D. & Zhao, Y. Metabolic regulation of gene expression through histone acylations. *Nature Reviews Molecular Cell Biology* **18,** 90–101 (2017).
28. Barnes, C. E., English, D. M. & Cowley, S. M. Acetylation & Co: an expanding repertoire of histone acylations regulates chromatin and transcription. *Essays Biochem* **63,** 97–107 (2019).
29. Mersfelder, E. L. & Parthun, M. R. The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. *Nucleic Acids Res* **34,** 2653–2662 (2006).
30. Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146,** 1016–1028 (2011).
31. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Research* **21,** 381–395 (2011).
32. Lee, D. Y., Hayes, J. J., Pruss, D. & Wolffe, A. P. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* **72,** 73–84 (1993).
33. Garcia-Ramirez, M., Rocchini, C. & Ausio, J. Modulation of chromatin folding by histone acetylation. *J Biol Chem* **270,** 17923–17928 (1995).
34. Dorigo, B., Schalch, T., Bystricky, K. & Richmond, T. J. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *Journal of Molecular Biology* **327,** 85–96 (2003).
35. Dorigo, B. *et al.* Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* **306,** 1571–1573 (2004).

36. Fan, J. Y., Rangasamy, D., Luger, K. & Tremethick, D. J. H2A.Z alters the nucleosome surface to promote HP1α-mediated chromatin fiber folding. *Molecular Cell* **16,** 655–661 (2004).
37. Ruan, K. *et al.* Histone H4 acetylation required for chromatin decompaction during DNA replication. *Scientific Reports* **5,** 12720 (2015).
38. Lee, D. Y., Hayes, J. J., Pruss, D. & Wolffe, A. P. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* **72,** 73–84 (1993).
39. Vettese-Dadey, M. *et al.* Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *EMBO J* **15,** 2508–2518 (1996).
40. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–45 (2000).
41. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–1080 (2001).
42. Lo, W. S. *et al.* Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to GCN5-mediated acetylation at lysine 14. *Mol Cell* **5,** 917–926 (2000).
43. Kirmizis, A. *et al.* Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **449,** 928–932 (2007).
44. Fischle, W. *et al.* Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438,** 1116–1122 (2005).
45. Dupont, C., Armant, D. R. & Brenner, C. A. Epigenetics: definition, mechanisms and clinical perspective. *Semin Reprod Med* **27,** 351–357 (2009).
46. Wei, J.-W., Huang, K., Yang, C. & Kang, C.-S. Non-coding RNAs as regulators in epigenetics (Review). *Oncol Rep* **37,** 3–9 (2017).
47. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406,** 593–599 (2000).
48. Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* **12,** 1052–1058 (2002).
49. Kwon, T. *et al.* Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. *EMBO J* **22,** 292–303 (2003).
50. Klose, R. J. & Zhang, Y. Regulation of histone methylation by demethylimination and demethylation. *Nature Reviews Molecular Cell Biology* **8,** 307–318 (2007).
51. Rotili, D. & Mai, A. Targeting histone demethylases: a new avenue for the fight against cancer. *Genes Cancer* **2,** 663–679 (2011).
52. Marmorstein, R. & Trievel, R. C. Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim Biophys Acta* **1789,** 58–68 (2009).
53. Hirota, T., Lipp, J. J., Toh, B.-H. & Peters, J.-M. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature* **438,** 1176–1180 (2005).

54. Goto, H., Yasui, Y., Nigg, E. A. & Inagaki, M. Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation. *Genes Cells* **7,** 11–17 (2002).
55. Dai, J., Sultan, S., Taylor, S. S. & Higgins, J. M. G. The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. *Genes Dev* **19,** 472–488 (2005).
56. Gil, R. S. & Vagnarelli, P. Protein phosphatases in chromatin structure and function. *Biochim Biophys Acta Mol Cell Res* **1866,** 90–101 (2019).
57. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431,** 873–878 (2004).
58. Nathan, D., Sterner, D. E. & Berger, S. L. Histone modifications: now summoning sumoylation. *Proc Natl Acad Sci U S A* **100,** 13118–13120 (2003).
59. Zha, J.-J., Tang, Y. & Wang, Y.-L. Role of mono-ADP-ribosylation histone modification (Review). *Exp Ther Med* **21,** 577 (2021).
60. Biel, M., Wascholowski, V. & Giannis, A. Epigenetics – an epicenter of gene regulation: histones and histone-modifying enzymes. *Angewandte Chemie International Edition* **44,** 3186–3216 (2005).
61. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399,** 491–496 (1999).
62. Shiama, N. The p300/CBP family: integrating signals with transcription factors and chromatin. *Trends in Cell Biology* **7,** 230–236 (1997).
63. Lundblad, J. R., Kwok, R. P. S., Laurance, M. E., Harter, M. L. & Goodman, R. H. Adenoviral EIA-associated protein p300 as a functional homologue of the transcriptional co-activator CBP. *Nature* **374,** 85–88 (1995).
64. Roth, S. Y., Denu, J. M. & Allis, C. D. Histone acetyltransferases. *Annual Review of Biochemistry* **70,** 81–120 (2001).
65. Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90,** 595–606 (1997).
66. Reed, S. M. & Quelle, D. E. p53 acetylation: regulation and consequences. *Cancers* (*Basel*) **7,** 30–69 (2014).
67. Shida, T., Cueva, J. G., Xu, Z., Goodman, M. B. & Nachury, M. V. The major α-tubulin K40 acetyltransferase αTAT1 promotes rapid ciliogenesis and efficient mechanosensation. *Proceedings of the National Academy of Sciences* **107,** 21517–21522 (2010).
68. Janke, C. & Montagnac, G. Causes and consequences of microtubule acetylation. *Curr Biol* **27,** R1287–R1292 (2017).
69. Seto, E. & Yoshida, M. Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb Perspect Biol* **6,** a018713 (2014).
70. Park, S.-Y. & Kim, J.-S. A short guide to histone deacetylases including recent progress on class II enzymes. *Experimental & Molecular Medicine* **52,** 204–212 (2020).

71. Yang, X.-J. & Seto, E. The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat Rev Mol Cell Biol* **9,** 206–218 (2008).
72. Park, H., Kim, Y., Park, D. & Jeoung, D. Nuclear localization signal domain of HDAC3 is necessary and sufficient for the expression regulation of MDR1. *BMB Rep* **47,** 342–347 (2014).
73. Guo, X. *et al.* Subcellular localization of class I histone deacetylases in the developing Xenopus tectum. *Front Cell Neurosci* **9,** 510 (2015).
74. Waltregny, D. *et al.* Expression of histone deacetylase 8, a class I histone deacetylase, is restricted to cells showing smooth muscle differentiation in normal human tissues. *Am J Pathol* **165,** 553–564 (2004).
75. Moser, M. A., Hagelkruys, A. & Seiser, C. Transcription and beyond: the role of mammalian class I lysine deacetylases. *Chromosoma* **123,** 67–78 (2014).
76. Yang, W.-M., Tsai, S.-C., Wen, Y.-D., Fejér, G. & Seto, E. Functional domains of histone deacetylase-3. *Journal of Biological Chemistry* **277,** 9447–9454 (2002).
77. De Ruijter, A. J. M., van Gennip, A. H., Caron, H. N., Kemp, S. & van Kuilenburg, A. P. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J* **370,** 737–749 (2003).
78. Wolfson, N. A., Pitcairn, C. A. & Fierke, C. A. HDAC8 substrates: Histones and beyond. *Biopolymers* **99,** 112–126 (2013).
79. McKinsey, T. A., Zhang, C. L. & Olson, E. N. Identification of a signal-responsive nuclear export sequence in class II histone deacetylases. *Mol Cell Biol* **21,** 6312–6321 (2001).
80. Hubbert, C. *et al.* HDAC6 is a microtubule-associated deacetylase. *Nature* **417,** 455–458 (2002).
81. Yang, C.-J. *et al.* Nuclear HDAC6 inhibits invasion by suppressing NF-κB/MMP2 and is inversely correlated with metastasis of non-small cell lung cancer. *Oncotarget* **6,** 30263–30276 (2015).
82. Hudson, G. M., Watson, P. J., Fairall, L., Jamieson, A. G. & Schwabe, J. W. R. Insights into the recruitment of class IIa histone deacetylases (HDACs) to the SMRT/NCoR transcriptional repression complex. *J Biol Chem* **290,** 18237–18244 (2015).
83. Parra, M. Class IIa HDACs – new insights into their functions in physiology and pathology. *The FEBS Journal* **282,** 1736–1744 (2015).
84. Kao, H.-Y., Lee, C.-H., Komarov, A., Han, C. C. & Evans, R. M. Isolation and characterization of mammalian HDAC10, a novel histone deacetylase. *J Biol Chem* **277,** 187–193 (2002).
85. Tong, J. J., Liu, J., Bertos, N. R. & Yang, X.-J. Identification of HDAC10, a novel class II human histone deacetylase containing a leucine-rich domain. *Nucleic Acids Res* **30,** 1114–1123 (2002).
86. Hai, Y., Shinsky, S. A., Porter, N. J. & Christianson, D. W. Histone deacetylase 10 structure and molecular function as a polyamine deacetylase. *Nat Commun* **8,** 15368 (2017).

87. Alhazzazi, T. Y., Kamarajan, P., Verdin, E. & Kapila, Y. L. SIRT3 and cancer: tumor promoter or suppressor? *Biochim Biophys Acta* **1816,** 80–88 (2011).
88. Michishita, E., Park, J. Y., Burneskis, J. M., Barrett, J. C. & Horikawa, I. Evolutionarily conserved and nonconserved cellular localizations and functions of human SIRT proteins. *Mol Biol Cell* **16,** 4623–4635 (2005).
89. Rahman, S. & Islam, R. Mammalian Sirt1: insights on its biological functions. *Cell Commun Signal* **9,** 11 (2011).
90. Gao, L., Cueto, M. A., Asselbergs, F. & Atadja, P. Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. *J Biol Chem* **277,** 25748–25755 (2002).
91. Liu, S.-S., Wu, F., Jin, Y.-M., Chang, W. Q. & Xu, T.-M. HDAC11: a rising star in epigenetics. *Biomedicine & Pharmacotherapy* **131,** 110607 (2020).
92. Kutil, Z. *et al.* Histone deacetylase 11 Is a fatty-acid deacylase. *ACS Chemical Biology* **13,** 685–693 (2018).
93. Finnin, M. S. *et al.* Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **401,** 188–193 (1999).
94. Wu, R., Lu, Z., Cao, Z. & Zhang, Y. Zinc chelation with hydroxamate in histone deacetylases modulated by water access to the linker binding channel. *J Am Chem Soc* **133,** 6110–6113 (2011).
95. Lombardi, P. M., Cole, K. E., Dowling, D. P. & Christianson, D. W. Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. *Curr Opin Struct Biol* **21,** 735–743 (2011).
96. Delcuve, G. P., Khan, D. H. & Davie, J. R. Roles of histone deacetylases in epigenetic regulation: emerging paradigms from studies with inhibitors. *Clin Epigenetics* **4,** 5 (2012).
97. Guenther, M. G., Barak, O. & Lazar, M. A. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Mol Cell Biol* **21,** 6091–6101 (2001).
98. Watson, P. J. *et al.* Insights into the activation mechanism of class I HDAC complexes by inositol phosphates. *Nature Communications* **7,** 11262 (2016).
99. Battaglia, S., Maguire, O. & Campbell, M. J. Transcription factor co-repressors in cancer biology: roles and targeting. *Int J Cancer* **126,** 2511–2519 (2010).
100. Shi, Y. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422,** 735–738 (2003).
101. Lee, M. G., Wynder, C., Cooch, N. & Shiekhattar, R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437,** 432–435 (2005).
102. Milazzo, G. *et al.* Histone deacetylases (HDACs): evolution, specificity, role in transcriptional complexes, and pharmacological actionability. *Genes* (*Basel*) **11** (2020).
103. Hakimi, M.-A. *et al.* A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc Natl Acad Sci U S A* **99,** 7420–7425 (2002).

104. Zalloum, W. A. & Zalloum, H. M. Exploring the active center of the LSD1/CoREST complex by molecular dynamics simulation utilizing its co-crystallized co-factor tetrahydrofolate as a probe. *J Chem Inf Model* **57,** 3022–3031 (2017).
105. Yoshida, M., Hayakawa, T. & Nakayama, J. Physiological roles of class I HDAC complex and histone demethylase. *Journal of Biomedicine and Biotechnology* **2011,** 129383 (2011).
106. Kelly, R. & Cowley, S. The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. *Biochemical Society transactions* **41,** 741–9 (2013).
107. Bantscheff, M. *et al.* Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nature Biotechnology* **29,** 255–265 (2011).
108. Hahnen, E. *et al.* Histone deacetylase inhibitors: possible implications for neurodegenerative disorders. *Expert Opin Investig Drugs* **17,** 169–184 (2008).
109. Duvic, M. & Zhang, C. Clinical and laboratory experience of vorinostat (suberoylanilide hydroxamic acid) in the treatment of cutaneous T-cell lymphoma. *British Journal of Cancer* **95,** S13–S19 (2006).
110. Bug, G. *et al.* Clinical trial of valproic acid and all-trans retinoic acid in patients with poor-risk acute myeloid leukemia. *Cancer* **104,** 2717–2725 (2005).
111. Eckschlager, T., Plch, J., Stiborova, M. & Hrabeta, J. Histone deacetylase inhibitors as anticancer drugs. *Int J Mol Sci* **18** (2017).
112. Boffa, L., Vidali, G., Mann, R. & Allfrey, V. Suppression of histone deacetylation in vivo and in vitro by sodium butyrate. *Journal of Biological Chemistry* **253,** 3364–3366 (1978).
113. Drummond, D. C. *et al.* Clinical development of histone deacetylase inhibitors as anticancer agents. *Annual Review of Pharmacology and Toxicology* **45,** 495–528 (2005).
114. Gore, S. D. & Carducci, M. A. Modifying histones to tame cancer: clinical development of sodium phenylbutyrate and other histone deacetylase inhibitors. *Expert Opinion on Investigational Drugs* **9,** 2923–2934 (2000).
115. Romoli, M. *et al.* Valproic acid and epilepsy: from molecular mechanisms to clinical evidences. *Curr Neuropharmacol* **17,** 926–946 (2019).
116. Finnin, M. S. *et al.* Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **401,** 188–193 (1999).
117. Tsuji, N., Kobayashi, M., Nagashima, K., Wakisaka, Y. & Koizumi, K. A new antifungal antibiotic, trichostatin. *J Antibiot* (*Tokyo*) **29,** 1–6 (1976).
118. Duvic, M. *et al.* Phase 2 trial of oral vorinostat (suberoylanilide hydroxamic acid, SAHA) for refractory cutaneous T-cell lymphoma (CTCL). *Blood* **109,** 31–39 (2007).
119. Saito, A. *et al.* A synthetic inhibitor of histone deacetylase, MS-27-275, with marked in vivo antitumor activity against human tumors. *Proceedings of the National Academy of Sciences* **96,** 4592–4597 (1999).
120. Suzuki, T. *et al.* Synthesis and histone deacetylase inhibitory activity of new benzamide derivatives. *Journal of Medicinal Chemistry* **42,** 3001–3003 (1999).

121. Fournel, M. *et al.* MGCD0103, a novel isotype-selective histone deacetylase inhibitor, has broad spectrum antitumor activity in vitro and in vivo. *Mol Cancer Ther* **7,** 759–768 (2008).
122. Lane, A. A. & Chabner, B. A. Histone deacetylase inhibitors in cancer therapy. *Journal of Clinical Oncology* **27,** 5459–5468 (2009).
123. Furumai, R. *et al.* FK228 (depsipeptide) as a natural prodrug that inhibits class I histone deacetylases. *Cancer Research* **62,** 4916–4921 (2002).
124. Miller, T. A., Witter, D. J. & Belvedere, S. Histone deacetylase inhibitors. *Journal of Medicinal Chemistry* **46,** 5097–5116 (2003).
125. Coiffier, B. *et al.* Romidepsin for the treatment of relapsed/refractory peripheral T-cell lymphoma: pivotal study update demonstrates durable responses. *Journal of Hematology & Oncology* **7,** 11 (2014).
126. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Research* **21,** 564–578 (2011).
127. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14,** 1025–1040 (2007).
128. Musselman, C. A., Lalonde, M.-E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19,** 1218–1227 (2012).
129. Musselman, C. A., Khorasanizadeh, S. & Kutateladze, T. G. Towards understanding methyllysine readout. *Biochim Biophys Acta* **1839,** 686–693 (2014).
130. Rona, G. B., Eleutherio, E. C. A. & Pinheiro, A. S. PWWP domains and their modes of sensing DNA and histone methylated lysines. *Biophys Rev* **8,** 63–74 (2016).
131. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–124 (2001).
132. Min, J., Zhang, Y. & Xu, R.-M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* **17,** 1823–1828 (2003).
133. Musselman, C. A. *et al.* Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nat Struct Mol Biol* **19,** 1266–1272 (2012).
134. Wang, Y. *et al.* Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Mol Cell* **33,** 428–437 (2009).
135. Yang, Y. *et al.* TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol Cell* **40,** 1016–1023 (2010).
136. Migliori, V. *et al.* Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat Struct Mol Biol* **19,** 136–144 (2012).
137. Zhao, Q. *et al.* PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* **16,** 304–311 (2009).
138. Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123,** 1213–1226 (2005).

139. Macdonald, N. *et al.* Molecular basis for the recognition of phosphorylated and phosphoacetylated histone H3 by 14-3-3. *Molecular Cell* **20,** 199–211 (2005).
140. Ishiyama, S. *et al.* Structure of the Dnmt1 reader module complexed with a unique two-mono-ubiquitin mark on histone H3 reveals the basis for DNA methylation maintenance. *Molecular Cell* **68,** 350–360.e7 (2017).
141. Zeng, L. *et al.* Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466,** 258–262 (2010).
142. Le Masson, I. *et al.* Yaf9, a novel NuA4 histone acetyltransferase subunit, is required for the cellular response to spindle stress in yeast. *Mol Cell Biol* **23,** 6086–6102 (2003).
143. Schulze, J. M., Wang, A. Y. & Kobor, M. S. YEATS domain proteins: a diverse family with many links to chromatin modification and transcription. *Biochem Cell Biol* **87,** 65–75 (2009).
144. Daser, A. & Rabbitts, T. H. Extending the repertoire of the mixed-lineage leukemia gene MLL in leukemogenesis. *Genes Dev* **18,** 965–974 (2004).
145. Zhao, D., Li, Y., Xiong, X., Chen, Z. & Li, H. YEATS domain - a histone acylation reader in health and disease. *Journal of Molecular Biology* **429,** 1994–2002 (2017).
146. Tamkun, J. W. *et al.* brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2SWI2. *Cell* **68,** 561–572 (1992).
147. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. *FEBS Letters* **586,** 2692–2704 (2012).
148. Fujisawa, T. & Filippakopoulos, P. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nature Reviews Molecular Cell Biology* **18,** 246–262 (2017).
149. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149,** 214–231 (2012).
150. Owen, D. J. *et al.* The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase GCN5p. *EMBO J* **19,** 6141–6149 (2000).
151. Miller, T. C. R. *et al.* A bromodomain–DNA interaction facilitates acetylation-dependent bivalent nucleosome recognition by the BET protein BRDT. *Nature Communications* **7,** 13855 (2016).
152. Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. *J Mol Biol* **304,** 355–370 (2000).
153. Hassan, A. H. *et al.* Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. *Biochem J* **402,** 125–133 (2007).
154. Mujtaba, S. *et al.* Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol Cell* **13,** 251–263 (2004).
155. Wei, L., Jamonnak, N., Choy, J., Wang, Z. & Zheng, W. Differential binding modes of the bromodomains of CREB-binding protein (CBP) and p300 with acetylated MyoD. *Biochem Biophys Res Commun* **368,** 279–284 (2008).

156. Wu, Q. *et al.* A chemical toolbox for the study of bromodomains and epigenetic signaling. *Nature Communications* **10,** 1915 (2019).
157. Lalonde, M.-E. *et al.* Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. *Genes Dev* **27,** 2009–2024 (2013).
158. Laue, K. *et al.* The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity. *Development* **135,** 1935–1946 (2008).
159. You, L. *et al.* The chromatin regulator Brpf1 regulates embryo development and cell proliferation. *J Biol Chem* **290,** 11349–11364 (2015).
160. Santoro, R., Li, J. & Grummt, I. The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nat Genet* **32,** 393–396 (2002).
161. Bortoluzzi, A., Amato, A., Lucas, X., Blank, M. & Ciulli, A. Structural basis of molecular recognition of helical histone H3 tail by PHD finger domains. *Biochem J* **474,** 1633–1651 (2017).
162. Zeng, L. *et al.* Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat Struct Mol Biol* **15,** 626–633 (2008).
163. Ali, M., Hom, R. A., Blakeslee, W., Ikenouye, L. & Kutateladze, T. G. Diverse functions of PHD fingers of the MLL/KMT2 subfamily. *Biochim Biophys Acta* **1843,** 366–371 (2014).
164. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and function of a human TAFII250 double bromodomain module. *Science* **288,** 1422–1425 (2000).
165. Xu, Y. *et al.* TAF1 plays a critical role in AML1-ETO driven leukemogenesis. *Nature Communications* **10,** 4925 (2019).
166. Li, A. G. *et al.* An acetylation switch in p53 mediates holo-TFIID recruitment. *Mol Cell* **28,** 408–421 (2007).
167. Yang, Z., He, N. & Zhou, Q. Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. *Mol Cell Biol* **28,** 967–976 (2008).
168. Itzen, F., Greifenberg, A. K., Bösken, C. A. & Geyer, M. Brd4 activates P-TEFb for RNA polymerase II CTD phosphorylation. *Nucleic Acids Res* **42,** 7577–7590 (2014).
169. Rahman, S. *et al.* The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol Cell Biol* **31,** 2641–2652 (2011).
170. Morgado-Pascual, J. L., Rayego-Mateos, S., Tejedor, L., Suarez-Alvarez, B. & Ruiz-Ortega, M. Bromodomain and extraterminal proteins as novel epigenetic targets for renal diseases. *Frontiers in Pharmacology* **10,** 1315 (2019).
171. Taniguchi, Y. The Bromodomain and Extra-Terminal Domain (BET) Family: Functional Anatomy of BET Paralogous Proteins. *Int J Mol Sci* **17** (2016).
172. Wang, N., Wu, R., Tang, D. & Kang, R. The BET family in immunity and disease. *Signal Transduction and Targeted Therapy* **6,** 23 (2021).
173. Belkina, A. C. & Denis, G. V. BET domain co-regulators in obesity, inflammation and cancer. *Nat Rev Cancer* **12,** 465–477 (2012).

174. French, C. A. *et al.* BRD4 bromodomain gene rearrangement in aggressive carcinoma with translocation t(15;19). *Am J Pathol* **159,** 1987–1992 (2001).
175. French, C. A. NUT midline carcinoma. *Cancer Genet Cytogenet* **203,** 16–20 (2010).
176. Alekseyenko, A. A. *et al.* The oncogenic BRD4-NUT chromatin regulator drives aberrant transcription within large topological domains. *Genes Dev* **29,** 1507–1523 (2015).
177. Grayson, A. R. *et al.* MYC, a downstream target of BRD-NUT, is necessary and sufficient for the blockade of differentiation in NUT midline carcinoma. *Oncogene* **33,** 1736–1742 (2014).
178. Reynoird, N. *et al.* Oncogenesis by sequestration of CBP/p300 in transcriptionally inactive hyperacetylated chromatin domains. *The EMBO Journal* **29,** 2943–2952 (2010).
179. Brennan, P., Filippakopoulos, P. & Knapp, S. The therapeutic potential of acetyl-lysine and methyl-lysine effector domains. *Drug Discovery Today: Therapeutic Strategies* **9,** e101–e110 (2012).
180. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–1073 (2010).
181. Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468,** 1119–1123 (2010).
182. Ferri, E., Petosa, C. & McKenna, C. E. Bromodomains: Structure, function and pharmacology of inhibition. *Biochemical Pharmacology* **106,** 1–18 (2016).
183. Pérez-Salvia, M. & Esteller, M. Bromodomain inhibitors and cancer therapy: from structures to applications. *Epigenetics* **12,** 323–339 (2017).
184. Shorstova, T., Foulkes, W. D. & Witcher, M. Achieving clinical success with BET inhibitors as anti-cancer agents. *British Journal of Cancer* **124,** 1478–1490 (2021).
185. Odore, E. *et al.* Phase I population pharmacokinetic assessment of the oral bromodomain inhibitor OTX015 in patients with haematologic malignancies. *Clin Pharmacokinet* **55,** 397–405 (2016).
186. Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468,** 1119–1123 (2010).
187. Gilan, O. *et al.* Selective targeting of BD1 and BD2 of the BET proteins in cancer and immunoinflammation. *Science* **368,** 387–394 (2020).
188. McLure, K. G. *et al.* RVX-208, an inducer of ApoA-I in humans, is a BET bromodomain antagonist. *PLoS One* **8,** e83190 (2013).
189. Faivre, E. J. *et al.* Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. *Nature* **578,** 306–310 (2020).
190. Slavish, P. J. *et al.* Bromodomain-selective BET inhibitors are potent antitumor agents against MYC-driven pediatric cancer. *Cancer Res* **80,** 3507–3518 (2020).
191. Picaud, S. *et al.* Promiscuous targeting of bromodomains by bromosporine identifies BET proteins as master regulators of primary transcription response in leukemia. *Science Advances* **2** (2016).

192. Picaud, S. *et al.* Generation of a selective small molecule inhibitor of the CBP/p300 bromodomain for leukemia therapy. *Cancer Res* **75,** 5106–5119 (2015).
193. Hay, D. A. *et al.* Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains. *J Am Chem Soc* **136,** 9308–9319 (2014).
194. Clark, P. G. K. *et al.* LP99: discovery and synthesis of the first selective BRD7/9 bromodomain inhibitor. *Angew Chem Weinheim Bergstr Ger* **127,** 6315–6319 (2015).
195. Theodoulou, N. H. *et al.* Discovery of I-BRD9, a selective cell active chemical probe for bromodomain containing protein 9 inhibition. *J Med Chem* **59,** 1425–1439 (2016).
196. Chen, P. *et al.* Discovery and characterization of GSK2801, a selective chemical probe for the bromodomains BAZ2A and BAZ2B. *J Med Chem* **59,** 1410–1424 (2016).
197. Yang, A., Cho, K. & Park, H.-S. Chemical biology approaches for studying post-translational modifications. *RNA Biol* **15,** 427–440 (2018).
198. Conibear, A. C. Deciphering protein post-translational modifications using chemical biology tools. *Nature Reviews Chemistry* **4,** 674–695 (2020).
199. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266,** 776–779 (1994).
200. Wan, Q. & Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angewandte Chemie International Edition* **46,** 9248–9252 (2007).
201. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proceedings of the National Academy of Sciences* **95,** 6705–6710 (1998).
202. Hackenberger, C. P. R. & Schwarzer, D. Chemoselective ligation and modification strategies for peptides and proteins. *Angewandte Chemie International Edition* **47,** 10030–10074 (2008).
203. Wu, H., Hu, Z. & Liu, X. Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc Natl Acad Sci U S A* **95,** 9226–9231 (1998).
204. Noren, C. J., Wang, J. & Perler, F. B. Dissecting the chemistry of protein splicing and its applications. *Angewandte Chemie International Edition* **39,** 450–466 (2000).
205. Mazmanian, S. K., Liu, G., Ton-That, H. & Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285,** 760–763 (1999).
206. Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc Natl Acad Sci U S A* **96,** 12424–12429 (1999).
207. Aparicio Pelaz, D. *et al.* Examining histone modification crosstalk using immobilized libraries established from ligation-ready nucleosomes. *Chem. Sci.* **11,** 9218–9225 (2020).

208. Aparicio Pelaz, D. *Establishing a chemical toolbox for designer nucleosomes and its applications for analysis of chromatin binding proteins.* PhD Thesis (Eberhard Karls Universität Tübingen, 2017).
209. Johansson, L., Gafvelin, G. & Arnér, E. S. J. Selenocysteine in proteins-properties and biotechnological use. *Biochim Biophys Acta* **1726,** 1–13 (2005).
210. Hao, B. *et al.* A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296,** 1462–1466 (2002).
211. Blight, S. K. *et al.* Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* **431,** 333–335 (2004).
212. Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc Natl Acad Sci U S A* **104,** 11268–11273 (2007).
213. Nguyen, D. P., Alai, M. M. G., Virdee, S. & Chin, J. W. Genetically directing ε-N, N-dimethyl-l-lysine in recombinant histones. *Chemistry & Biology* **17,** 1072–1076 (2010).
214. Yanagisawa, T. *et al.* Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem Biol* **15,** 1187–1197 (2008).
215. Saleh, A. M., Wilding, K. M., Calve, S., Bundy, B. C. & Kinzer-Ursem, T. L. Non-canonical amino acid labeling in proteomics and biotechnology. *J Biol Eng* **13,** 43 (2019).
216. Wang, Z. A. *et al.* A versatile approach for site-specific lysine acylation in proteins. *Angewandte Chemie International Edition* **56,** 1643–1647 (2017).
217. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding Nε-acetyllysine in recombinant proteins. *Nature Chemical Biology* **4,** 232–234 (2008).
218. Rogerson, D. T. *et al.* Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nature Chemical Biology* **11,** 496–503 (2015).
219. De la Torre, D. & Chin, J. W. Reprogramming the genetic code. *Nature Reviews Genetics* **22,** 169–184 (2021).
220. Burke, T. R., Smyth, M. S., Nomizu, M., Otaka, A. & Roller, P. R. Preparation of fluoro- and hydroxy-4-(phosphonomethyl)-D,L-phenylalanine suitably protected for solid-phase synthesis of peptides containing hydrolytically stable analogs of O-phosphotyrosine. *The Journal of Organic Chemistry* **58,** 1336–1340 (1993).
221. Liu, F., Park, J.-E., Lee, K. S. & Burke, T. R. J. Preparation of orthogonally protected (2S, 3R)-2-amino-3-methyl-4-phosphonobutyric acid (Pmab) as a phosphatase-stable phosphothreonine mimetic and its use in the synthesis of Polo-box domain-binding peptides. *Tetrahedron* **65,** 9673–9679 (2009).
222. Fujimoto, H. *et al.* A possible overestimation of the effect of acetylation on lysine residues in KQ mutant analysis. *J Comput Chem* **33,** 239–246 (2012).
223. Kamieniarz, K. & Schneider, R. Tools to Tackle Protein Acetylation. *Chemistry & Biology* **16,** 1027–1029 (2009).

224. Guo, J., Wang, J., Lee, J. S. & Schultz, P. G. Site-Specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angewandte Chemie International Edition* **47,** 6399–6401 (2008).
225. Dose, A. *et al.* Interrogating substrate selectivity and composition of endogenous histone deacetylase complexes with chemical probes. *Angew Chem Int Ed Engl* **55,** 1192–1195 (2016).
226. Dose, A. *Untersuchung von Histon-Deacetylasen mittels chemischer Analysestrategien.* PhD Thesis (Eberhard Karls Universität Tübingen, 2016).
227. Sindlinger, J. *Synthesis and application of chemical probes to investigate substrate selectivity of histone deacetylase complexes on a proteome-wide level.* PhD Thesis (Eberhard Karls Universität Tübingen, 2018).
228. Kirchgäßner, S. *Synthesis and characterization of peptide-based probes for bromodomains.* Master Thesis (Eberhard Karls Universität Tübingen, 2016).
229. Schwarzer, D., Seidel, J. S., Kirchgäßner, S. B. & Bierlmeier, J. P. Acetyl-lysine mimicry, as well as its manufacture and use. *DE Patent,* Eberhard Karls Universität Tübingen, DE102018214919B4 (2018).
230. Dent, W. H., Erickson, W. R., Fields, S. C., Parker, M. H. & Tromiczak, E. G. 9-BBN: an amino acid protecting group for functionalization of amino acid side chains in organic solvents. *Organic Letters* **4,** 1249–1251 (2002).
231. Castanedo, G. M., Seng, P. S., Blaquiere, N., Trapp, S. & Staben, S. T. Rapid synthesis of 1,3,5-substituted 1,2,4-triazoles from carboxylic acids, amidines, and hydrazines. *The Journal of Organic Chemistry* **76,** 1177–1179 (2011).
232. Staben, S. T. & Blaquiere, N. Four-component synthesis of fully substituted 1,2,4-triazoles. *Angewandte Chemie International Edition* **49,** 325–328 (2010).
233. Bommarius, A. S., Drauz, K., Klenk, H. & Wandrey, C. Operational stability of enzymes. Acylase-catalyzed resolution of N-acetyl amino acids to enantiomerically pure L-amino acids. *Ann N Y Acad Sci* **672,** 126–136 (1992).
234. Uttamsingh, V., Keller, D. A. & Anders, M. W. Acylase I-catalyzed deacetylation of N-acetyl-l-cysteine and S-alkyl-N-acetyl-l-cysteines. *Chemical Research in Toxicology* **11,** 800–809 (1998).
235. Kühn, A. *Genetic encoding of HDAC-trapping amino acids.* Master Thesis (Eberhard Karls Universität Tübingen, 2020).
236. Seidel, J. *Development of enzyme-specific chemical probes for high-throughput substrate profiling of histone deacetylases.* PhD Thesis (Eberhard Karls Universität Tübingen, 2021).
237. Sánchez, A., Calderón, E. & Vazquez, A. Using the 9-BBN group as a transient protective group for the functionalization of reactive chains of α-amino acids. *Synthesis* **45,** 1364–1372 (2013).
238. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. *FEBS Lett* **586,** 2692–2704 (2012).
239. Scheffler, M. *Study of interaction between bromodomains and histone-tail mimicking peptides.* Bachelor Thesis (Eberhard Karls Universität Tübingen, 2018).

240. Chhabra, S. R. *et al.* An appraisal of new variants of Dde amine protecting group for solid phase peptide synthesis. *Tetrahedron Letters* **39,** 1603–1606 (1998).
241. Díaz-Mochón, J. J., Bialy, L. & Bradley, M. Full orthogonality between Dde and Fmoc: the direct synthesis of PNA-peptide conjugates. *Organic Letters* **6,** 1127–1129 (2004).
242. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149,** 214–231 (2012).
243. Subach, O. M. *et al.* Conversion of red fluorescent protein into a bright blue probe. *Chem Biol* **15,** 1116–1124 (2008).
244. Shagin, D. A. *et al.* GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Mol Biol Evol* **21,** 841–850 (2004).
245. Bartoschik, T. *et al.* Near-native, site-specific and purification-free protein labeling for quantitative protein interaction analysis by MicroScale Thermophoresis. *Scientific Reports* **8** (2018).
246. Jost, J. O., Hanswillemenke, A. & Schwarzer, D. A miniaturized readout strategy for endogenous histone deacetylase activity. *Mol. BioSyst.* **11,** 1820–1823 (7 2015).
247. Filonov, G. S. *et al.* Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nature Biotechnology* **29,** 757–761 (2011).
248. Nikić, I. *et al.* Debugging eukaryotic genetic code expansion for site-specific Click-PAINT super-resolution microscopy. *Angewandte Chemie International Edition* **55,** 16172–16176 (2016).
249. Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E. A. Genetically encoded copper-free click chemistry. *Angewandte Chemie International Edition* **50,** 3878–3881 (2011).
250. Markwardt, M. L. *et al.* An improved cerulean fluorescent protein with enhanced brightness and reduced reversible photoswitching. *PLOS ONE* **6,** 1–11 (2011).
251. Ormö, M. *et al.* Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273,** 1392 (1996).
252. Luo, J., Su, F., Chen, D., Shiloh, A. & Gu, W. Deacetylation of p53 modulates its effect on cell growth and apoptosis. *Nature* **408,** 377–381 (2000).
253. Hubbert, C. *et al.* HDAC6 is a microtubule-associated deacetylase. *Nature* **417,** 455–458 (2002).
254. Bajar, B. T. *et al.* Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Scientific Reports* **6,** 20889 (2016).
255. Wen, W., Meinkoth, J. L., Tsien, R. Y. & Taylor, S. S. Identification of a signal for rapid export of proteins from the nucleus. *Cell* **82,** 463–473 (1995).
256. Bajar, B. T., Wang, E. S., Zhang, S., Lin, M. Z. & Chu, J. A guide to fluorescent protein FRET pairs. *Sensors* (*Basel*) **16** (2016).
257. Schmidt, T. G. *et al.* Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expression and Purification* **92,** 54–61 (2013).

258. Ooi, L. & Wood, I. C. Chromatin crosstalk in development and disease: lessons from REST. *Nature Reviews Genetics* **8,** 544–554 (2007).
259. 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. *Journal of Biological Chemistry* **282,** 4408–4416 (2007).
260. Mahler, H. *Development of a 'tool box' for generating designer nucleosomes in high throughput fashion.* PhD Thesis (Georg-August-Universität Göttingen, 2016).
261. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276,** 19–42 (1998).
262. Williamson, D. J., Fascione, M. A., Webb, M. E. & Turnbull, W. B. Efficient N-terminal labeling of proteins by use of sortase. *Angew Chem Int Ed Engl* **51,** 9377–9380 (2012).
263. Eot-Houllier, G., Fulcrand, G., Watanabe, Y., Magnaghi-Jaulin, L. & Jaulin, C. Histone deacetylase 3 is required for centromeric H3K4 deacetylation and sister chromatid cohesion. *Genes Dev* **22,** 2639–2644 (2008).
264. Park, J.-A. *et al.* Deacetylation and methylation at histone H3 lysine 9 (H3K9) coordinate chromosome condensation during cell cycle progression. *Mol Cells* **31,** 343–349 (2011).
265. Li, F. *et al.* Histone deacetylase 1 (HDAC1) negatively regulates thermogenic program in brown adipocytes via coordinated regulation of histone H3 lysine 27 (H3K27) deacetylation and methylation. *J Biol Chem* **291,** 4523–4536 (2016).
266. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2,** 1896–1906 (2007).
267. Flynn, E. M. *et al.* A subset of human bromodomains recognizes butyryllysine and crotonyllysine histone peptide modifications. *Structure* **23,** 1801–1814 (2015).
268. Ren, X. *et al.* Histone benzoylation serves as an epigenetic mark for DPF and YEATS family proteins. *Nucleic Acids Research* **49,** 114–126 (2020).
269. Morinière, J. *et al.* Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* **461,** 664–668 (2009).
270. Marks, P. A. Discovery and development of SAHA as an anticancer agent. *Oncogene* **26,** 1351–1356 (2007).
271. Jamonnak, N., Fatkins, D. G., Wei, L. & Zheng, W. Nε-Methanesulfonyl-lysine as a non-hydrolyzable functional surrogate for Nε-acetyl-lysine. *Org. Biomol. Chem.* **5,** 892–896 (2007).
272. Huang, R. *et al.* Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation. *J Am Chem Soc* **132,** 9986–9987 (2010).
273. Sekirnik Née Measures, A. R. *et al.* Isoxazole-derived amino acids are bromodomain-binding acetyl-lysine mimics: incorporation into histone H4 peptides and histone H3. *Angew Chem Int Ed Engl* **55,** 8353–8357 (2016).

274. Schiedel, M. *et al.* Chemical Epigenetics: The Impact of Chemical and Chemical Biology Techniques on Bromodomain Target Validation. *Angewandte Chemie International Edition* **58,** 17930–17952 (2019).
275. Moreno-Yruela, C. *et al.* Hydroxamic acid-modified peptide microarrays for profiling isozyme-selective interactions and inhibition of histone deacetylases. *Nature Communications* **12,** 62 (2021).
276. Seidel, J. *et al.* Peptide-based 2-aminophenylamide probes for targeting endogenous class I histone deacetylase complexes. *Chembiochem* **20,** 3001–3005 (2019).
277. Morrison, E. A., Bowerman, S., Sylvers, K. L., Wereszczynski, J. & Musselman, C. A. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife* **7** (2018).
278. Vermeulen, M. *et al.* In vitro targeting reveals intrinsic histone tail specificity of the Sin3/histone deacetylase and N-CoR/SMRT corepressor complexes. *Mol Cell Biol* **24,** 2364–2372 (2004).
279. Yamane, K. *et al.* JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125,** 483–495 (2006).
280. London, R. E. The structural basis of XRCC1-mediated DNA repair. *DNA Repair* (*Amst*) **30,** 90–103 (2015).
281. Walker, J. R., Corpina, R. A. & Goldberg, J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412,** 607–614 (2001).
282. Li, B. & Comai, L. Functional interaction between Ku and the werner syndrome protein in DNA end processing. *J Biol Chem* **275,** 28349–28352 (2000).
283. Kershnar, E., Wu, S. Y. & Chiang, C. M. Immunoaffinity purification and functional characterization of human transcription factor IIH and RNA polymerase II from clonal cell lines that conditionally express epitope-tagged subunits of the multiprotein complexes. *J Biol Chem* **273,** 34444–34453 (1998).
284. Greber, B. J., Toso, D. B., Fang, J. & Nogales, E. The complete structure of the human TFIIH core complex. *Elife* **8** (2019).
285. Araki, M. *et al.* Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *J Biol Chem* **276,** 18665–18672 (2001).
286. Lukas, J., Lukas, C. & Bartek, J. More than just a focus: the chromatin response to DNA damage and its role in genome integrity maintenance. *Nature Cell Biology* **13,** 1161–1169 (2011).
287. Misteli, T. & Soutoglou, E. The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nature Reviews Molecular Cell Biology* **10,** 243–254 (2009).
288. Miller, K. M. *et al.* Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. *Nat Struct Mol Biol* **17,** 1144–1151 (2010).

289. Umlauf, D. *et al.* The human TREX-2 complex is stably associated with the nuclear pore basket. *Journal of Cell Science* **126,** 2656–2667 (2013).
290. Kehat, I., Accornero, F., Aronow, B. J. & Molkentin, J. D. Modulation of chromatin position and gene expression by HDAC4 interaction with nucleoporins. *J Cell Biol* **193,** 21–29 (2011).
291. Bai, X.-T. *et al.* Trans-repressive effect of NUP98-PMX1 on PMX1-regulated c-FOS gene through recruitment of histone deacetylase 1 by FG repeats. *Cancer Research* **66,** 4584–4590 (2006).
292. Tsukamoto, T. *et al.* Cloning and characterization of two human cDNAs encoding the mRNA capping enzyme. *Biochem Biophys Res Commun* **243,** 101–108 (1998).
293. Bertram, K. *et al.* Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. *Cell* **170,** 701–713.e11 (2017).
294. Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structure of a human catalytic step I spliceosome. *Science* **359,** 537–545 (2018).
295. Januszyk, K. & Lima, C. D. The eukaryotic RNA exosome. *Curr Opin Struct Biol* **24,** 132–140 (2014).
296. Rahhal, R. & Seto, E. Emerging roles of histone modifications and HDACs in RNA splicing. *Nucleic Acids Research* **47,** 4911–4926 (2019).
297. Coletta, R. D. *et al.* The Six1 homeoprotein stimulates tumorigenesis by reactivation of cyclin A1. *Proc Natl Acad Sci U S A* **101,** 6478–6483 (2004).
298. Wu, K. *et al.* Cell fate determination factor DACH1 inhibits c-Jun-induced contact-independent growth. *Mol Biol Cell* **18,** 755–767 (2007).
299. Kida, Y., Maeda, Y., Shiraishi, T., Suzuki, T. & Ogura, T. Chick Dach1 interacts with the Smad complex and Sin3a to control AER formation and limb development along the proximodistal axis. *Development* **131,** 4179–4187 (2004).
300. Frank, S. R. *et al.* MYC recruits the TIP60 histone acetyltransferase complex to chromatin. *EMBO Rep* **4,** 575–580 (2003).
301. Kwon, H. & Green, M. R. The RNA polymerase I transcription factor, upstream binding factor, interacts directly with the TATA box-binding protein. *J Biol Chem* **269,** 30140–30146 (1994).
302. Comai, L. *et al.* Reconstitution of transcription factor SL1: exclusive binding of TBP by SL1 or TFIID subunits. *Science* **266,** 1966–1972 (1994).
303. Sharma, A. L. *et al.* CBF-1 promotes the establishment and maintenance of HIV latency by recruiting Polycomb repressive complexes, PRC1 and PRC2, at HIV LTR. *Viruses* **12** (2020).
304. Marfella, C. G. A. & Imbalzano, A. N. The Chd family of chromatin remodelers. *Mutat Res* **618,** 30–40 (2007).
305. Newton, A. H. & Pask, A. J. CHD9 upregulates RUNX2 and has a potential role in skeletal evolution. *BMC Molecular and Cell Biology* **21,** 27 (2020).
306. Seidel, J. *Development of a hydroxamic acid building block for high-throughput synthesis of peptide-based HDAC affinity probes.* Master Thesis (Eberhard Karls Universität Tübingen, 2017).

307. Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *Journal of the American Chemical Society* **85,** 2149–2154 (1963).
308. HUFFMAN, C. W. Formylation of amines. *The Journal of Organic Chemistry* **23,** 727–729 (May 1958).
309. Neises, B. & Steglich, W. Simple method for the esterification of carboxylic acids. *Angewandte Chemie International Edition in English* **17,** 522–524 (1978).
310. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66,** 133–144 (2010).
311. Diederichs, K. Some aspects of quantitative analysis and correction of radiation damage. *Acta Crystallogr D Biol Crystallogr* **62,** 96–101 (2006).
312. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63,** 32–41 (2007).
313. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66,** 213–221 (2010).
314. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66,** 486–501 (2010).
315. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research* **11,** 1475–1489 (1983).
316. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26,** 1367–1372 (2008).
317. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13,** 731–740 (2016).
318. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43,** e47–e47 (2015).

Acknowledgements

For me, it is a pleasure to thank the people who made this PhD thesis possible. To begin with, this work would not have been possible without my supervisor Prof. Dr. Dirk Schwarzer, who offered me the opportunity to perform my PhD thesis in his laboratory and to continue the bromodomain project. His continuous support and guidance during the work on this thesis were helpful for me in numerous ways.

I'm also grateful to Prof. Dr. Thorsten Stafforst for being my second supervisor.

Many thanks go to Cengiz Koç, Natascha Bartlick and Michael Braun from the group of Prof. Dr. Thilo Stehle for solving the crystal structures.

Thanks to Christopher Reinkemeier from the group of Prof. Dr. Edward A. Lemke at the Johannes Gutenberg University of Mainz for performing the PylRS screens and for providing the reporter and PylRS plasmids.

I would like to thank Zhadyra Yerkesh from the group of Prof. Dr. Wolfgang Fischle at the King Abdullah University of Science and Technology in Saudi Arabia for generating and providing the SML-ready nucleosomal core particles.

I also want to thank Jürgen Eirich from the group of Prof. Dr. Iris Finkemeier at the University of Münster for performing the mass spectrometry and data processing of the proteomics experiments.

Many thanks to the current and former members of the Schwarzer lab for all the relaxing coffee breaks, dinner 4 1s, a wonderful working atmosphere and the many always valuable advises: Alex, Julia, Lena, Diego, Jan, Julian, Stefan, Yandan, Luisa and Katharina. Special thanks to Julia for guiding me into the depths of HDAC research and western blotting and for all the discussions of problems. Thanks to Julian for sharing most of the PhD time with me and for your support in chemical questions (9-BBN!). Thanks to Lena and Jan, who guided me especially at the beginning of the thesis. Thanks to Luisa for reading and commenting on this work. I also like to thank Katharina for all the support in the lab by providing buffers, gels, cells and so on. I am very grateful to Maren Scheffler and Annika Kühn for the supporting work during their bachelor and master thesis.

I would also like to thank all my friends and colleagues, most importantly Matze, Gerd, Michi, Pascal, Daniel and Moritz for the great time together.

My final words go to my family. Without the support and benefits of my family, this work would not have been possible. Thanks to my father, for the review and correction of the thesis.

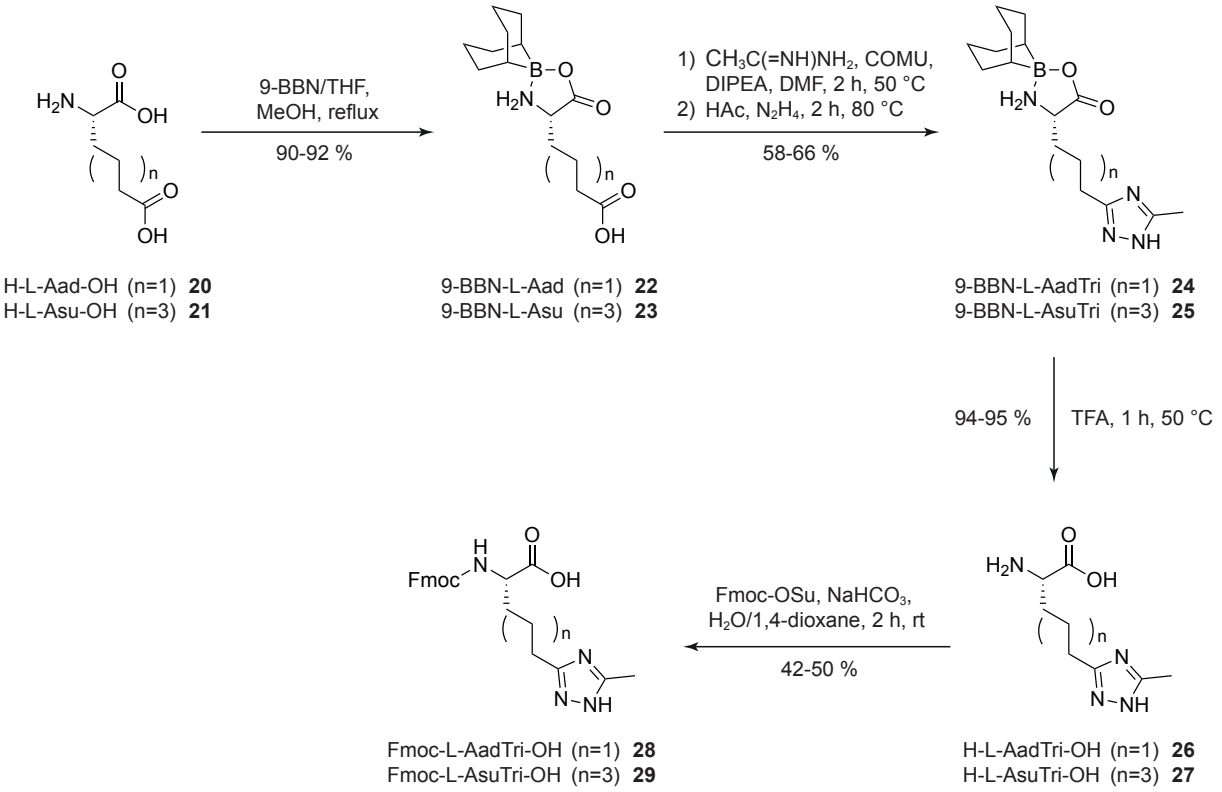
Appendices

A Supporting figures and tables

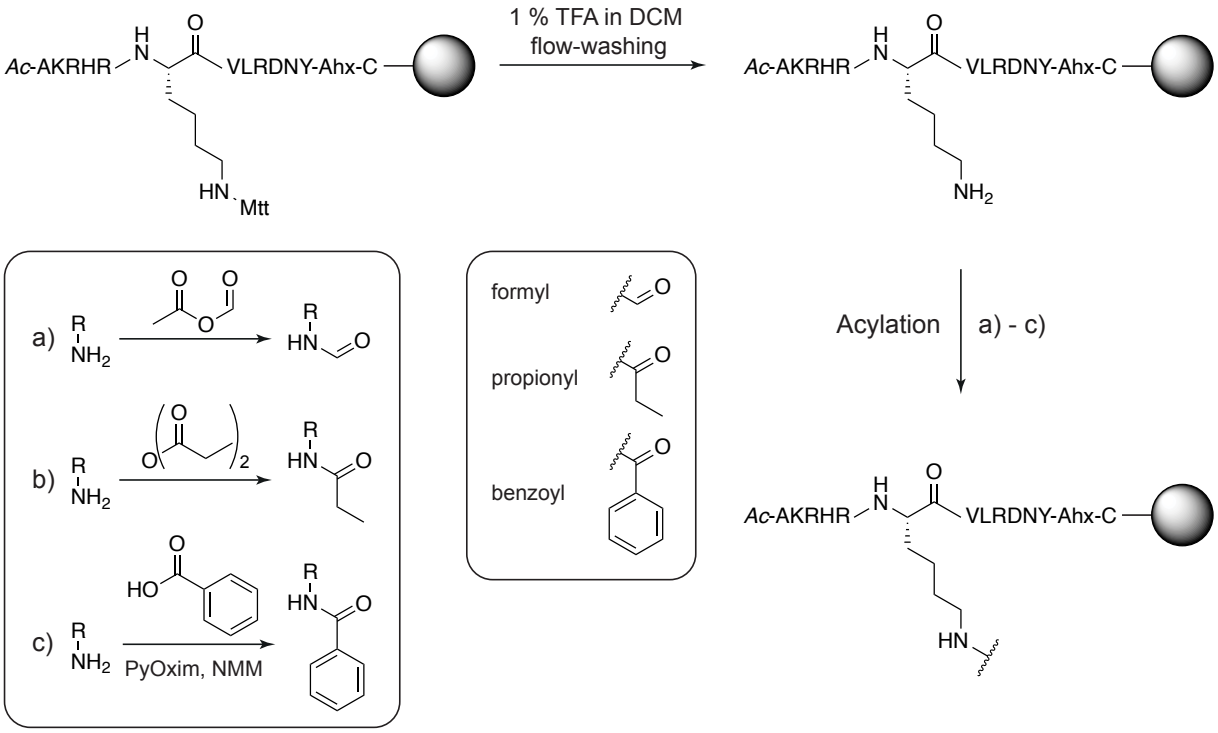
Table A.1: Overview of synthesized building blocks and their abbreviations.

Nr.	Building Block	Abbreviation	LC-MS figure	NMR figure
2	9-BBN-DL-Apm	-	C.1	C.10
3	9-BBN-DL-Apm(1,2,4-triazole)	9-BBN-DL-ApmFTri	C.1	C.11
4	9-BBN-DL-Apm(3-methyl-1,2,4-triazole)	9-BBN-DL-ApmTri	C.1	C.12
5	9-BBN-DL-Apm(3-ethyl-1,2,4-triazole)	9-BBN-DL-ApmPTri	C.2	C.13
6	9-BBN-DL-Apm(3-phenyl-1,2,4-triazole)	9-BBN-DL-ApmBTri	C.2	C.14
7	H-DL-Apm(1,2,4-triazole)-OH	ApmFTri	C.2	C.15
8	H-DL-Apm(3-methyl-1,2,4-triazole)-OH	ApmTri	C.2	-
9	H-DL-Apm(3-ethyl-1,2,4-triazole)-OH	ApmPTri	C.2	C.16
10	H-DL-Apm(3-phenyl-1,2,4-triazole)-OH	ApmBTri	C.3	C.17
11	Fmoc-DL-Apm(1,2,4-triazole)-OH	Fmoc-ApmFTri	C.3	C.18
12	Fmoc-DL-Apm(3-methyl-1,2,4-triazole)-OH	Fmoc-ApmTri	C.3	C.19
13	Fmoc-DL-Apm(3-ethyl-1,2,4-triazole)-OH	Fmoc-ApmPTri	C.3	C.20
14	Fmoc-DL-Apm(3-phenyl-1,2,4-triazole)-OH	Fmoc-ApmBTri	C.4	C.21
15	Ac-DL-Apm(3-methyl-1,2,4-triazole)-OH	Ac-DL-ApmTri	C.4	-
16	H-L-Apm(3-methyl-1,2,4-triazole)-OH	L-ApmTri	C.7	C.22
17	Fmoc-L-Apm(3-methyl-1,2,4-triazole)-OH	Fmoc-L-ApmTri	C.4	C.23
18	H-D-Apm(3-methyl-1,2,4-triazole)-OH	D-ApmTri	-	-
19	Fmoc-D-Apm(3-methyl-1,2,4-triazole)-OH	Fmoc-D-ApmTri	C.4	-
20	H-L-Aad-OH	-	-	-
21	H-L-Asu-OH	-	-	-
22	9-BBN-L-Aad	-	C.5	C.24
23	9-BBN-L-Asu	-	C.5	C.25
24	9-BBN-L-Aad(3-methyl-1,2,4-triazole)	9-BBN-L-AadTri	C.5	C.26
25	9-BBN-L-Asu(3-methyl-1,2,4-triazole)	9-BBN-L-AsuTri	C.5	C.27
26	H-L-Aad(3-methyl-1,2,4-triazole)-OH	L-AadTri	C.6	C.28
27	H-L-Asu(3-methyl-1,2,4-triazole)-OH	L-AsuTri	C.6	C.29
28	Fmoc-L-Aad(3-methyl-1,2,4-triazole)-OH	Fmoc-L-AadTri	C.6	C.30
29	Fmoc-L-Asu(3-methyl-1,2,4-triazole)-OH	Fmoc-L-AsuTri	C.6	C.31
30	Ac-DL-Apm-OH	-	C.6	C.32
31	H-L-Apm-OH	-	-	-
32	9-BBN-L-Apm	-	C.7	-

Nr.	Building Block	Abbreviation	LC-MS figure	NMR figure
33	9-BBN-L-Apm(3-methyl-1,2,4-triazole)	9-BBN-L-ApmTri	C.7	-
34	H-DL-Asu-OH	-	-	-
35	Ac-DL-Asu-OH	-	C.7	-
36	9-BBN-L-Asu(NH-OH)	9-BBN-AsuHd	C.8	-
37	H-L-Asu(NH-OH)-OH	AsuHd	C.8	C.33
38	9-BBN-L-Asu(NH-OTrt)	-	C.8	-
39	H-L-Asu(NH-OTrt)-OH	-	C.8	-
40	Fmoc-L-Asu(NH-OTrt)-OH	-	C.9	-



Scheme A.1: Synthesis scheme of the building blocks Fmoc-L-AadTri-OH (28) and Fmoc-L-AsuTri-OH (29).



Scheme A.2: Installation of acylation modifications on lysine 20 of the H4K20 peptide probe sequence for the synthesis of H4K20for, H4K20prop and H4K20benz. Reaction a) shows the formylation via acetic formic mixed anhydride, b) illustrates the propionylation via propionic anhydride and c) the coupling of benzoic acid with PyOxim as activator.

Table A.2: Overview of recombinantly expressed proteins in *E. coli*.

Protein	PDB code	Fluorescent label	Purification Tags	MW [kDa]	ε [$M^{-1}cm^{-1}$]
BAZ2B	3G0L	TagBFP	$^{N}His_6$, $^{C}Strep$	44.04	51590
BRD3(2)	2OO1	TurboYFP	$^{N}His_6$, $^{C}Strep$	44.38	50810
BRD3(2)	2OO1	-	$^{N}His_6$	15.36	16055
BRD4(1)	2OSS	TagGFP2	$^{N}His_6$, $^{C}Strep$	46.34	54695
BRD4(1)	2OSS	-	$^{N}His_6$	17.16	27055
BRD4(2)	2OUO	TurboYFP	$^{N}His_6$, $^{C}Strep$	46.14	49445
CREBBP	3DWY	TurboYFP	$^{N}His_6$, $^{C}Strep$	45.31	60195
Sortase A	1IJA	-	$^{C}His_6$	17.79	14440

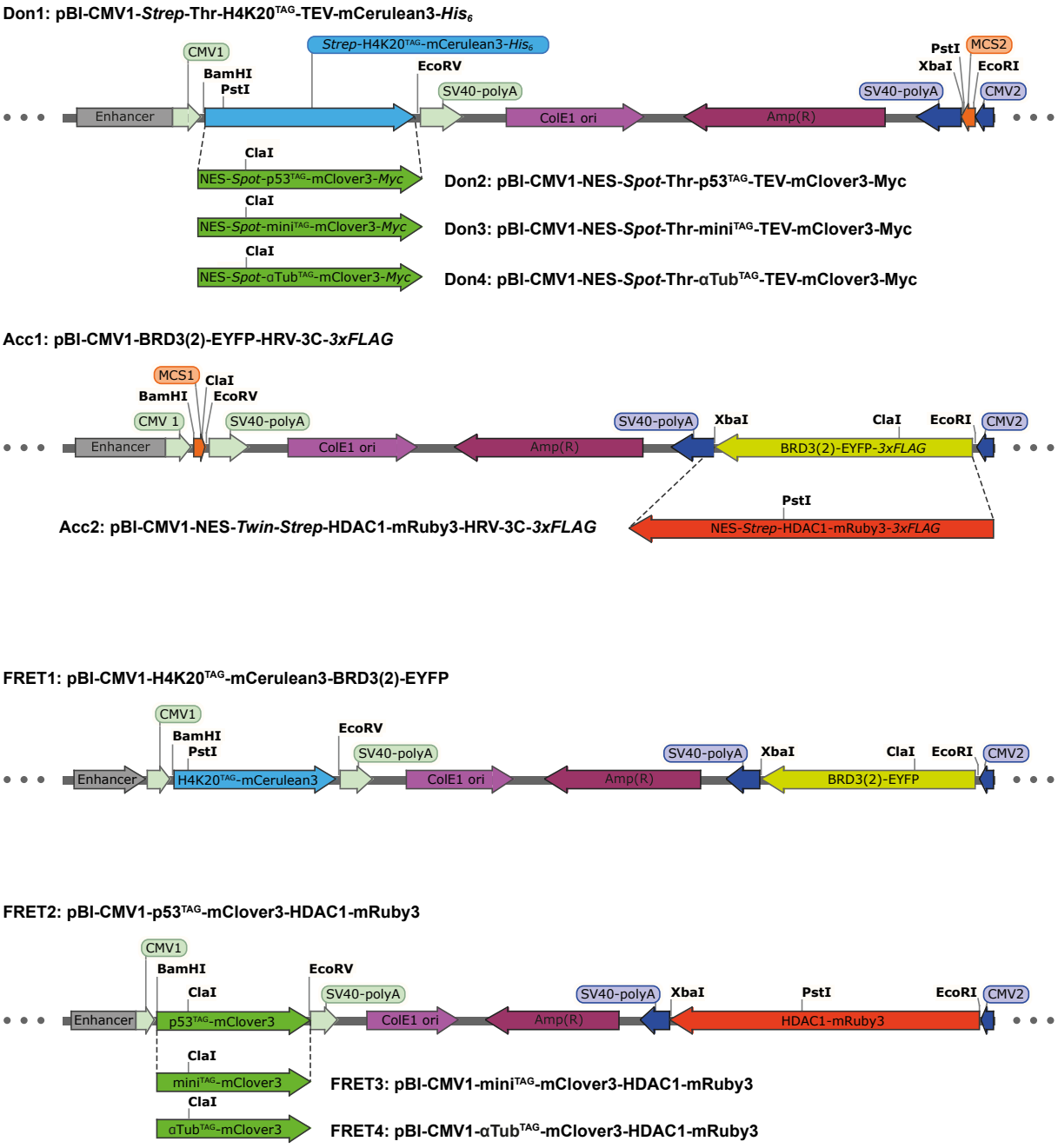


Figure A.1: Linear plasmid maps for the expression of Donor and Acceptor constructs in mammalian cells. Fusion proteins are colored in the color of the fluorescent label. Restriction sites used for cloning are highlighted in bold.

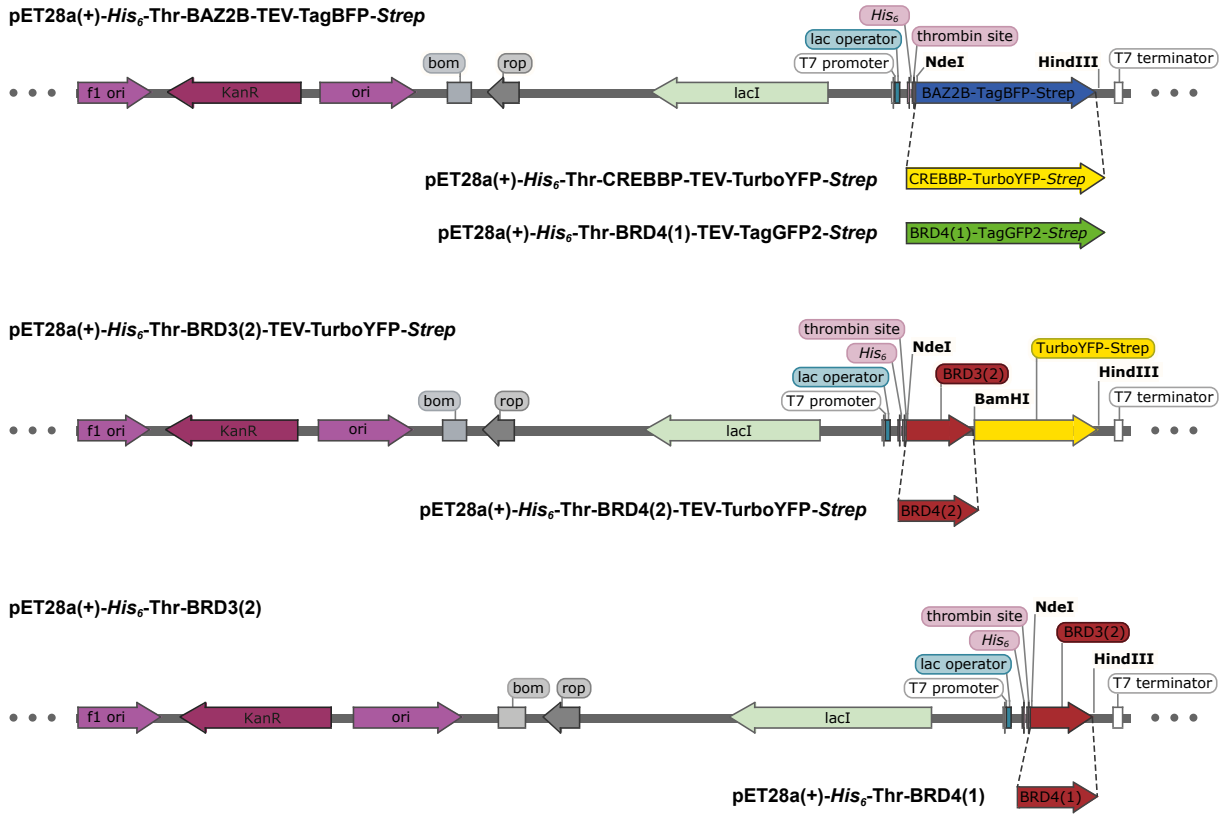


Figure A.2: Linear plasmid maps for the recombinant expression of bromodomain constructs in *E. coli*. Incorporated synthetic genes are colored in the color of the fluorescent label. BRDs are colored in red. Restriction sites used for cloning are highlighted in bold.

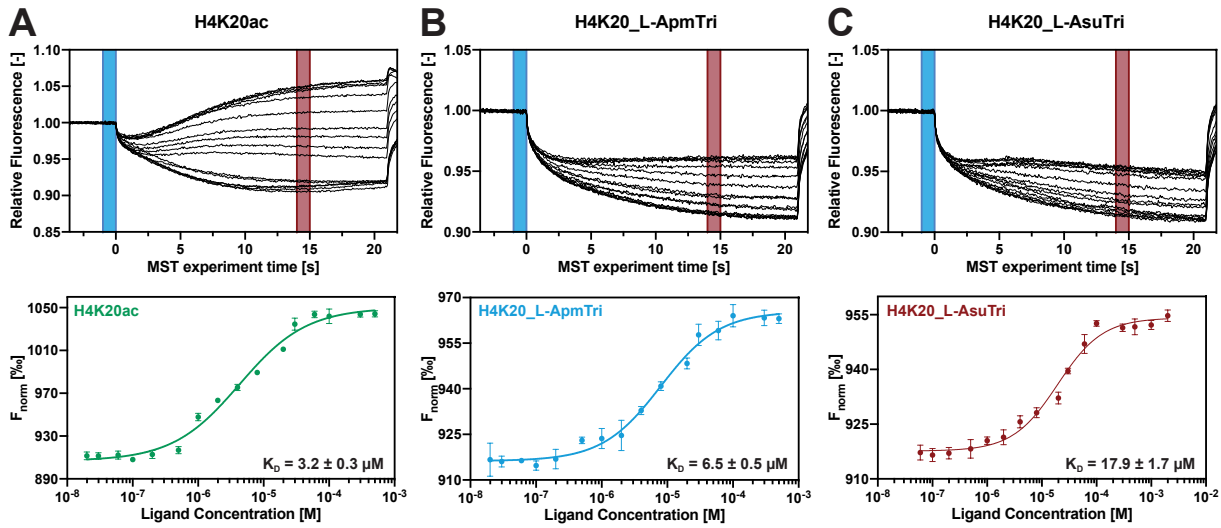


Figure A.3: MST measurements of BRD4(1) with H4K5K8 peptides. Representative MST traces (top panel) and resulting dose-response curves (bottom panel) of RED-tris-NTA-BRD3(2) towards H4K20ac (**A**), H4K20_L-ApmTri (**B**) and H4K20_L-AsuTri (**C**). Cold (blue) and hot (red) regions for the calculation of normalized fluorescence (F_{norm}) are colored. F_{norm} data are shown as mean ± s.d ($n \geq 3$). K_D values were determined by fitting the dose-response curves as described in 4.2.4.15.

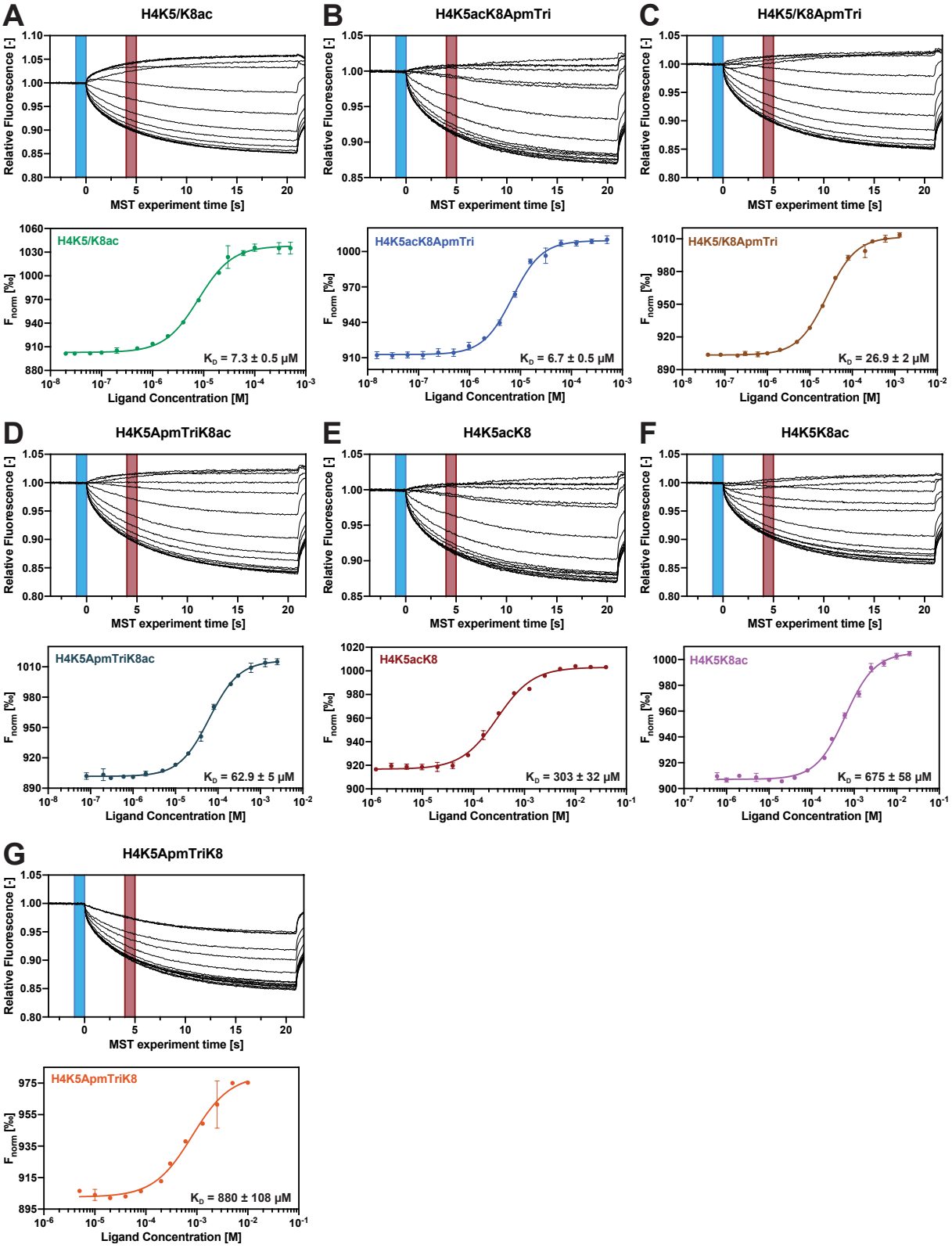


Figure A.4: MST measurements of BRD3(2) with H4K20 peptides. Representative MST traces (top panel) and resulting dose-response curves (bottom panel) of RED-tris-NTA-BRD3(2) towards H4K5/K8ac (**A**), H4K5acK8ApmTri (**B**), H4K5/K8ApmTri (**C**), H4K5ApmTriK8ac (**D**), H4K5acK8 (**E**), H4K5K8ac (**F**) and H4K5ApmTriK8 (**G**). Cold (blue) and hot (red) regions for the calculation of normalized fluorescence (F_{norm}) are colored. F_{norm} data are shown as mean ± s.d ($n \geq 3$). K_D values were determined by fitting the dose-response curves as described in 4.2.4.15.

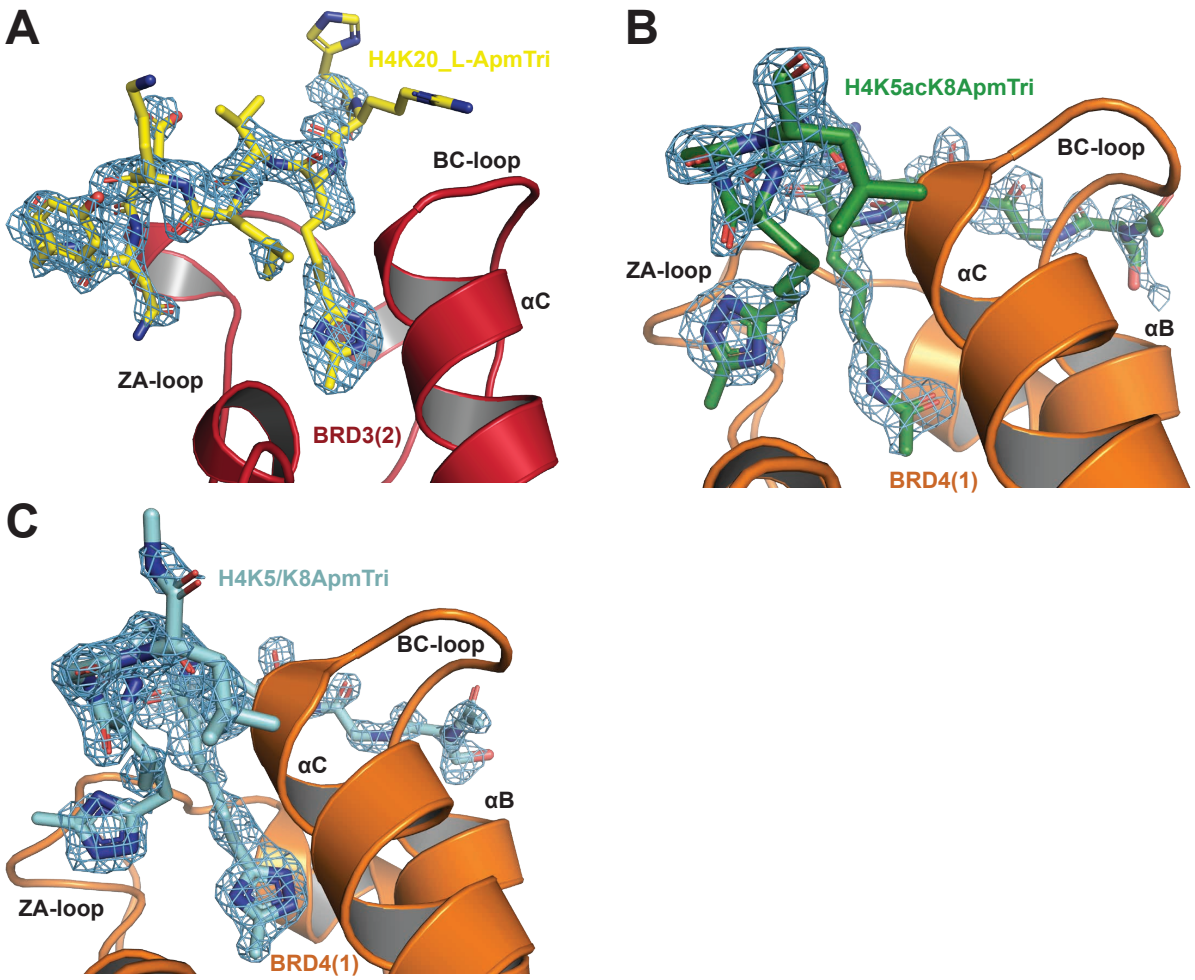


Figure A.5: Electron density maps (Fo-Fc) of L-ApmTri peptide ligands bound to BRD3(2) and BRD4(1). A Crystal structure of BRD3(2) (red) in complex with H4K20_L-ApmTri (yellow). **B** Crystal structure of BRD4(1) (orange) in complex with H4K5acK8ApmTri (green). **C** Crystal structure of BRD4(1) in complex with H4K5/K8ApmTri (light blue). Electron density maps were contoured at 3σ.

Table A.3: Data collection and refinement statistics for BRD crystal structures.

		BRD3(1) + H4K20_L-ApmTri	BRD4(1) + H4K5acK8ApmTri	BRD4(1) + H4K5/K8ApmTri
Data collection				
Wavelength [Å]		0.999884	1.000029	0.999867
Resolution [Å]		50-1.85 (1.97-1.86)	50-1.85 (1.96-1.85)	50-1.58 (1.67-1.58)
Space group		$P6_122$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimension	a, b, c [Å]	79.52, 79.52, 95.14	39.54, 51.81, 170.89	42.59, 46.82, 59.02
	α, β, γ(°)	90.0, 90.0, 120.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Measured reflections		341966 (45932)	235804 (38076)	122351 (19781)
Unique reflections		26066 (3756)	30618 (4783)	16702 (2628)
Redundancy		13.1 (12.2)	6.2 (6.4)	7.3 (7.5)
Completeness [%]		90.6 (81.2)	99.0 (97.8)	99.8 (99.7)
$I/\sigma(I)$		17.87 (1.01)	9.94 (1.33)	11.47 (1.21)
$CC_{1/2}$		99.9 (72.3)	99.7 (52.2)	99.9 (52.1)
R_{means} (%)		9.4 (267.9)	18.8 (170.6)	14.8 (194.5)
Wilson B [$Å^2$]		45.15	29.81	25.24
Refinement				
Resolution [Å]		1.85	1.85	1.58
Number of atoms	Protein A/B/C	910	1045 / 1040 /1069	1074
	Ligand A/B/C	86	15 / 55 / 74	81
	Waters	52	294	103
	Others	-	1	5
Average B-Factors [$Å^2$]	Protein A/B/C	47.5	26.6 / 29.0 / 23.4	19.6
	Ligand A/B/C	54.3	49.0 / 39.7 / 31.4	28.5
	Waters	49.0	32.5	34
	Others	-	32.0	41.4
R_{free} flag set total / %		1225 / 4.95	1531 / 5.0	1002 / 6.0
R_{work}/R_{free} [%]		20.76 / 24.94	18.20 / 24.11	17.09 / 19.95
R.m.s.d bond angles [°]		1.194	1.192	1.045
R.m.s.d bond length [Å]		0.003	0.012	0.010
Ramachandran [%]	favored	100.00	98.14	99.22
	allowed	0.00	1.86	0.078
	outliers	0.00	0.00	0.00

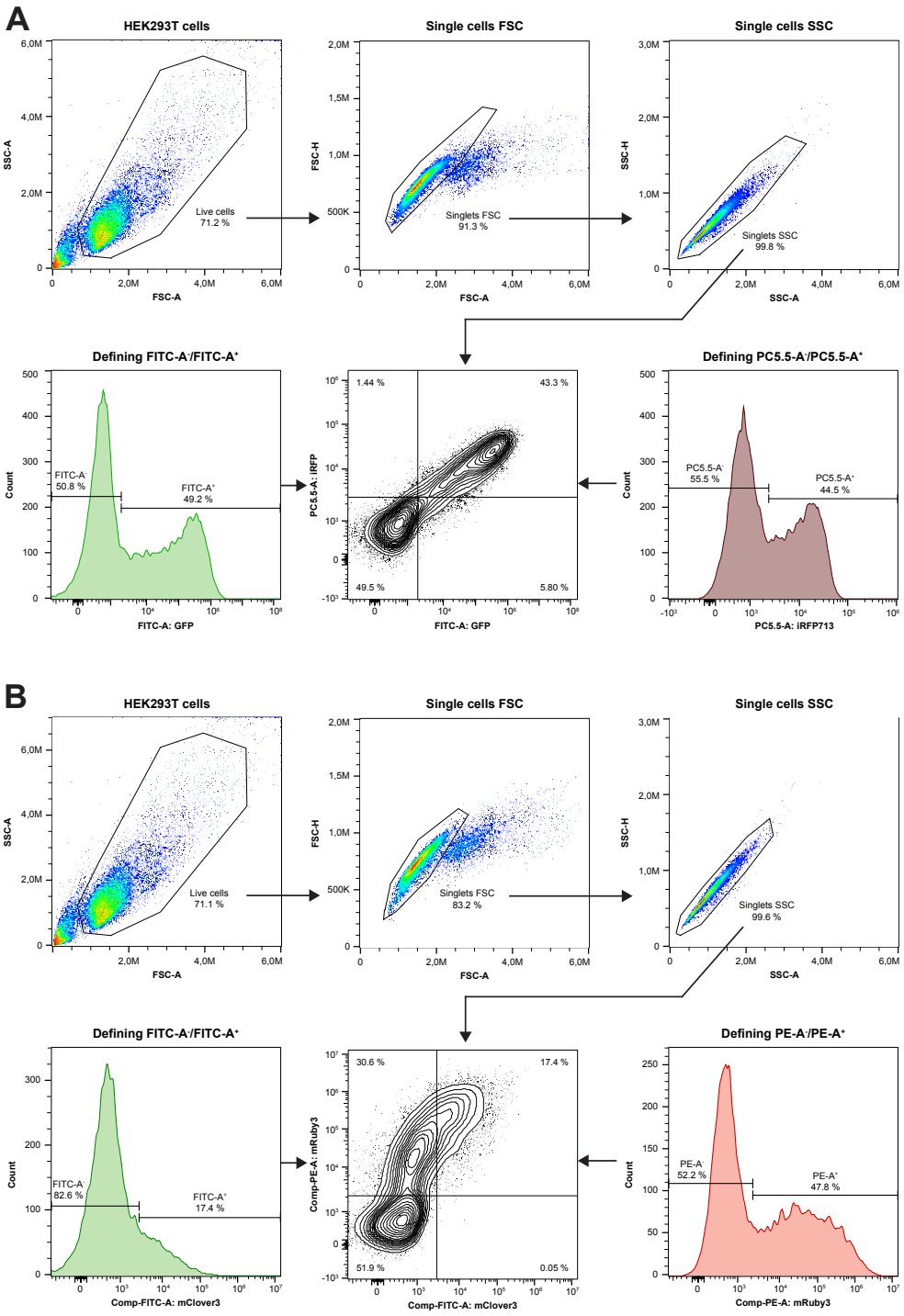


Figure A.6: Sequential gating strategies for flow cytometric data. A Gating scheme for cells expressing the iRFP-GFP^{Y39TAG} reporter construct shown exemplary on cells transfected with PylRS and the reporter plasmid and treated with 250 µM BocK. **B** Gating scheme for cells expressing proteins labeled with mClover3 and mRuby3 shown exemplary on cells transfected with PylRS^{AF} and the FRET4 plasmid and treated with 250 µM BocK. Based on FSC-A vs SSC-A plot cells were gated for live cells followed by identifying single cells using the ratios of height to width in FSC and SSC (Singlets FSC, Singlets SSC). GFP and mClover3 were measured using FITC-A filter settings, iRFP713 was measured in the PC5.5-A channel and mRuby3 in the PE-A channel. PC5.5$^+$, FITC$^+$ and PE-A$^+$ cell populations were defined based on fluorescence intensities measured for non-transfected, single and double positive cells. Final data were visualized in two-parameter contour blots.

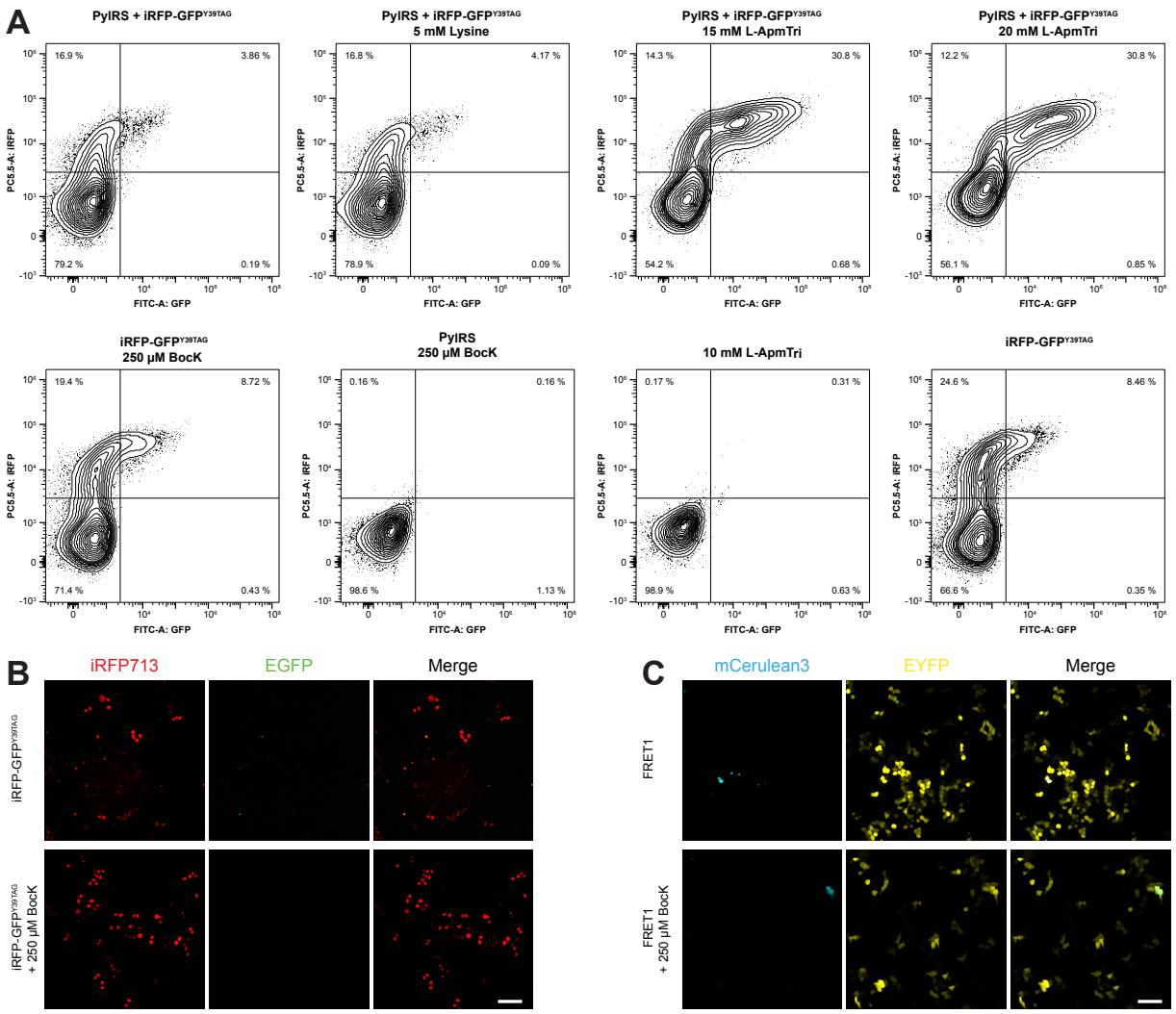


Figure A.7: Amber suppression controls of ncAA incorporation into the reporter construct and FRET1. A Representative two-parameter contour blots of several L-ApmTri concentrations, 5 mM lysine and control expressions lacking different components (ncAA, PylRS, reporter plasmid). **B** Fluorescence microscopy images of living HEK293T cells transfected with the reporter construct and treated with 250 µM BocK. Scale bar, 100 µM. **C** Fluorescence microscopy images of living HEK293T cells transfected with FRET1 and treated with 250 µM BocK. Scale bar, 100 µM.

Table A.4: Overview of suppliers and their locations. In this work, the short versions for the suppliers are used.

Supplier	Short	Location
3M Company	3M	Maplewood, MN, United States
Abcam PLC	Abcam	Cambridge, United Kingdom
Agilent Technologies, Inc.	Agilent	Santa Clara, CA, United States
AVESTIN, Inc.	Avestin	Ottawa, Ontario, Canada
Bachem Holding AG	Bachem	Bubendorf BL, Switzerland
Beckman Coulter, Inc.	Beckman Coulter	Brea, CA, United States
Bio-Rad Laboratories, Inc	Bio-Rad	Hercules, CA, United States
Biosolve B.V.	Biosolve	Valkenswaard, Netherlands
Biotage Sweden AB	Biotage	Uppsala, Sweden
Biozym Scientific GmbH	Biozym	Oldendorf, Germany
Bruker Corporation	Bruker	Billerica, MA, United States
Carbolution Chemicals GmbH	Carbolution	St. Ingbert, Germany
Carl Roth GmbH + Co. KG	Carl Roth	Karlsruhe, Germany
Carl Zeiss AG	Carl Zeiss	Oberkochen, Germany
ChromoTek GmbH	ChromoTek	Planegg-Martinsried, Germany
Clontech Laboratories Inc (now Takara)	Clontech	Mountain View, CA, United States
Dr. A. Maisch HPLC GmbH	Dr. Maisch	Ammerbuch-Entringen, Germany
Eurofins Scientific SE	Eurofins	Hamburg, Germany
Frank Gutjahr Chromatographie	-	Balingen, Germany
FUJIFILM Wako Chemicals Europe GmbH	FUJIFILM Wako Chemicals	Neuss, Germany
GATC Biotech AG (now Eurofins)	GATC Biotech	Konstanz, Germany
GE Healthcare	-	Chicago, IL, United States
GL Biochem (Shanghai) Ltd.	GL Biochem	Shanghai, China
Greiner AG	Greiner	Kremsmünster, Österreich
IBA Lifesciences GmbH	IBA	Göttingen, Germany
IKA Werke GmbH & Co. KG	IKA	Staufen, Germany
Infors AG	Infors	Bottmingen, Switzerland
Iris Biotech GmbH	Iris Biotech	Marktredwitz, Germany
J.T. Baker Inc. (Avantor)	J.T. Baker	Deventer, Netherlands
LI-COR Biosciences Inc.	LI-COR	Lincoln, NE, United States
Martin Christ GmbH	Martin Christ	Osterode am Harz, Germany
Merck KGaA	Merck	Darmstadt, Germany
NanoTemper Technologies GmbH	NanoTemper	Munich, Germany
New England Biolabs, Inc.	New England Biolabs	Ipswich, MA, United States
PEQLAB Biotechnologie GmbH (now VWR)	Peqlab	Erlangen, Germany
Phenomenex Inc.	Phenomenex	Torrance, CA, United States
Polyplus Transfection	-	Illkirch, France
Promega Corporation	Promega	Madison, WI, United States
QIAGEN N.V.	Qiagen	Venlo, Netherlands
Rapp Polymere GmbH	Rapp Polymere	Tübingen, Germany
Santa Cruz Biotechnology, Inc.	Santa Cruz	Dallas, TX, United States
Sarstedt AG & Co. KG	Sarstedt	Nümbrecht, Germany
Shimadzu Corp.	Shimadzu	Kyōto, Japan
Sigma-Aldrich Corp. (now Merck KGaA)	Sigma-Aldrich	St. Louis, MO, United States
Spectrum Laboratories Inc.	Spectrum Laboratories	Rancho Dominguez, CA, United States
Takara Bio USA, Inc.	Takara	Mountain View, CA, United States
Tecan Trading AG	Tecan	Männedorf, Switzerland
Thermo Fisher Scientific Inc.	Thermo Fisher Scientific	Waltham, MA, United States
Th. Geyer GmbH & Co. KG	Th. Geyer	Renningen, Germany
VWR International (Avantor)	VWR	Radnor, PA, United States
Yokogawa Denki K.K.	Yokogawa Denki	Musashino, Japan

B Protein sequences

B.1 Sequences of protein constructs recombinantly expressed in *E. coli*

Table B.1: Sequences of recombinantly expressed proteins in *E. coli*.

His_6-BAZ2B-TagBFP-*Strep*

```
  1 MGSSHHHHHH SSGLVPRGSH MSVKKPKRDD SKDLALCSMI LTEMETHEDA WPFLLPVNLK
 61 LVPGYKKVIK KPMDFSTIRE KLSSGQYPNL ETFALDVRLV FDNCETFNED DSDIGRAGHN
121 MRKYFEKKWT DTFKVSGGEN LYFQSGGELI KENMHMKLYM EGTVDNHHFK CTSEGEGKPY
181 EGTQTMRIKV VEGGPLPFAF DILATSFLYG SKTFINHTQG IPDFFKQSFP EGFTWERVTT
241 YEDGGVLTAT QDTSLQDGCL IYNVKIRGVN FTSNGPVMQK KTLGWEAFTE TLYPADGGLE
301 GRNDMALKLV GGSHLIANIK TTYRSKKPAK NLKMPGVYYV DYRLERIKEA NNETYVEQHE
361 VAVARYCDLP SKLGHKLNGG GWSHPQFEK
```

His_6-BRD3(2)-TurboYFP-*Strep*

```
  1 MGSSHHHHHH SSGLVPRGSH MGKLSEHLRY CDSILREMLS KKHAAYAWPF YKPVDAEALE
 61 LHDYHDIIKH PMDLSTVKRK MDGREYPDAQ GFAADVRLMF SNCYKYNPPD HEVVAMARKL
121 QDVFEMRFAK MPGGENLYFQ SGGSSGALLF HGKIPYVVEM EGNVDGHTFS IRGKGYGDAS
181 VGKVDAQFIC TTGDVPVPWS TLVTTLTYGA QCFAKYGPEL KDFYKSCMPD GYVQERTITF
241 EGDGNFKTRA EVTFENGSVY NRVKLNGQGF KKDGHVLGKN LEFNFTPHCL YIWGDQANHG
301 LKSAFKICHE ITGSKGDFIV ADHTQMNTPI GGGPVHVPEY HHMSYHVKLS KDVTDHRDNM
361 SLKETVRAVD CRKTYDFDAG SGDTSGGGWS HPQFEK
```

His_6-BRD3(2)

```
  1 MGSSHHHHHH SSGLVPRGSH MGKLSEHLRY CDSILREMLS KKHAAYAWPF YKPVDAEALE
 61 LHDYHDIIKH PMDLSTVKRK MDGREYPDAQ GFAADVRLMF SNCYKYNPPD HEVVAMARKL
121 QDVFEMRFAK MP
```

His_6-BRD4(1)-TagGFP2-*Strep*

```
  1 MGSSHHHHHH SSGLVPRGSH MNPPPPETSN PNKPKRQTNQ LQYLLRVVLK TLWKHQFAWP
 61 FQQPVDAVKL NLPDYYKIIK TPMDMGTIKK RLENNYYWNA QECIQDFNTM FTNCYIYNKP
121 GDDIVLMAEA LEKLFLQKIN ELPTEEGGEN LYFQSGGMSG GEELFAGIVP VLIELDGDVH
181 GHKFSVRGEG EGDADYGKLE IKFICTTGKL PVPWPTLVTT LCYGIQCFAR YPEHMKMNDF
241 FKSAMPEGYI QERTIQFQDD GKYKTRGEVK FEGDTLVNRI ELKGKDFKED GNILGHKLEY
301 SFNSHNVYIR PDKANNGLEA NFKTRHNIEG GGVQLADHYQ TNVPLGDGPV LIPINHYLST
361 QTKISKDRNE ARDHMVLLES FSACCHTHGM DELYRGGGWS HPQFEK
```

His_6-BRD4(1)

```
  1 MGSSHHHHHH SSGLVPRGSH MNPPPPETSN PNKPKRQTNQ LQYLLRVVLK TLWKHQFAWP
 61 FQQPVDAVKL NLPDYYKIIK TPMDMGTIKK RLENNYYWNA QECIQDFNTM FTNCYIYNKP
121 GDDIVLMAEA LEKLFLQKIN ELPTEE
```

His_6-BRD4(2)-TurboYFP-*Strep*

```
  1 MGSSHHHHHH SSGLVPRGSH MKDVPDSQQH PAPEKSSKVS EQLKCCSGIL KEMFAKKHAA
 61 YAWPFYKPVD VEALGLHDYC DIIKHPMDMS TIKSKLEARE YRDAQEFGAD VRLMFSNCYK
121 YNPPDHEVVA MARKLQDVFE MRFAKMPDEG GENLYFQSGG SSGALLFHGK IPYVVEMEGN
181 VDGHTFSIRG KGYGDASVGK VDAQFICTTG DVPVPWSTLV TTLTYGAQCF AKYGPELKDF
241 YKSCMPDGYV QERTITFEGD GNFKTRAEVT FENGSVYNRV KLNGQGFKKD GHVLGKNLEF
301 NFTPHCLYIW GDQANHGLKS AFKICHEITG SKGDFIVADH TQMNTPIGGG PVHVPEYHHM
361 SYHVKLSKDV TDHRDNMSLK ETVRAVDCRK TYDFDAGSGD TSGGGWSHPQ FEK
```

His_6-CREBBP-TurboYFP-*Strep*

```
  1 MGSSHHHHHH SSGLVPRGSH MRKKIFKPEE LRQALMPTLE ALYRQDPESL PFRQPVDPQL
 61 LGIPDYFDIV KNPMDLSTIK RKLDTGQYQE PWQYVDDVWL MFNNAWLYNR KTSRVYKFCS
121 KLAEVFEQEI DPVMQSLGGG ENLYFQSGGS SGALLFHGKI PYVVEMEGNV DGHTFSIRGK
181 GYGDASVGKV DAQFICTTGD VPVPWSTLVT TLTYGAQCFA KYGPELKDFY KSCMPDGYVQ
241 ERTITFEGDG NFKTRAEVTF ENGSVYNRVK LNGQGFKKDG HVLGKNLEFN FTPHCLYIWG
301 DQANHGLKSA FKICHEITGS KGDFIVADHT QMNTPIGGGP VHVPEYHHMS YHVKLSKDVT
361 DHRDNMSLKE TVRAVDCRKT YDFDAGSGDT SGGGWSHPQF EK
```

Sortase A-His_6

```
  1 MQAKPQIPKD KSKVAGYIEI PDADIKEPVY PGPATPEQLN RGVSFAEENE SLDDQNISIA
 61 GHTFIDRPNY QFTNLKAAKK GSMVYFKVGN ETRKYKMTSI RDVKPTDVGV LDEQKGKDKQ
121 LTLITCDDYN EKTGVWEKRK IFVATEVKLE HHHHHH
```

B.2 Sequences of protein constructs for mammalian cell expression

Table B.2: Sequences of recombinantly expressed proteins in *E. coli*.

Don1: *Strep*-H4K20^{TAG}-mCerulean3-His_6

```
  1 MASWSHPQFE KGGSGGSLVP RGSLQGGAKR HR*VLRDNGG ENLYFQSRSL VSKGEELFTG
 61 VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT GKLPVPWPTL VTTLSWGVQC
121 FARYPDHMKQ HDFFKSAMPE GYVQERTIFF KDDGNYKTRA EVKFEGDTLV NRIELKGIDF
181 KEDGNILGHK LEYNAIHGNV YITADKQKNG IKANFGLNCN IEDGSVQLAD HYQQNTPIGD
241 GPVLLPDNHY LSTQSKLSKD PNEKRDHMVL LEFVTAAGIT LGMDELYKGG SGGSHHHHHH
```

Acc1: BRD3(2)-EYFP-*3xFLAG*

```
  1 MAGKLSEHLR YCDSILREML SKKHAAYAWP FYKPVDAEAL ELHDYHDIIK HPMDLSTVKR
 61 KMDGREYPDA QGFAADVRLM FSNCYKYNPP DHEVVAMARK LQDVFEMRFA KMPIDGGSGG
121 MVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY GKLTLKFICT TGKLPVPWPT
181 LVTTFGYGLQ CFARYPDHMK QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL
241 VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN GIKVNFKIRH NIEDGSVQLA
301 DHYQQNTPIG DGPVLLPDNH YLSYQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYKG
361 GLEVLFQGPG GDYKDHDGDY KDHDIDYKDD DDK
```

Don2: NES-*Spot*-p53^{TAG}-mClover3-*Myc*

```
  1 MASLALKLAG LDIGGPDRVR AVSHWSSGGS GGSLVPRGSQ LGGTSRHK*L MFKTEGGENL
 61 YFQSIDVSKG EELFTGVVPI LVELDGDVNG HKFSVRGEGE GDATNGKLTL KFICTTGKLP
121 VPWPTLVTTF GYGVACFSRY PDHMKQHDFF KSAMPEGYVQ ERTISFKDDG TYKTRAEVKF
181 EGDTLVNRIE LKGIDFKEDG NILGHKLEYN FNSHYVYITA DKQKNCIKAN FKIRHNVEDG
241 SVQLADHYQQ NTPIGDGPVL LPDNHYLSHQ SKLSKDPNEK RDHMVLLEFV TAAGITHGMD
301 ELYKGGSGGS EQKLISEEDL
```

Don3: NES-*Spot*-mini^{TAG}-mClover3-*Myc*

```
  1 MASLALKLAG LDIGGPDRVR AVSHWSSGGS GGSLVPRGSQ LGGGGSGGSG G*GGSGGSGG
 61 SGGENLYFQS IDVSKGEELF TGVVPILVEL DGDVNGHKFS VRGEGEGDAT NGKLTLKFIC
121 TTGKLPVPWP TLVTTFGYGV ACFSRYPDHM KQHDFFKSAM PEGYVQERTI SFKDDGTYKT
181 RAEVKFEGDT LVNRIELKGI DFKEDGNILG HKLEYNFNSH YVYITADKQK NCIKANFKIR
241 HNVEDGSVQL ADHYQQNTPI GDGPVLLPDN HYLSHQSKLS KDPNEKRDHM VLLEFVTAAG
301 ITHGMDELYK GGSGGSEQKL ISEEDL
```

Don4: NES-*Spot*-αTub^{TAG}-mClover3-*Myc*

```
  1 MASLALKLAG LDIGGPDRVR AVSHWSSGGS GGSLVPRGSQ LGGPDGQMPS D*TIGGGDDS
 61 GGENLYFQSI DVSKGEELFT GVVPILVELD GDVNGHKFSV RGEGEGDATN GKLTLKFICT
121 TGKLPVPWPT LVTTFGYGVA CFSRYPDHMK QHDFFKSAMP EGYVQERTIS FKDDGTYKTR
181 AEVKFEGDTL VNRIELKGID FKEDGNILGH KLEYNFNSHY VYITADKQKN CIKANFKIRH
241 NVEDGSVQLA DHYQQNTPIG DGPVLLPDNH YLSHQSKLSK DPNEKRDHMV LLEFVTAAGI
301 THGMDELYKG GSGGSEQKLI SEEDL
```

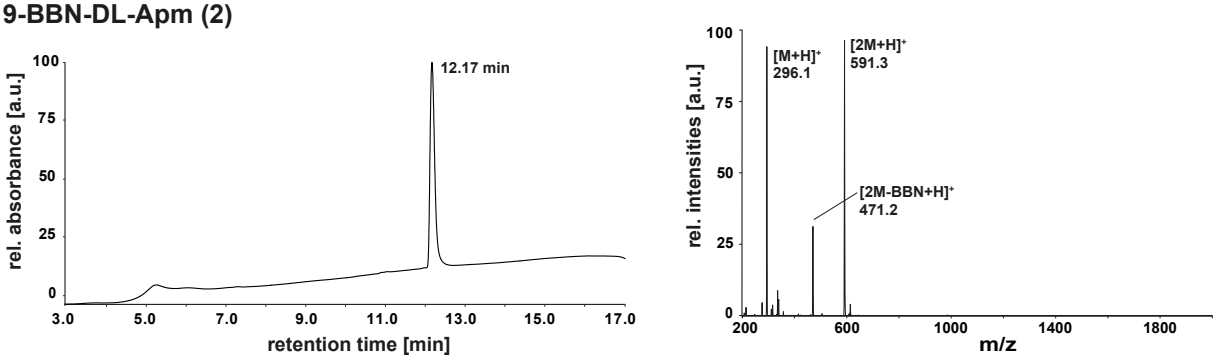
Acc2: NES-*Twin-Strep*-HDAC1-mRuby3-*3xFLAG*

```
  1 MASLALKLAG LDIGGWSHPQ FEKGGGSGGG SGGSAWSHPQ FEKSGGGSAQ TQGTRRKVCY
 61 YYDGDVGNYY YGQGHPMKPH RIRMTHNLLL NYGLYRKMEI YRPHKANAEE MTKYHSDDYI
121 KFLRSIRPDN MSEYSKQMQR FNVGEDCPVF DGLFEFCQLS TGGSVASAVK LNKQQTDIAV
181 NWAGGLHHAK KSEASGFCYV NDIVLAILEL LKYHQRVLYI DIDIHHGDGV EEAFYTTDRV
241 MTVSFHKYGE YFPGTGDLRD IGAGKGKYYA VNYPLRDGID DESYEAIFKP VMSKVMEMFQ
301 PSAVVLQCGS DSLSGDRLGC FNLTIKGHAK CVEFVKSFNL PMLMLGGGGY TIRNVARCWT
361 YETAVALDTE IPNELPYNLQ GGSGGSVSKG EELIKENMRM KVVMEGSVNG HQFKCTGEGE
421 YETAVALDTE IPNELPYNLQ GGSGGSVSKG EELIKENMRM KVVMEGSVNG HQFKCTGEGE
481 GRPYEGVQTM RIKVIEGGPL PFAFDILATS FMYGSRTFIK YPADIPDFFK QSFPEGFTWE
541 GGLRGYTDIA LKVDGGGHLH CNFVTTYRSK KTVGNIKMPG VHAVDHRLER IEESDNETYV
601 VQREVAVAKY SNLGGGMDEL YKGGLEVLFQ GPGGDYKDHD GDYKDHDIDY KDDDDK
```

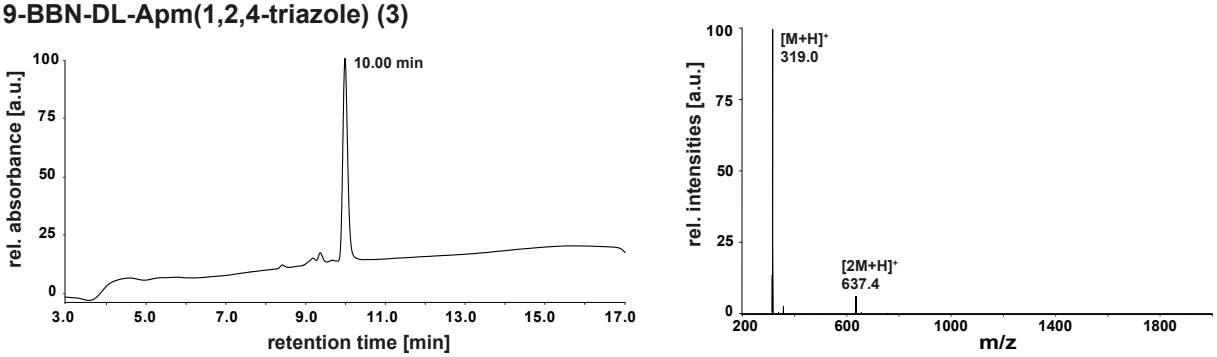
C Analytical results

C.1 Analytical results of building block synthesis

9-BBN-DL-Apm (2)



9-BBN-DL-Apm(1,2,4-triazole) (3)



9-BBN-DL-Apm(3-methyl-1,2,4-triazole) (4)

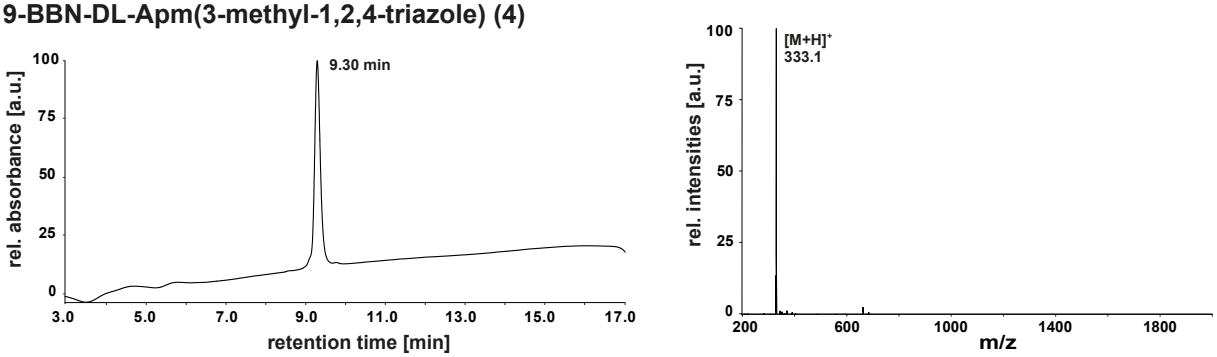
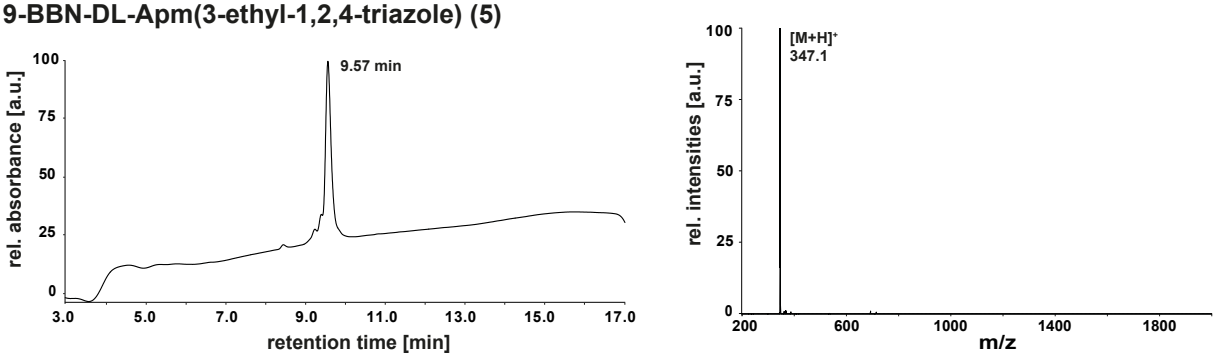
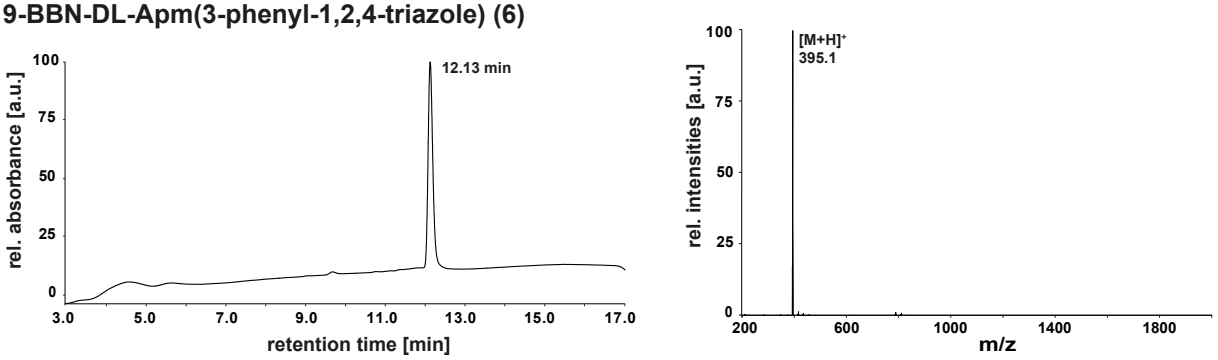


Figure C.1: LC-MS analysis of synthesized building blocks 9-BBN-DL-Apm (2), 9-BBN-DL-ApmFTri (3) and 9-BBN-DL-ApmTri (4).

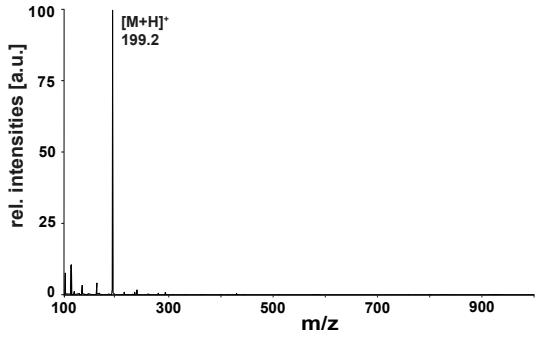
9-BBN-DL-Apm(3-ethyl-1,2,4-triazole) (5)



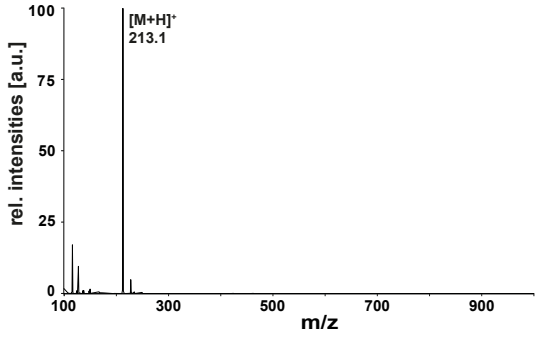
9-BBN-DL-Apm(3-phenyl-1,2,4-triazole) (6)



H-DL-Apm(1,2,4-triazole)-OH (7)



H-DL-Apm(3-methyl-1,2,4-triazole)-OH (8)



H-DL-Apm(3-ethyl-1,2,4-triazole)-OH (9)

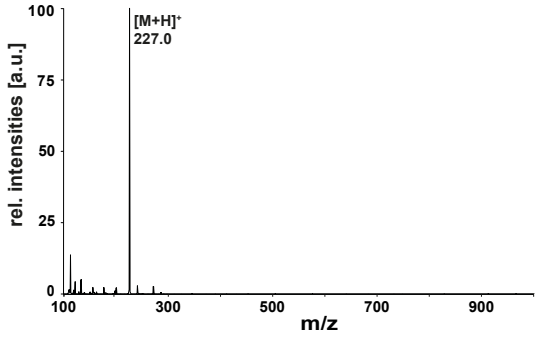
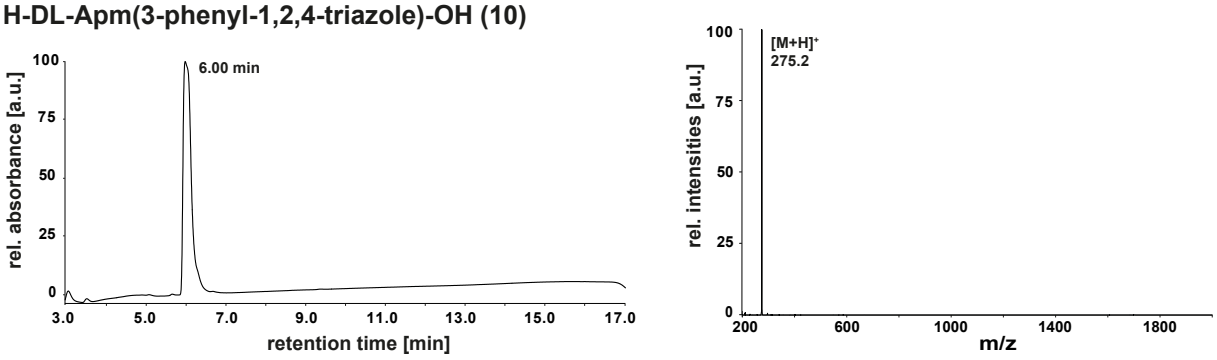
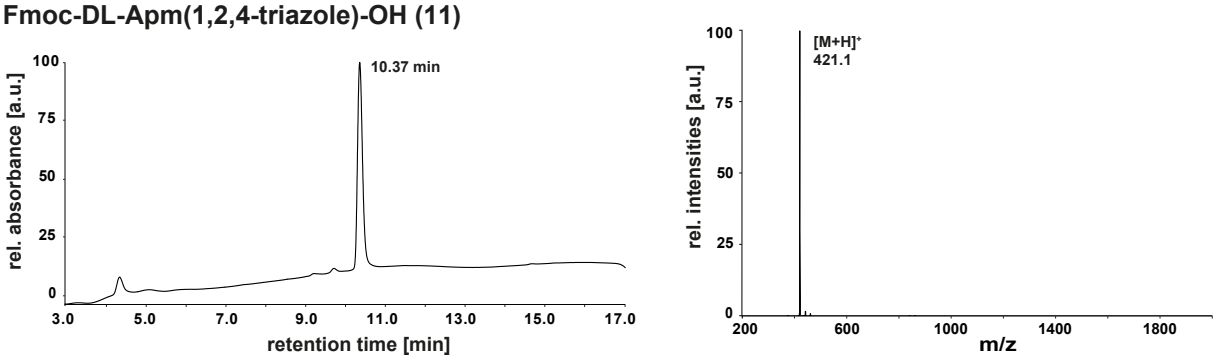


Figure C.2: LC-MS analysis of synthesized building blocks 9-BBN-DL-ApmPTri (5), 9-BBN-DL-ApmBTri (6), ApmFTri (7), ApmTri (8) and ApmPTri (9).

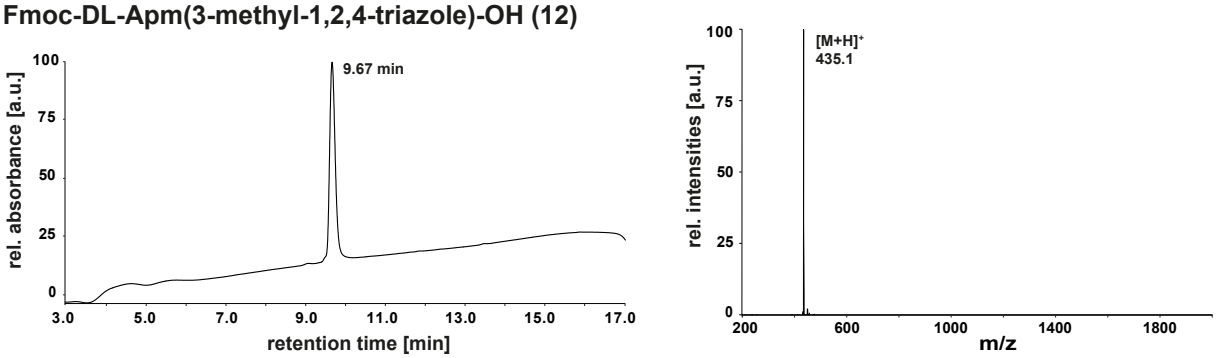
H-DL-Apm(3-phenyl-1,2,4-triazole)-OH (10)



Fmoc-DL-Apm(1,2,4-triazole)-OH (11)



Fmoc-DL-Apm(3-methyl-1,2,4-triazole)-OH (12)



Fmoc-DL-Apm(3-ethyl-1,2,4-triazole)-OH (13)

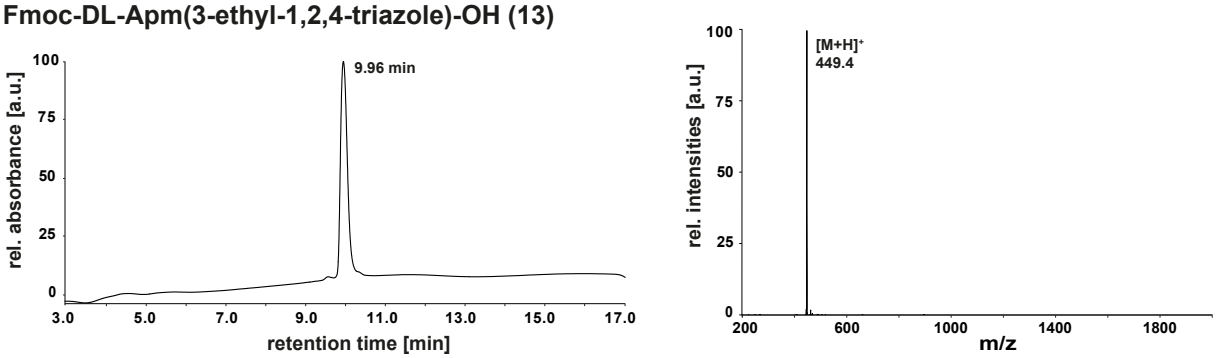
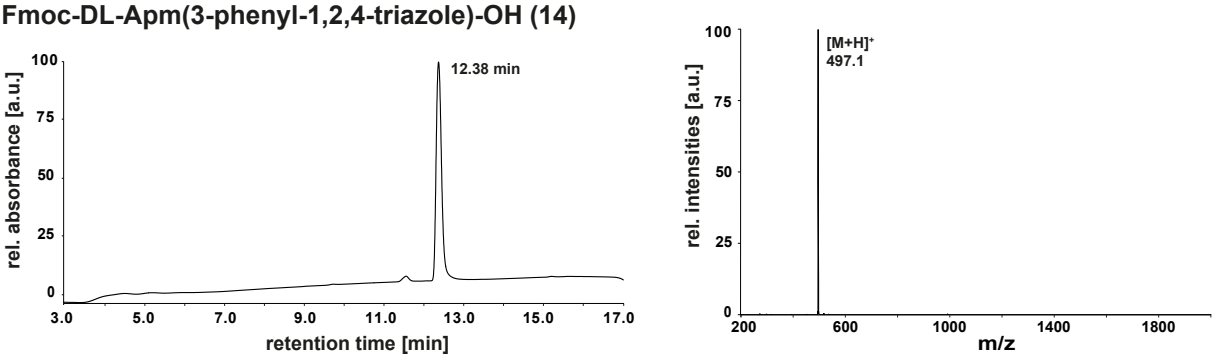
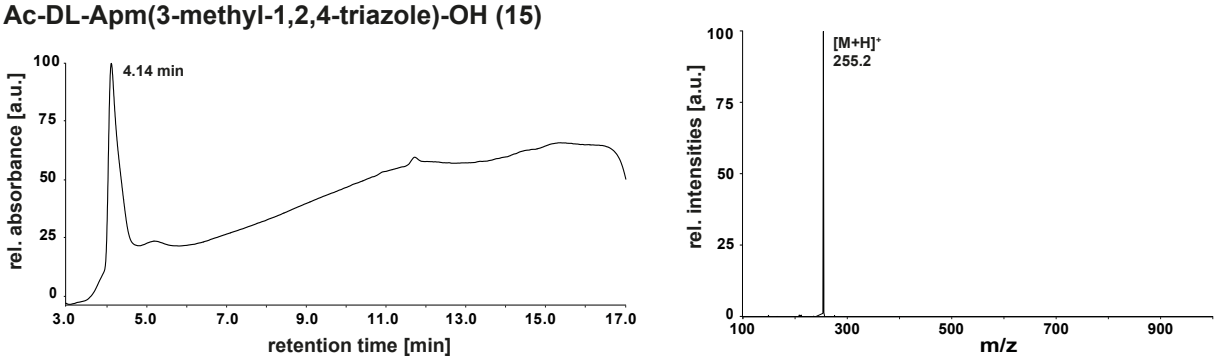


Figure C.3: LC-MS analysis of synthesized building blocks ApmBTri (10), Fmoc-ApmFTri (11), Fmoc-ApmTri (12) and Fmoc-ApmPTri (13).

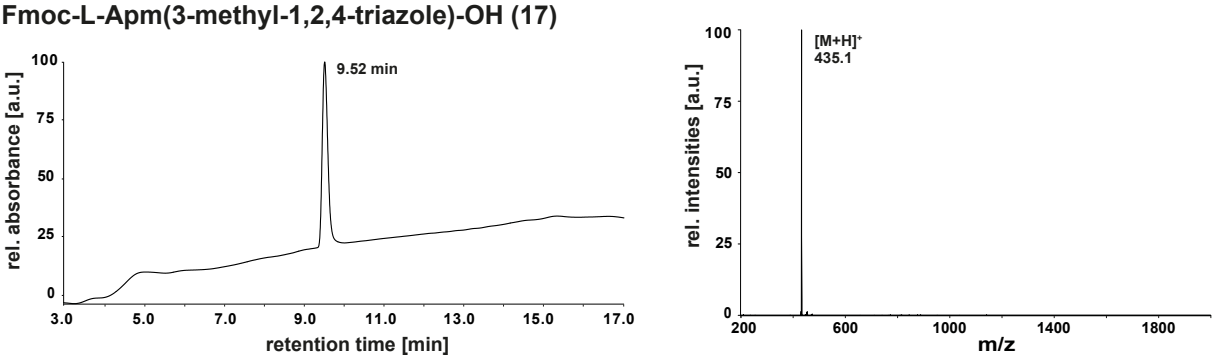
Fmoc-DL-Apm(3-phenyl-1,2,4-triazole)-OH (14)



Ac-DL-Apm(3-methyl-1,2,4-triazole)-OH (15)



Fmoc-L-Apm(3-methyl-1,2,4-triazole)-OH (17)



Fmoc-D-Apm(3-methyl-1,2,4-triazole)-OH (19)

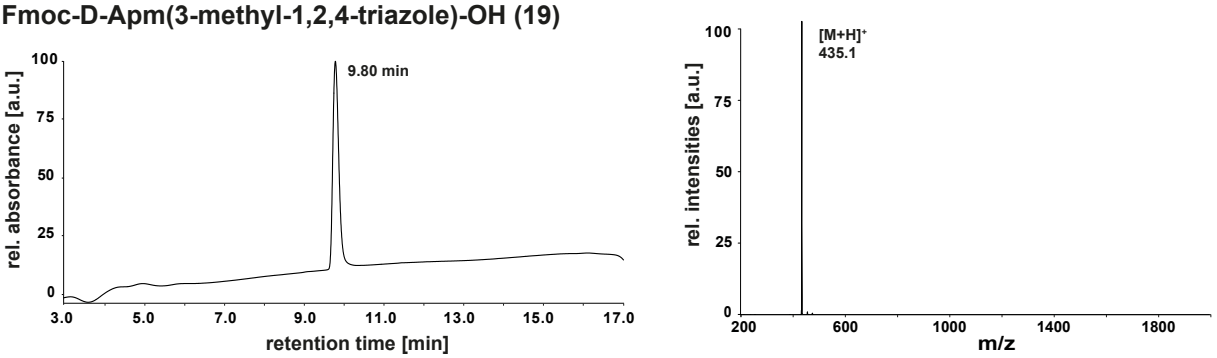
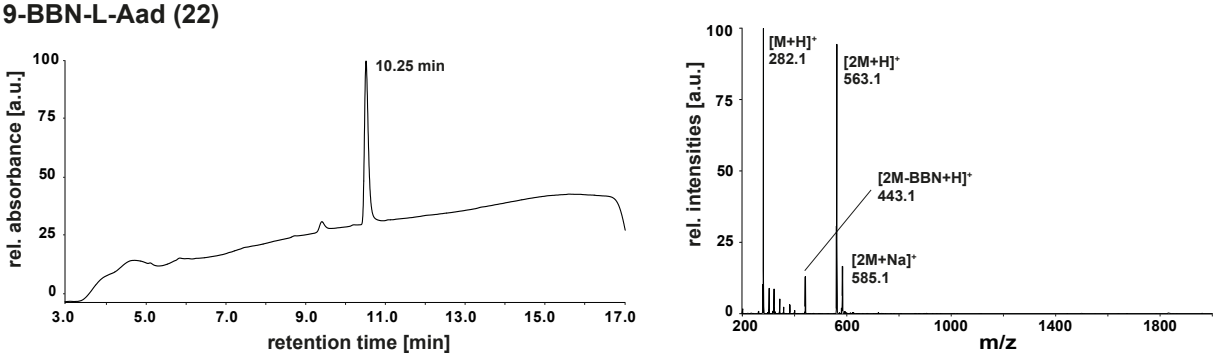
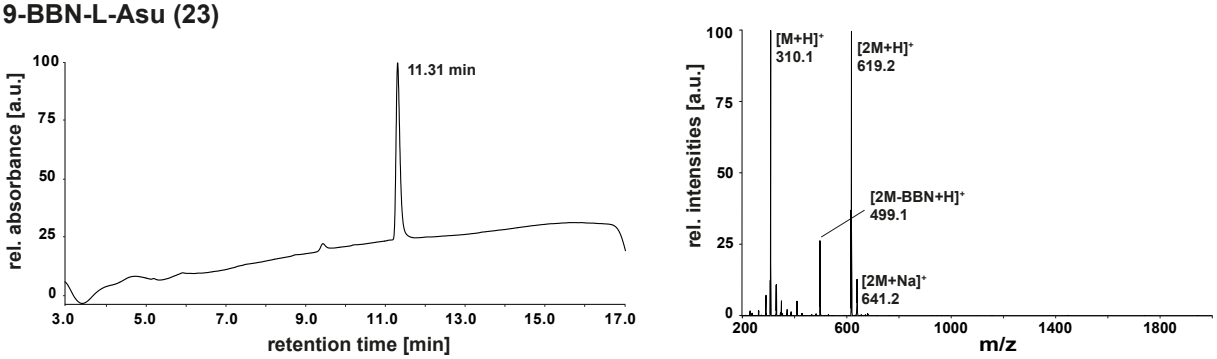


Figure C.4: LC-MS analysis of synthesized building blocks Fmoc-ApmBTri (14), Ac-DL-ApmTri (15), Fmoc-L-ApmTri (17) and Fmoc-D-ApmTri (19).

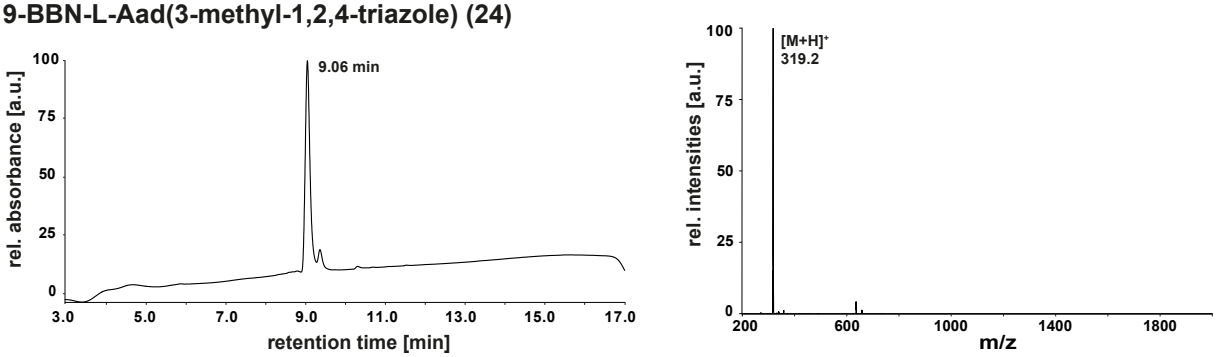
9-BBN-L-Aad (22)



9-BBN-L-Asu (23)



9-BBN-L-Aad(3-methyl-1,2,4-triazole) (24)



9-BBN-L-Asu(3-methyl-1,2,4-triazole) (25)

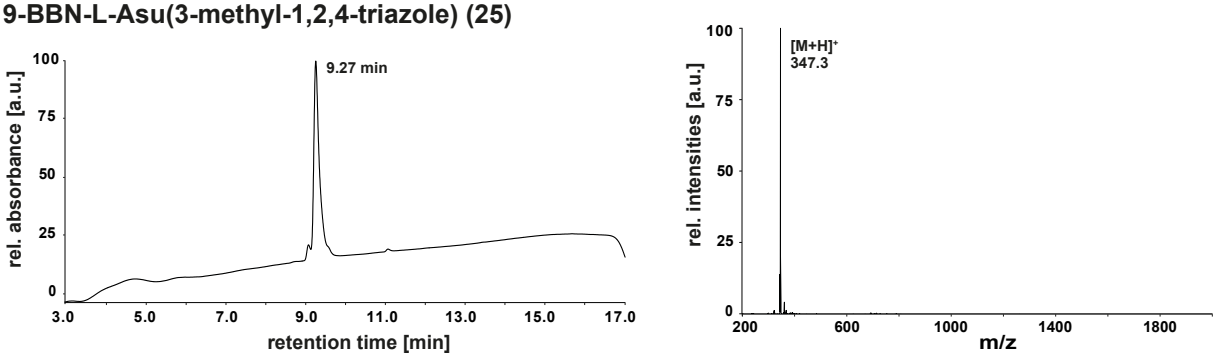
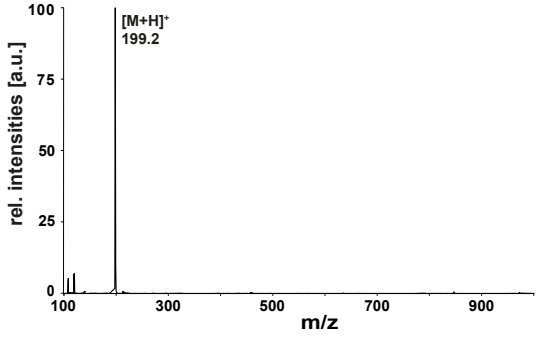
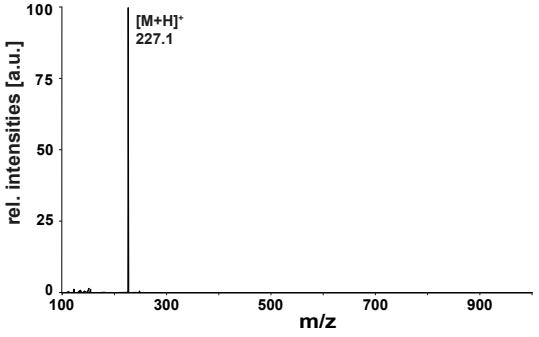


Figure C.5: LC-MS analysis of synthesized building blocks 9-BBN-L-Aad (22), 9-BBN-L-Asu (23), 9-BBN-L-AadTri (24) and 9-BBN-L-AsuTri (25).

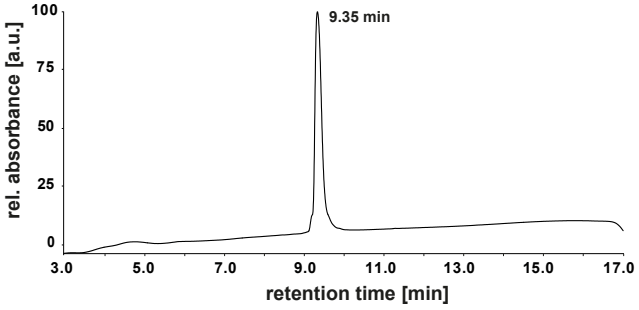
H-L-Aad(3-methyl-1,2,4-triazole)-OH (26)

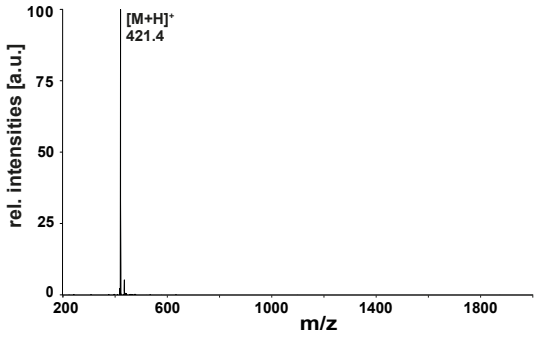


H-L-Asu(3-methyl-1,2,4-triazole)-OH (27)

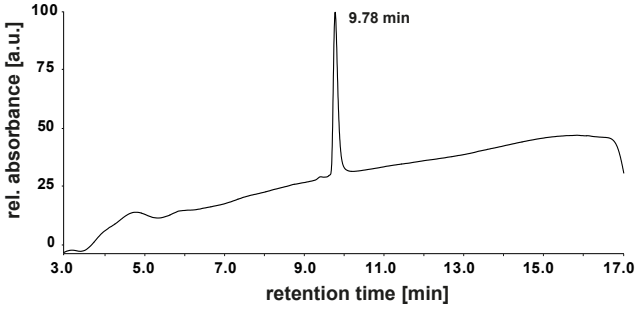


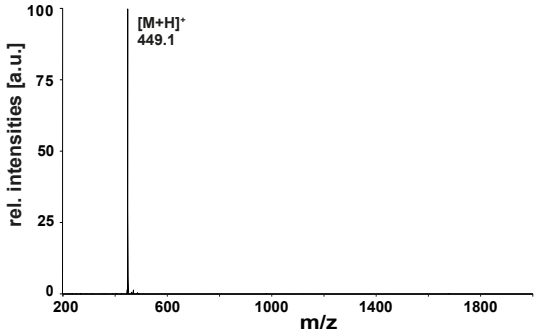
Fmoc-L-Aad(3-methyl-1,2,4-triazole)-OH (28)



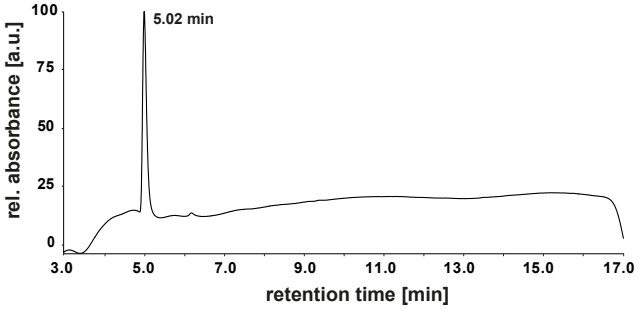


Fmoc-L-Asu(3-methyl-1,2,4-triazole)-OH (29)





Ac-DL-Apm-OH (30)



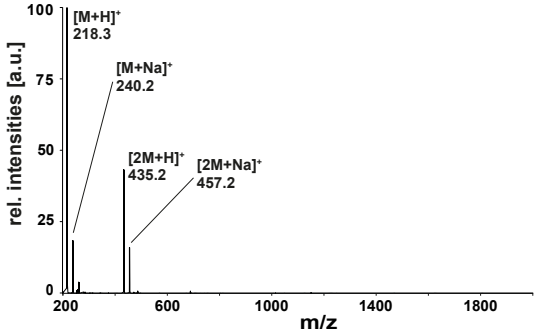
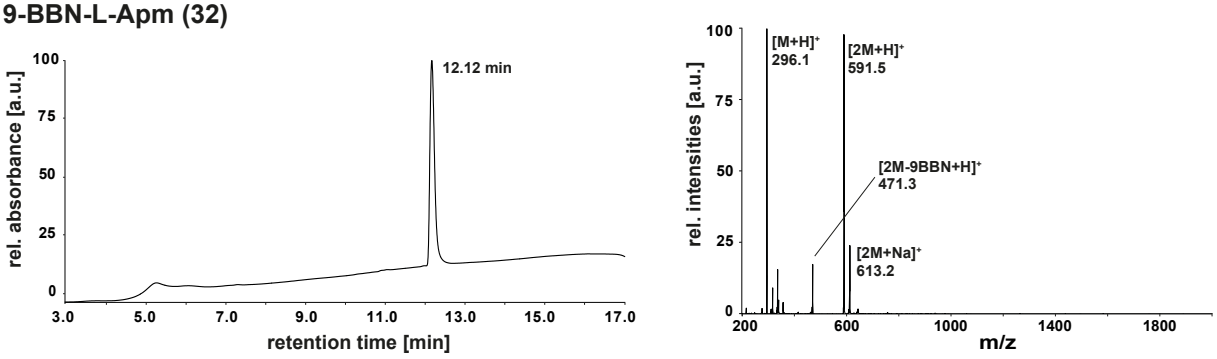
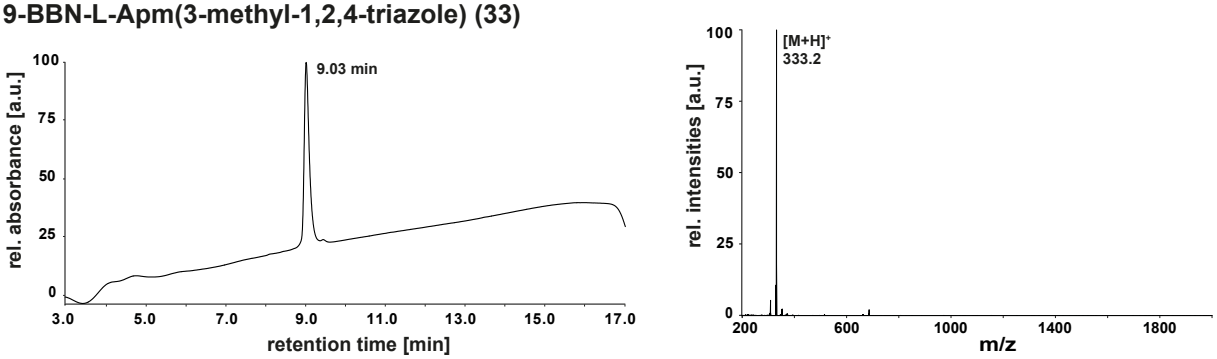


Figure C.6: LC-MS analysis of synthesized building blocks L-AadTri (26), L-AsuTri (27), Fmoc-L-AadTri (28), Fmoc-L-AsuTri (29) and Ac-DL-Apm-OH (30).

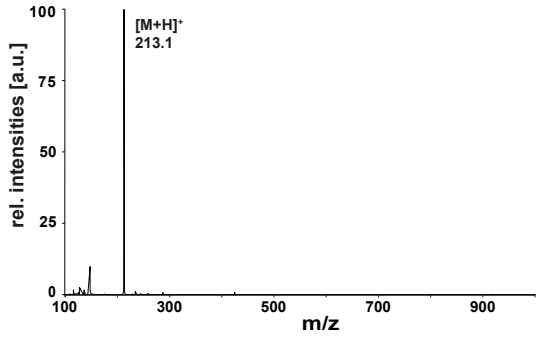
9-BBN-L-Apm (32)



9-BBN-L-Apm(3-methyl-1,2,4-triazole) (33)



H-L-Apm(3-methyl-1,2,4-triazole)-OH (16)



Ac-DL-Asu-OH (35)

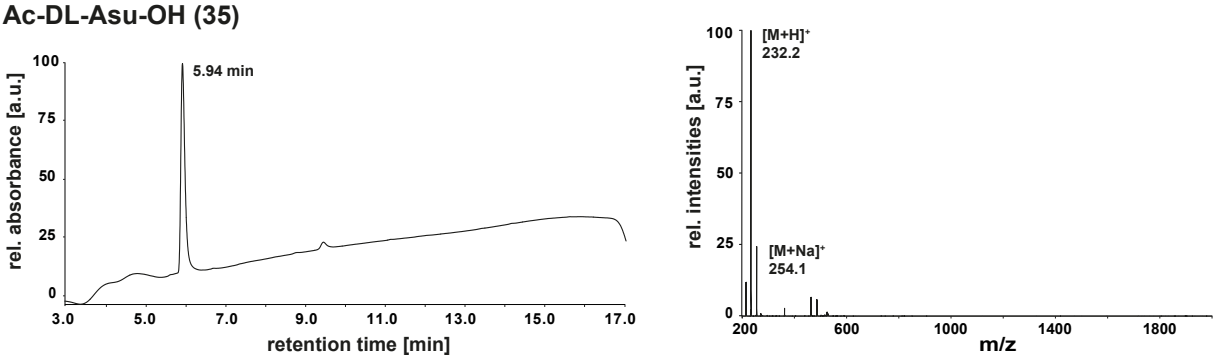
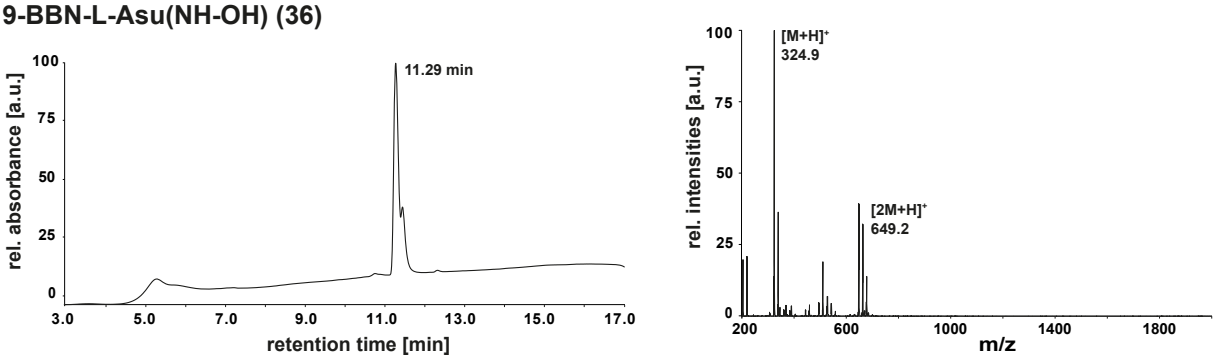
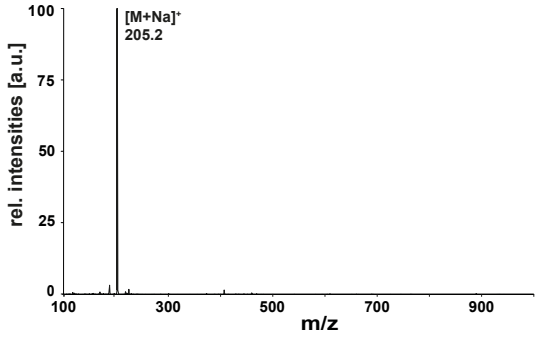


Figure C.7: LC-MS analysis of synthesized building blocks 9-BBN-L-Apm (32), 9-BBN-L-ApmTri (33), L-ApmTri (16) and Ac-DL-Asu-OH (35).

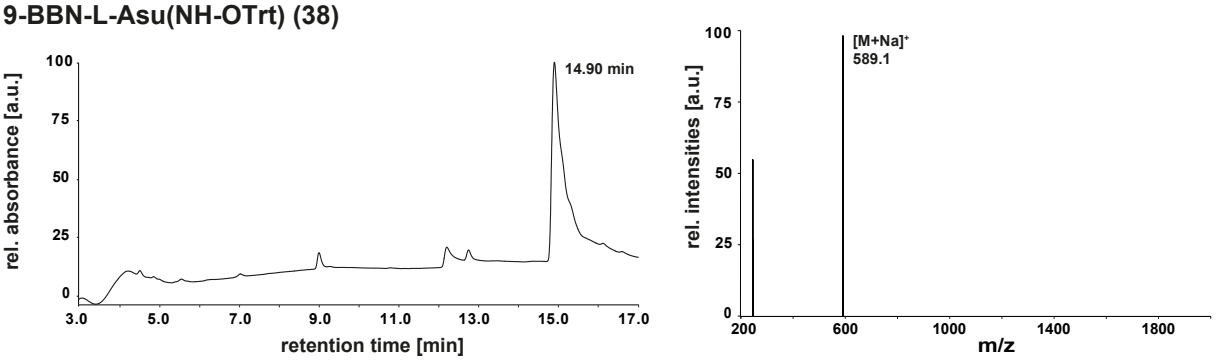
9-BBN-L-Asu(NH-OH) (36)



H-L-Asu(NH-OH)-OH (37)



9-BBN-L-Asu(NH-OTrt) (38)



H-L-Asu(NH-OTrt)-OH (39)

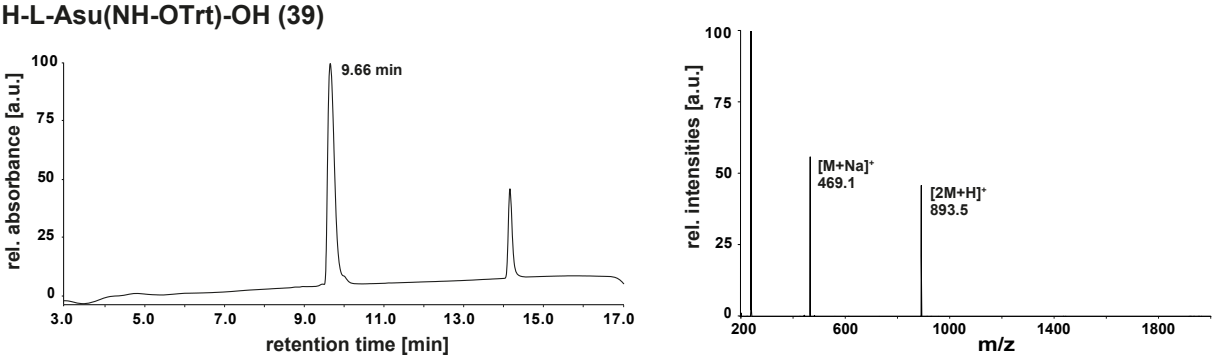


Figure C.8: LC-MS analysis of synthesized building blocks 9-BBN-AsuHd (36), AsuHd (37), 9-BBN-L-Asu(NH-OTrt) (38) and H-L-Asu(NH-OTrt)-OH (39).

Fmoc-L-Asu(NH-OTrt)-OH (40)

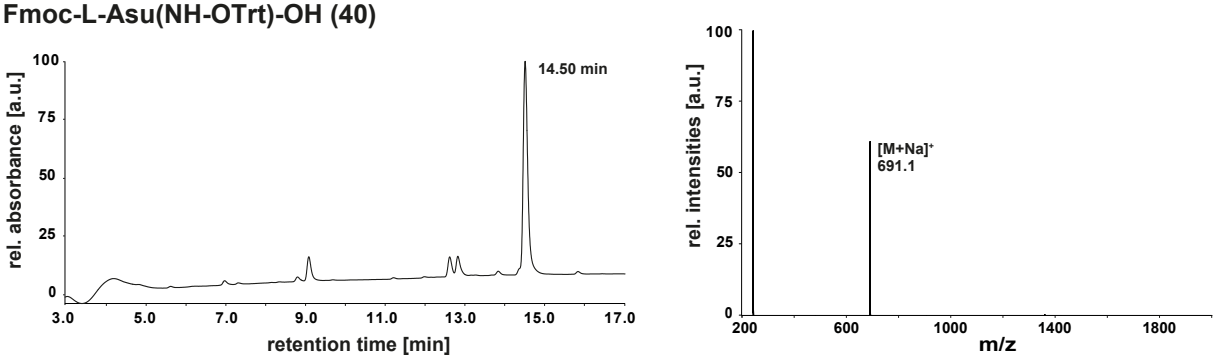
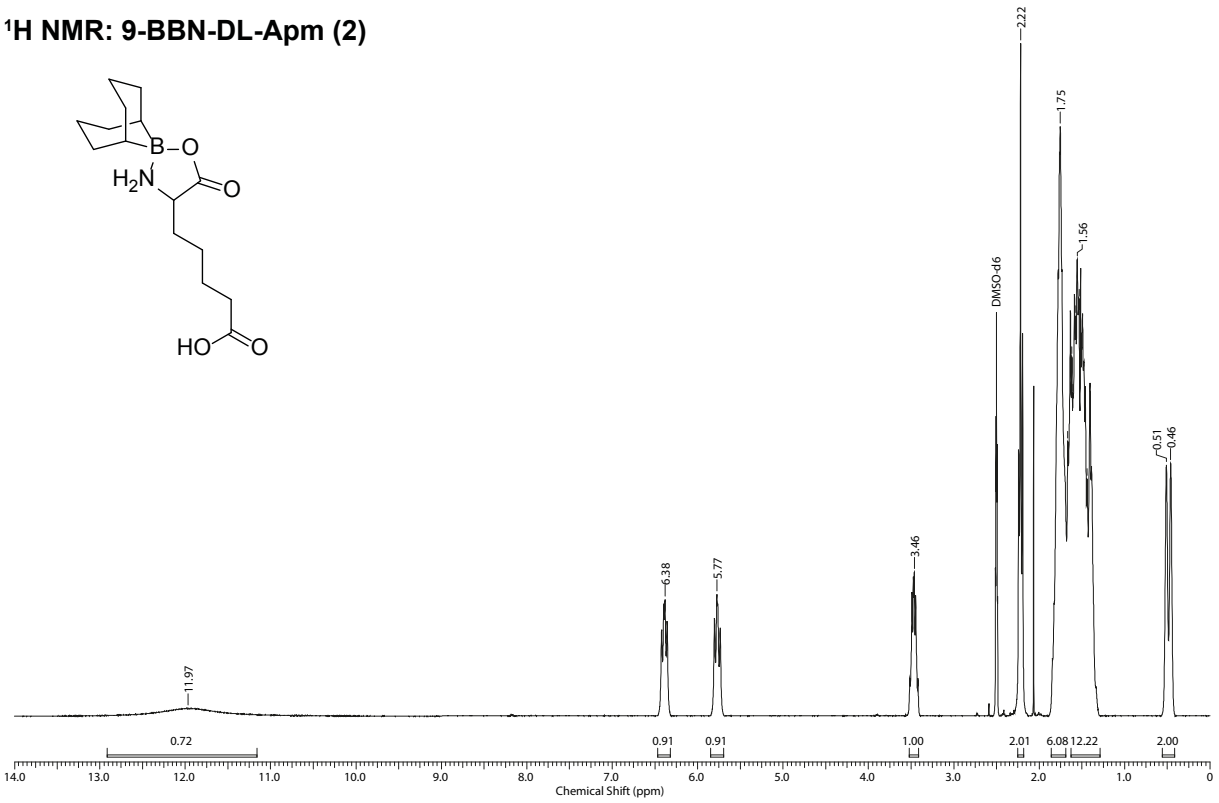


Figure C.9: LC-MS analysis of synthesized building block Fmoc-L-Asu(NH-OTrt)-OH (40).

C.2 NMR spectra

^{1}H NMR: 9-BBN-DL-Apm (2)



^{13}C NMR: 9-BBN-DL-Apm (2)

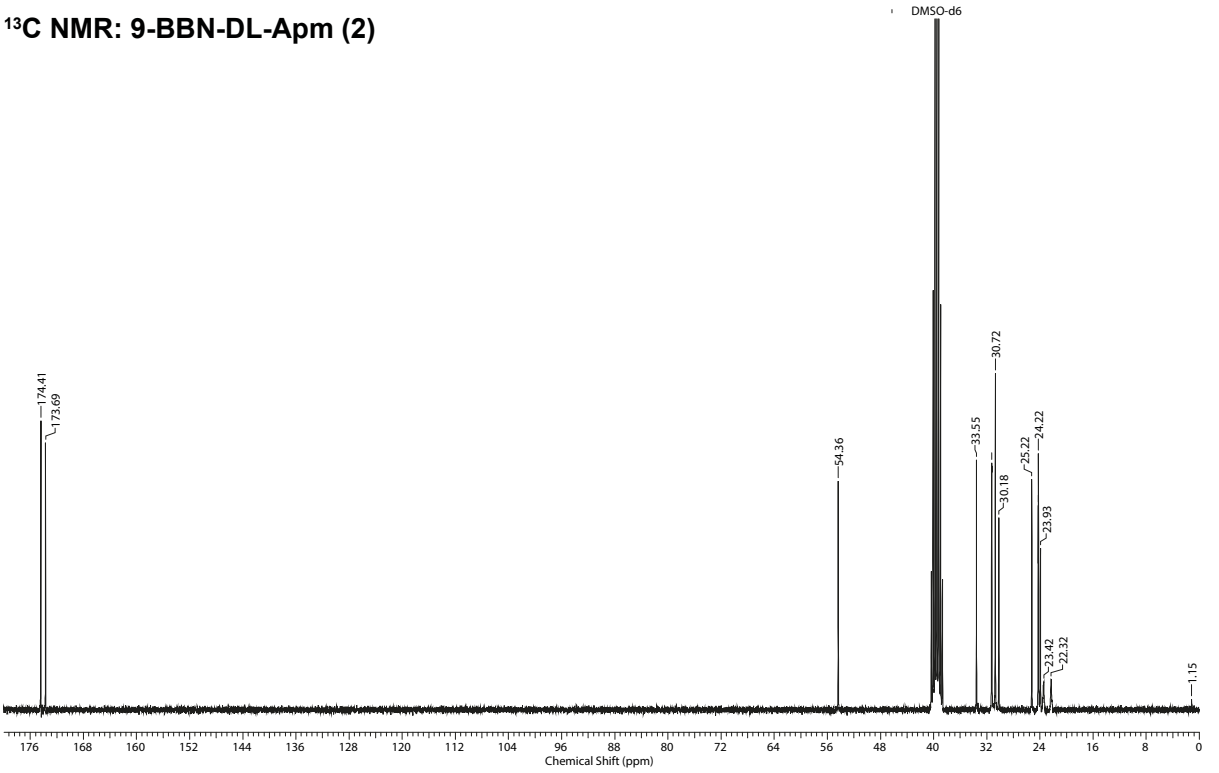
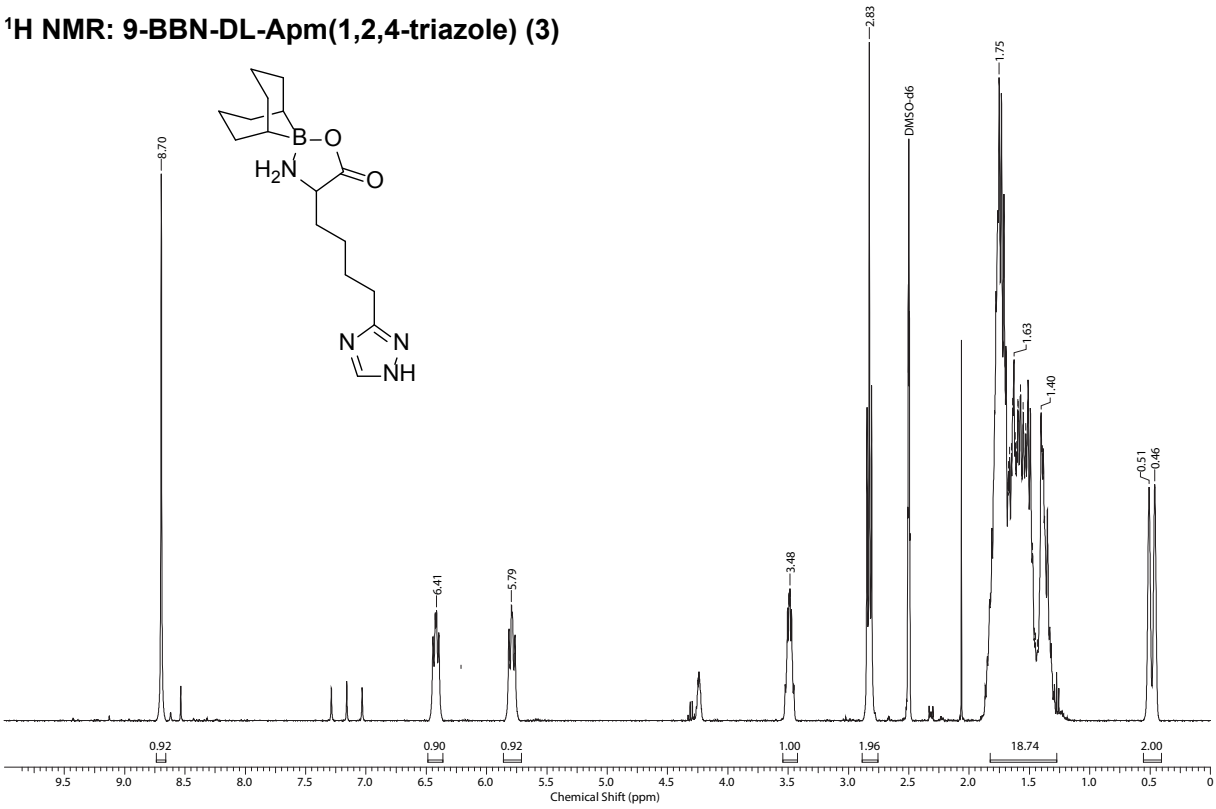


Figure C.10: ^{1}H NMR (300 MHz; DMSO-d_6) and ^{13}C NMR (75 MHz; DMSO-d_6) spectra of 9-BBN-DL-Apm (2).

^{1}H NMR: 9-BBN-DL-Apm(1,2,4-triazole) (3)



^{13}C NMR: 9-BBN-DL-Apm(1,2,4-triazole) (3)

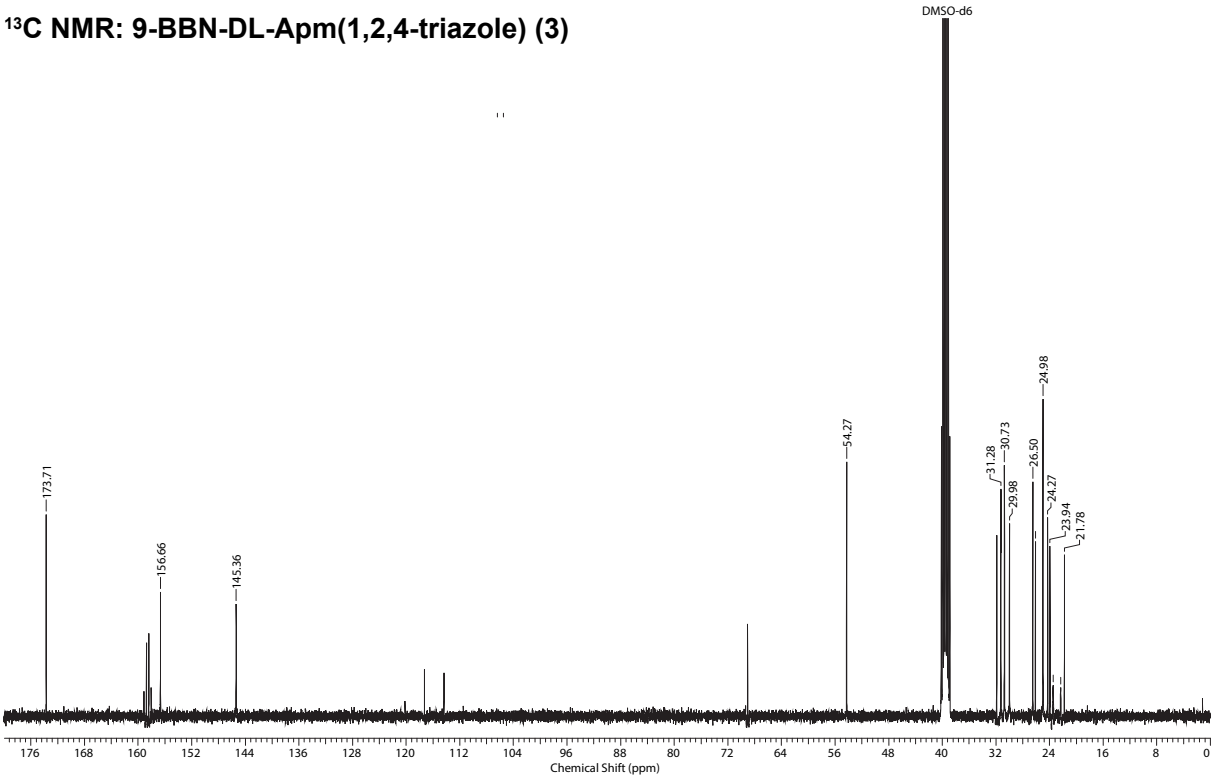
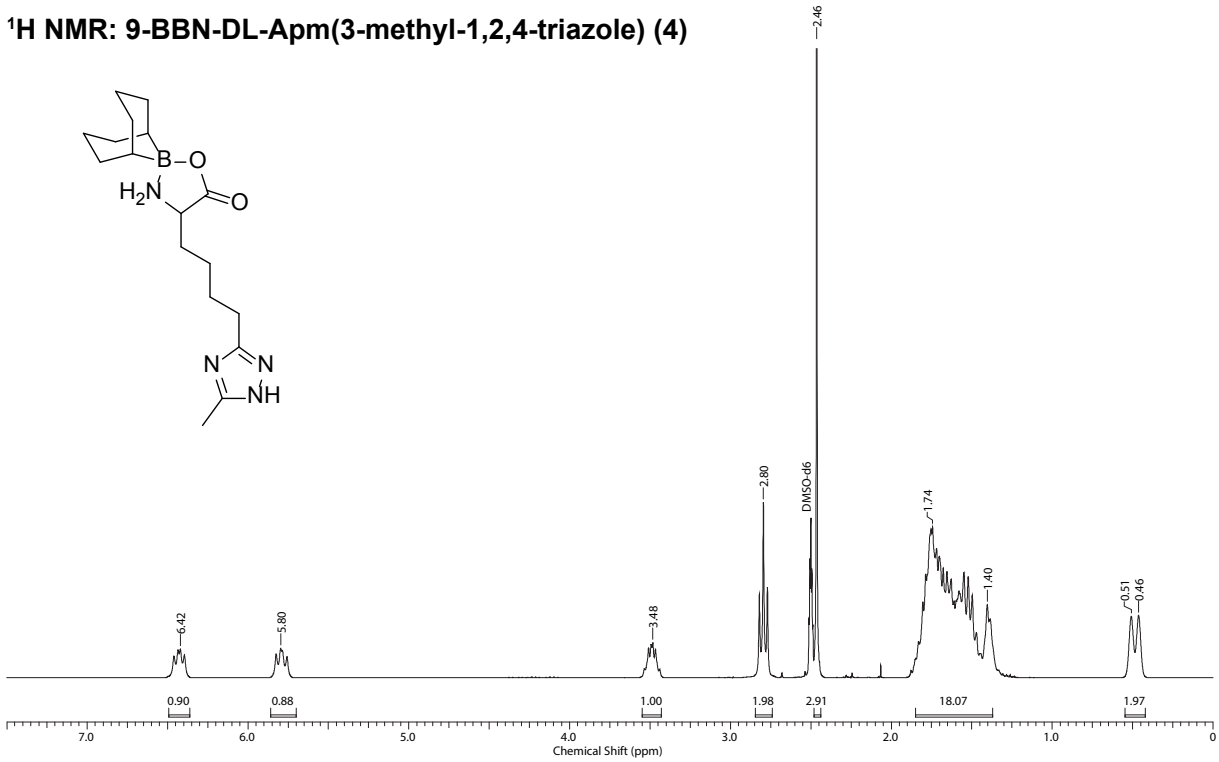


Figure C.11: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (100 MHz; DMSO-d_6) spectra of 9-BBN-DL-Apm(1,2,4-triazole) (9-BBN-DL-ApmFTri) (3).

^{1}H NMR: 9-BBN-DL-Apm(3-methyl-1,2,4-triazole) (4)



^{13}C NMR: 9-BBN-DL-Apm(3-methyl-1,2,4-triazole) (4)

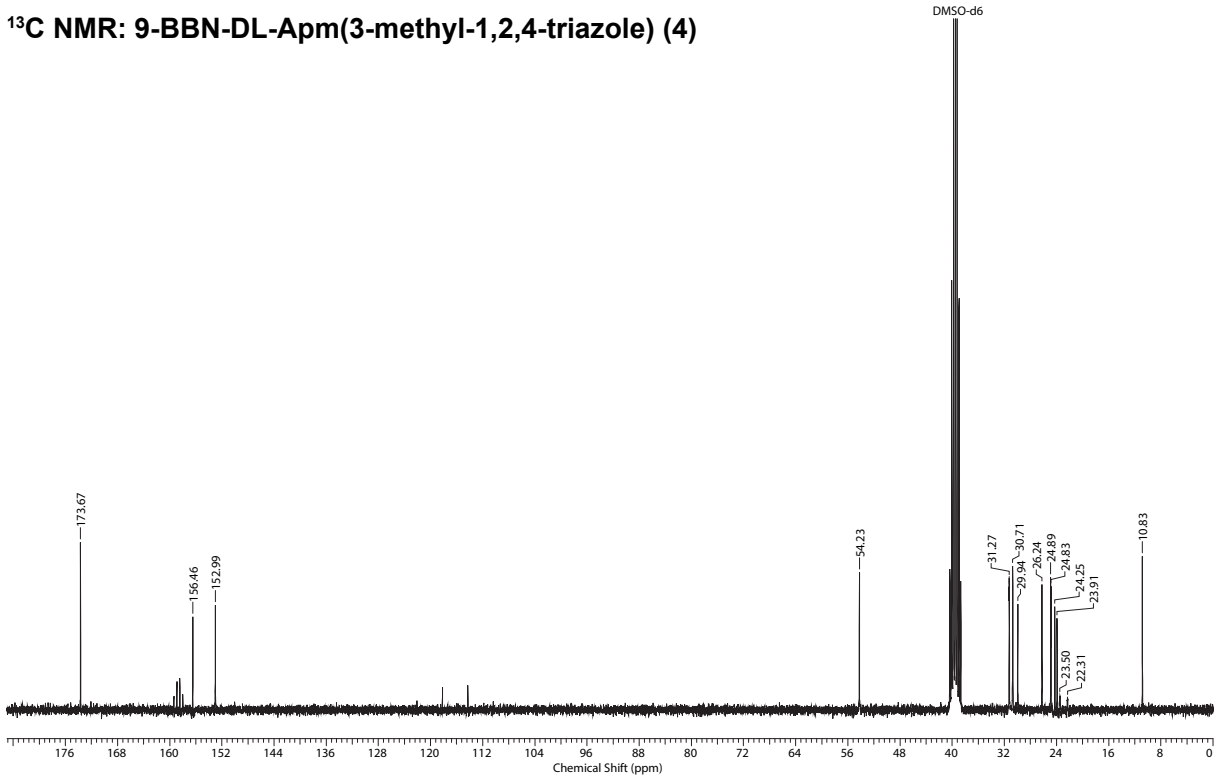
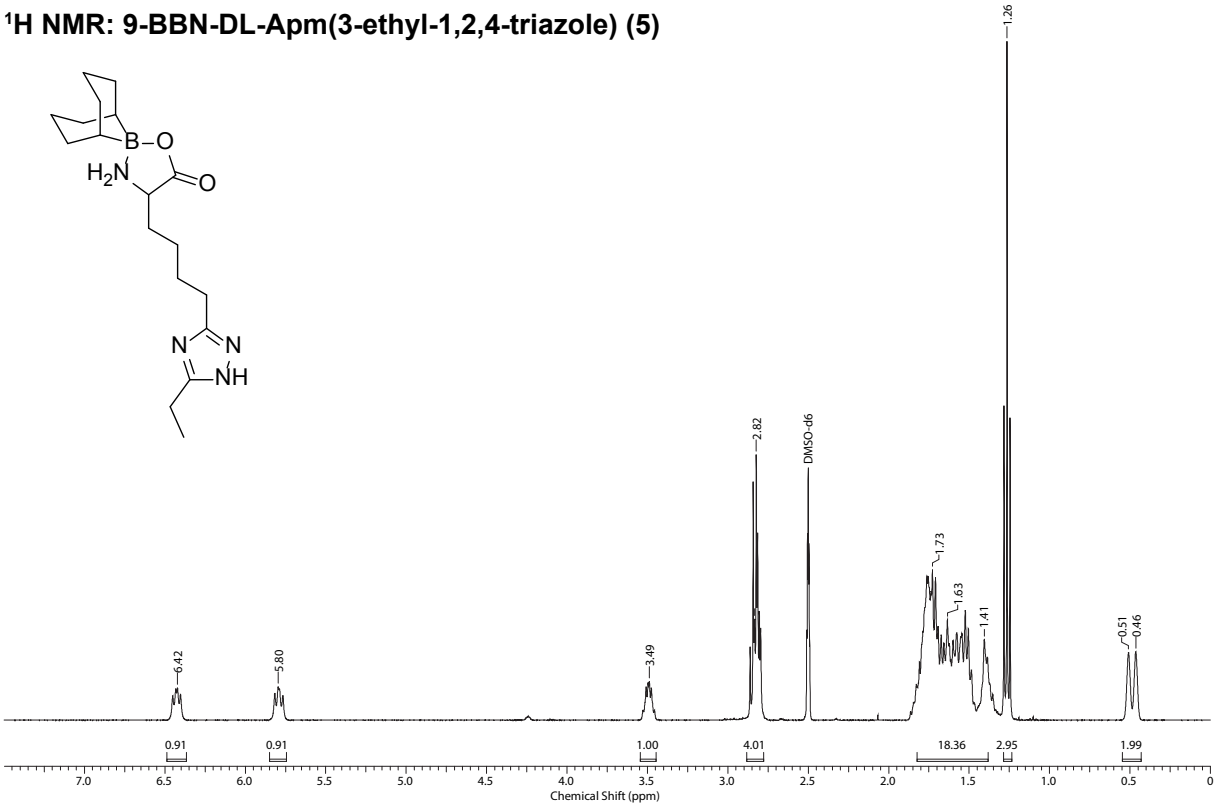


Figure C.12: ^{1}H NMR (300 MHz; DMSO-d_6) and ^{13}C NMR (75 MHz; DMSO-d_6) spectra of 9-BBN-DL-Apm(3-methyl-1,2,4-triazole) (9-BBN-DL-ApmTri) (4).

^{1}H NMR: 9-BBN-DL-Apm(3-ethyl-1,2,4-triazole) (5)



^{13}C NMR: 9-BBN-DL-Apm(3-ethyl-1,2,4-triazole) (5)

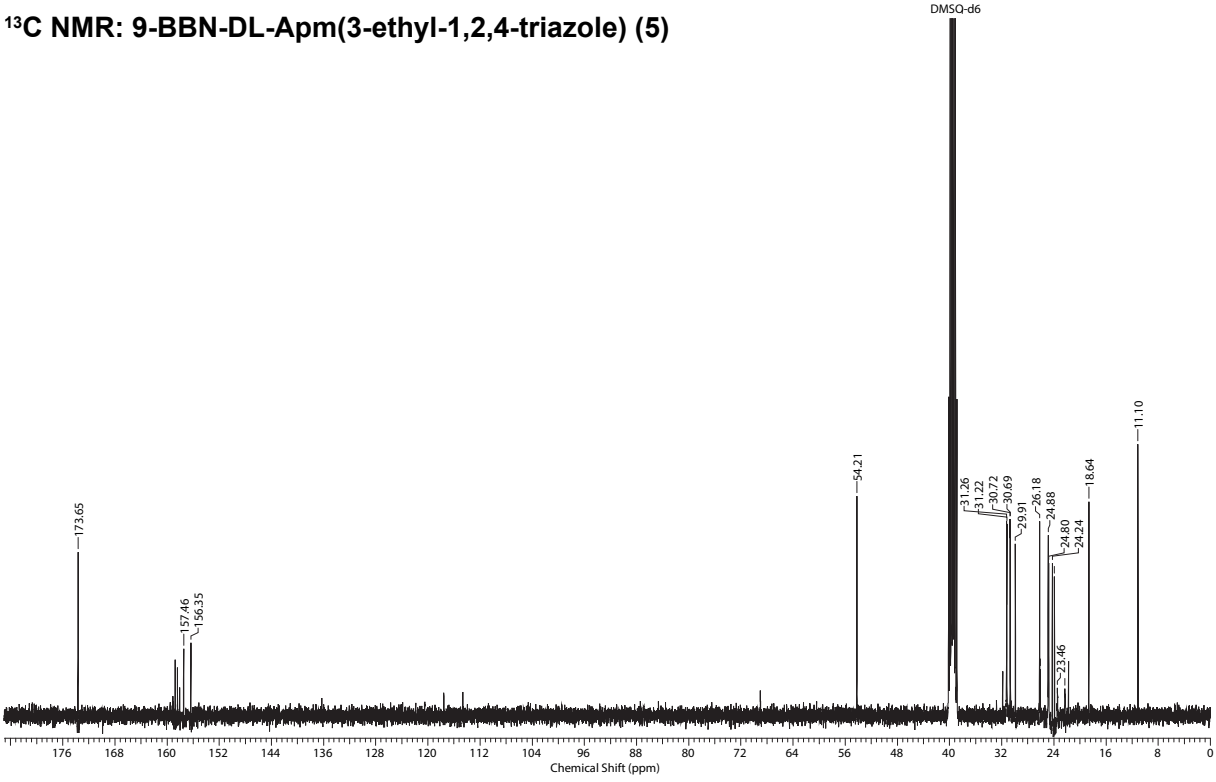
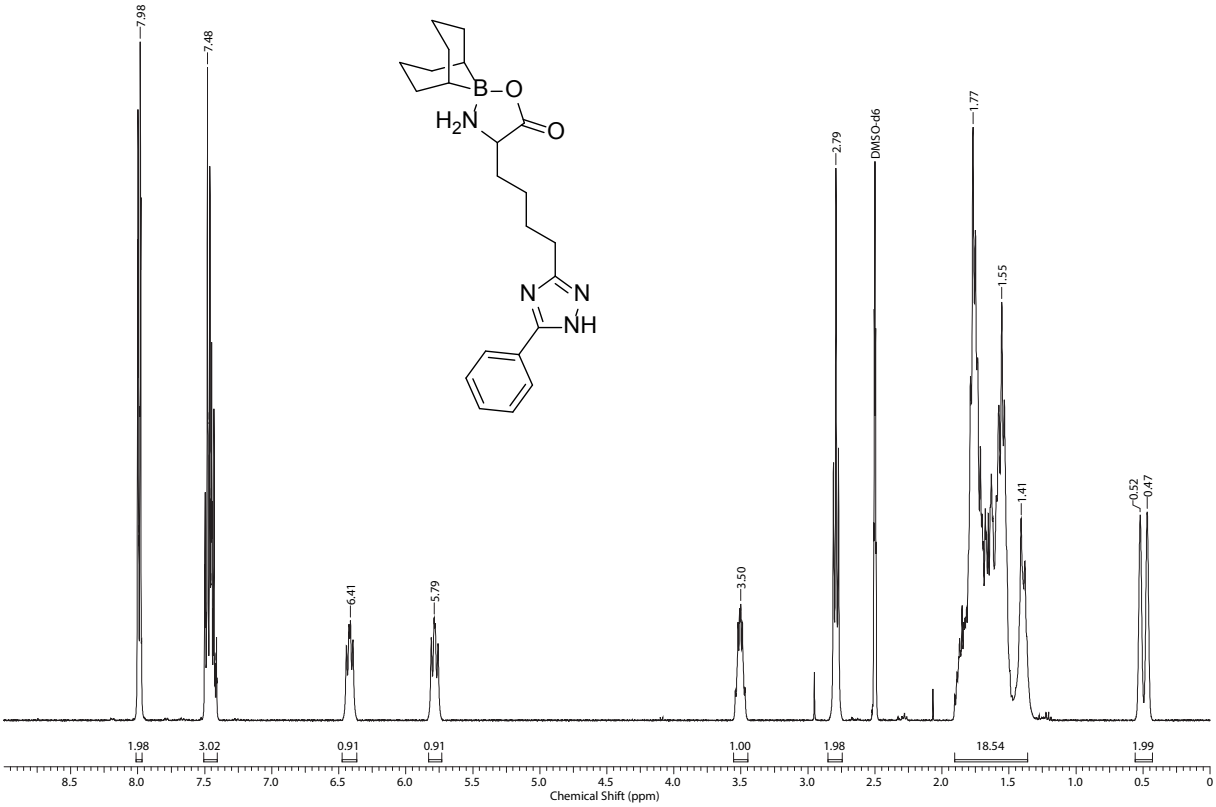


Figure C.13: ^{1}H NMR (300 MHz; DMSO-d_6) and ^{13}C NMR (75 MHz; DMSO-d_6) spectra of 9-BBN-DL-Apm(3-ethyl-1,2,4-triazole) (9-BBN-DL-ApmPTri) (5).

^{1}H NMR: 9-BBN-DL-Apm(3-phenyl-1,2,4-triazole) (6)



^{13}C NMR: 9-BBN-DL-Apm(3-phenyl-1,2,4-triazole) (6)

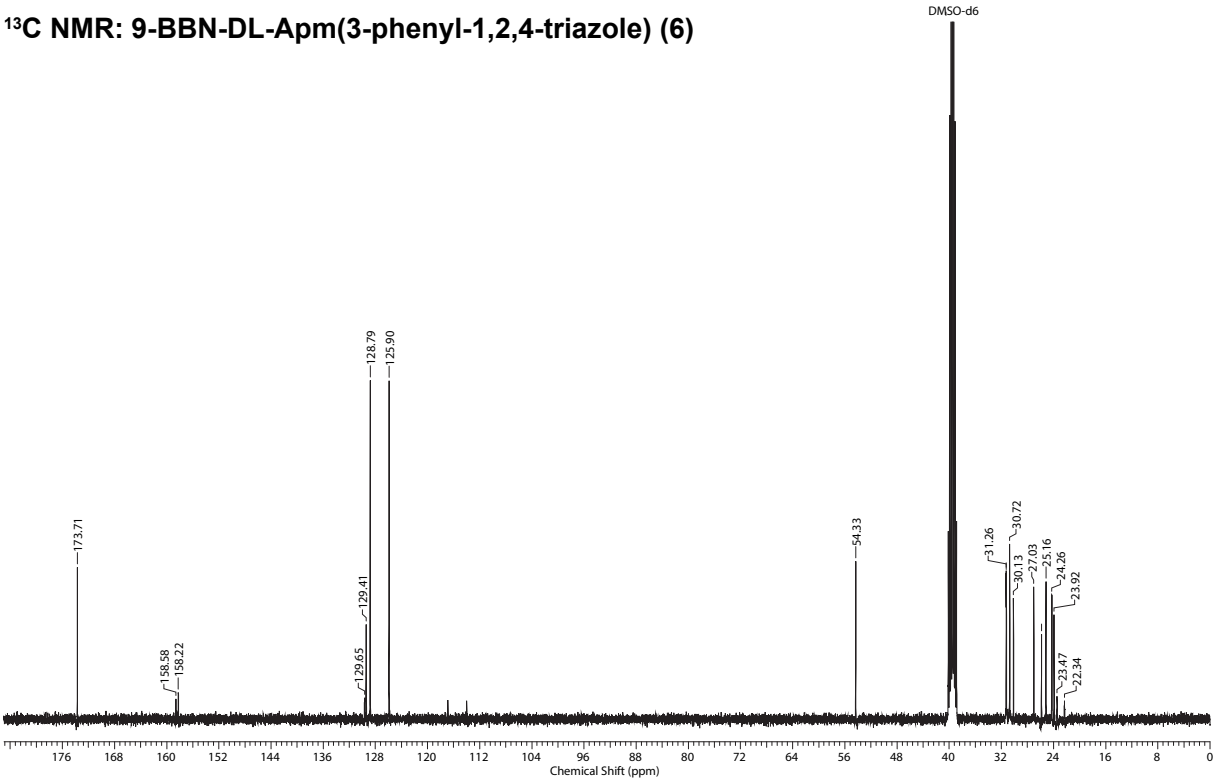
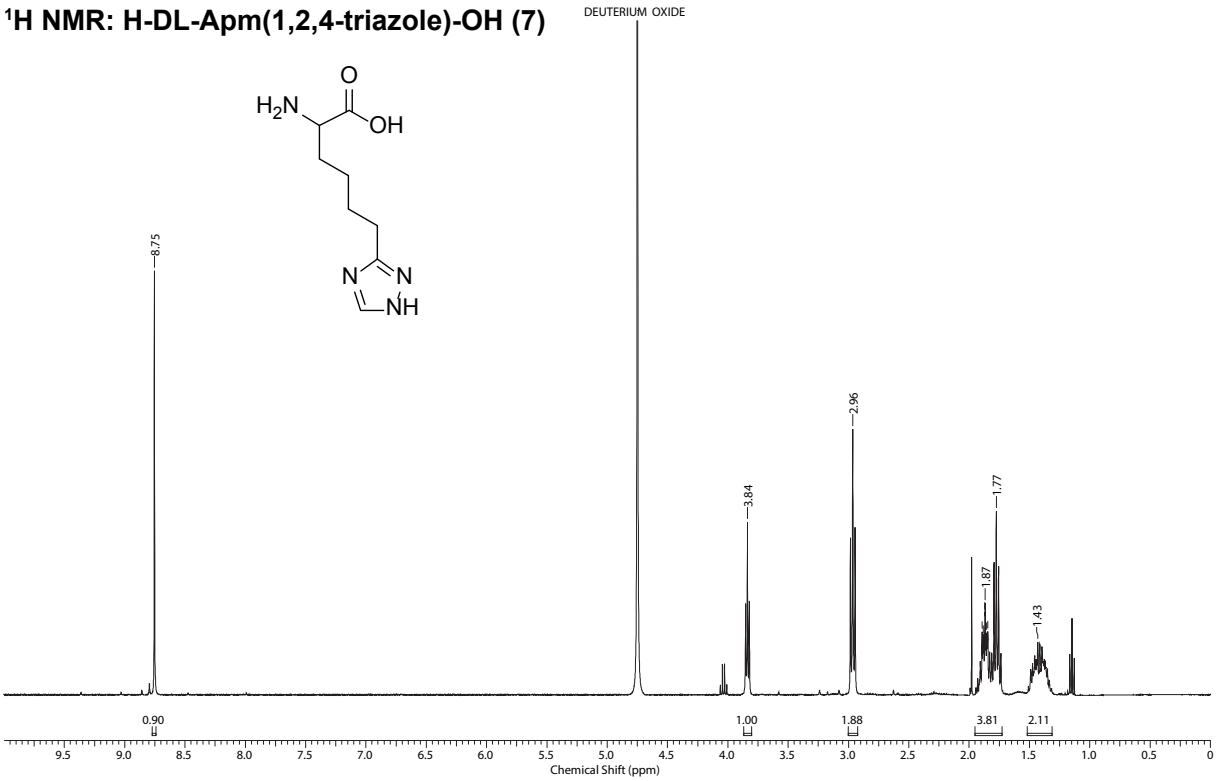


Figure C.14: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (100 MHz; DMSO-d_6) spectra of 9-BBN-DL-Apm(3-phenyl-1,2,4-triazole) (9-BBN-DL-ApmBTri) (6).

^{1}H NMR: H-DL-Apm(1,2,4-triazole)-OH (7)



^{13}C NMR: H-DL-Apm(1,2,4-triazole)-OH (7)

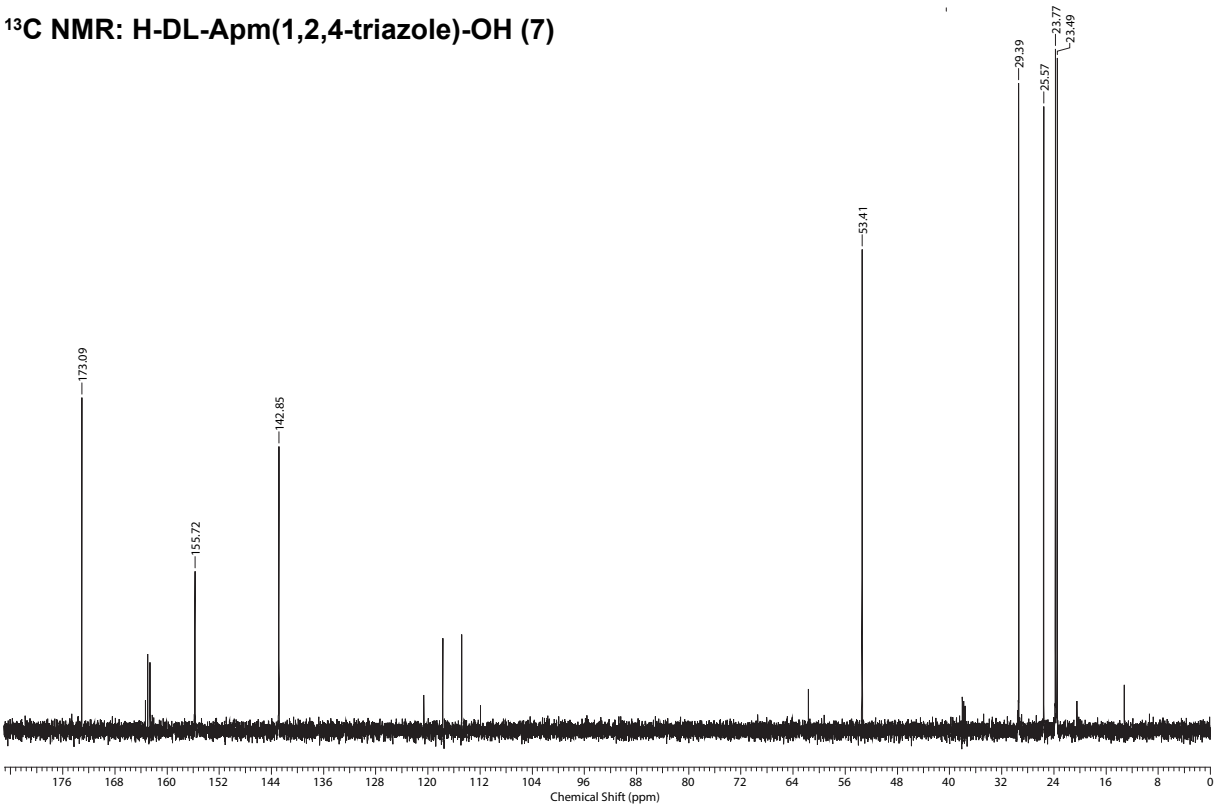
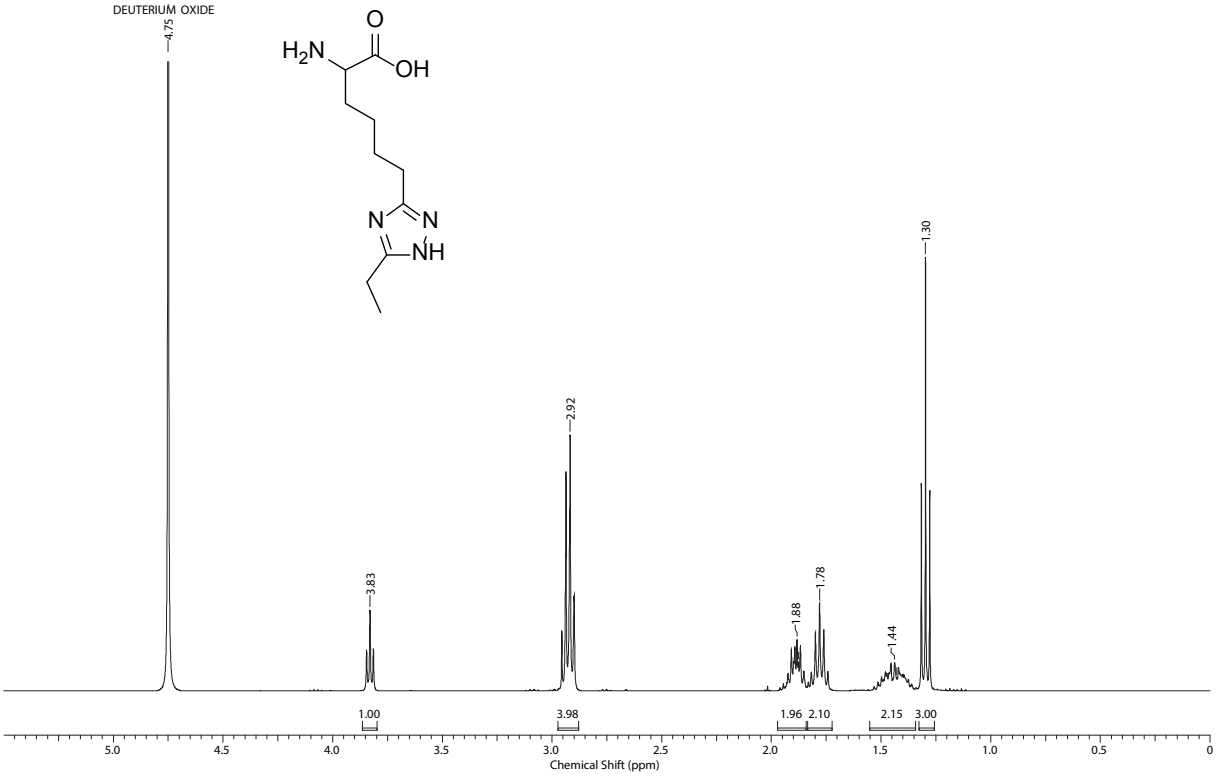


Figure C.15: ^{1}H NMR (400 MHz; D_2O) and ^{13}C NMR (100 MHz; D_2O) spectra of H-DL-Apm(1,2,4-triazole) (ApmFTri) (7).

^{1}H NMR: H-DL-Apm(3-ethyl-1,2,4-triazole)-OH (9)



^{13}C NMR: H-DL-Apm(3-ethyl-1,2,4-triazole)-OH (9)

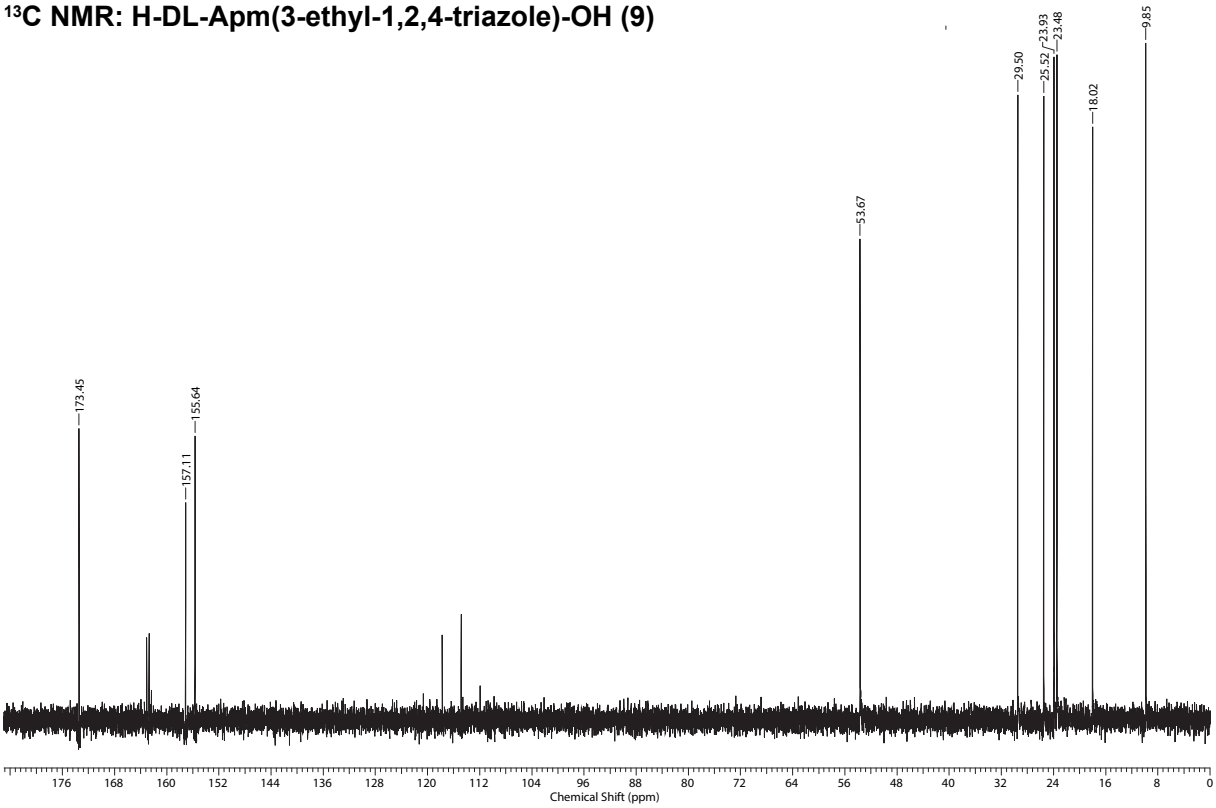
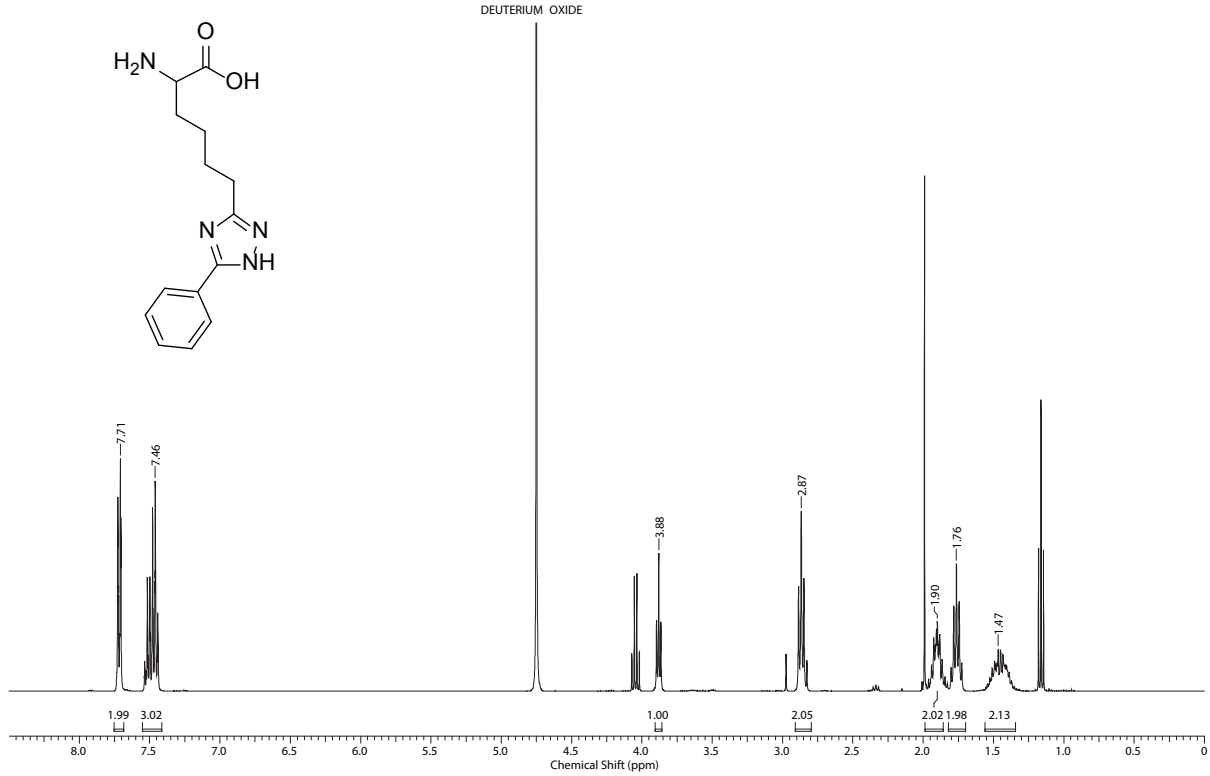


Figure C.16: ^{1}H NMR (400 MHz; D_2O) and ^{13}C NMR (100 MHz; D_2O) spectra of H-DL-Apm(3-ethyl-1,2,4-triazole)-OH (ApmPTri) (9).

^{1}H NMR: H-DL-Apm(3-phenyl-1,2,4-triazole)-OH (10)



^{13}C NMR: H-DL-Apm(3-phenyl-1,2,4-triazole)-OH (10)

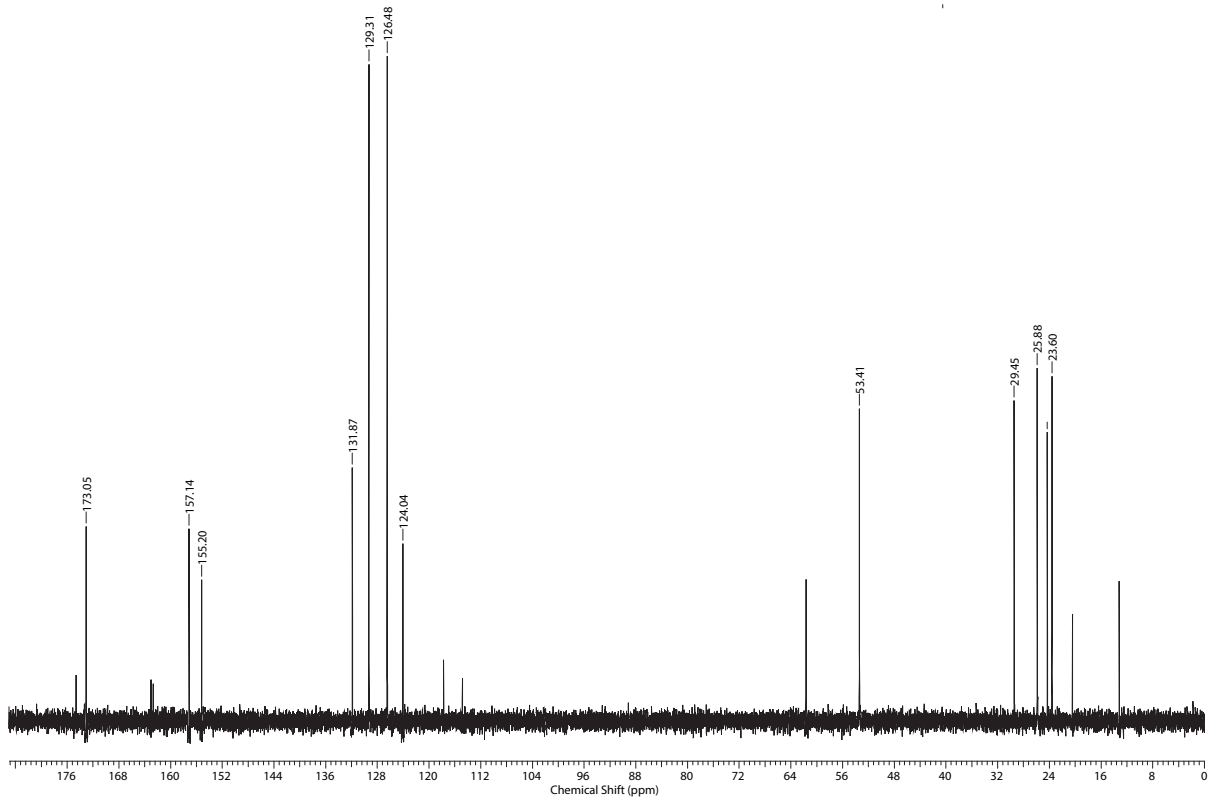
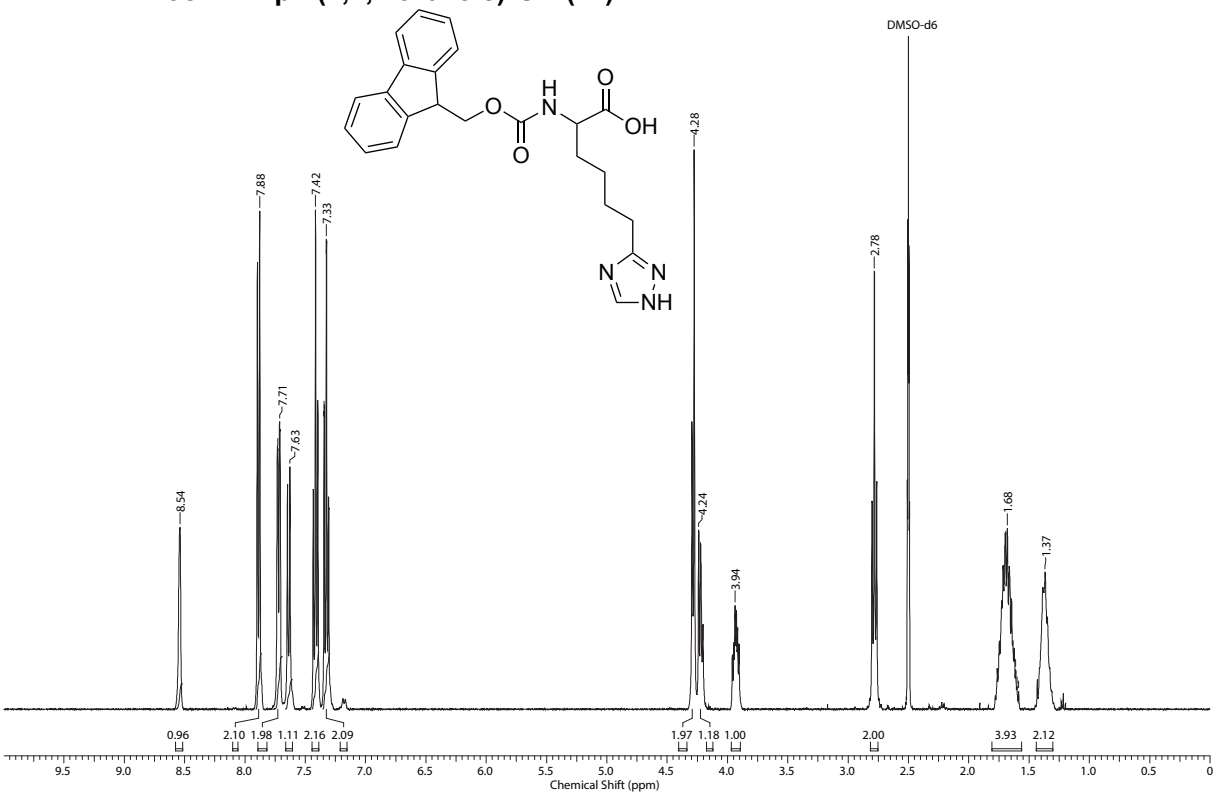


Figure C.17: ^{1}H NMR (400 MHz; D_2O) and ^{13}C NMR (100 MHz; D_2O) spectra of H-DL-Apm(3-phenyl-1,2,4-triazole)-OH (ApmBTri) (10).

^{1}H NMR: Fmoc-DL-Apm(1,2,4-triazole)-OH (11)



^{13}C NMR: Fmoc-DL-Apm(1,2,4-triazole)-OH (11)

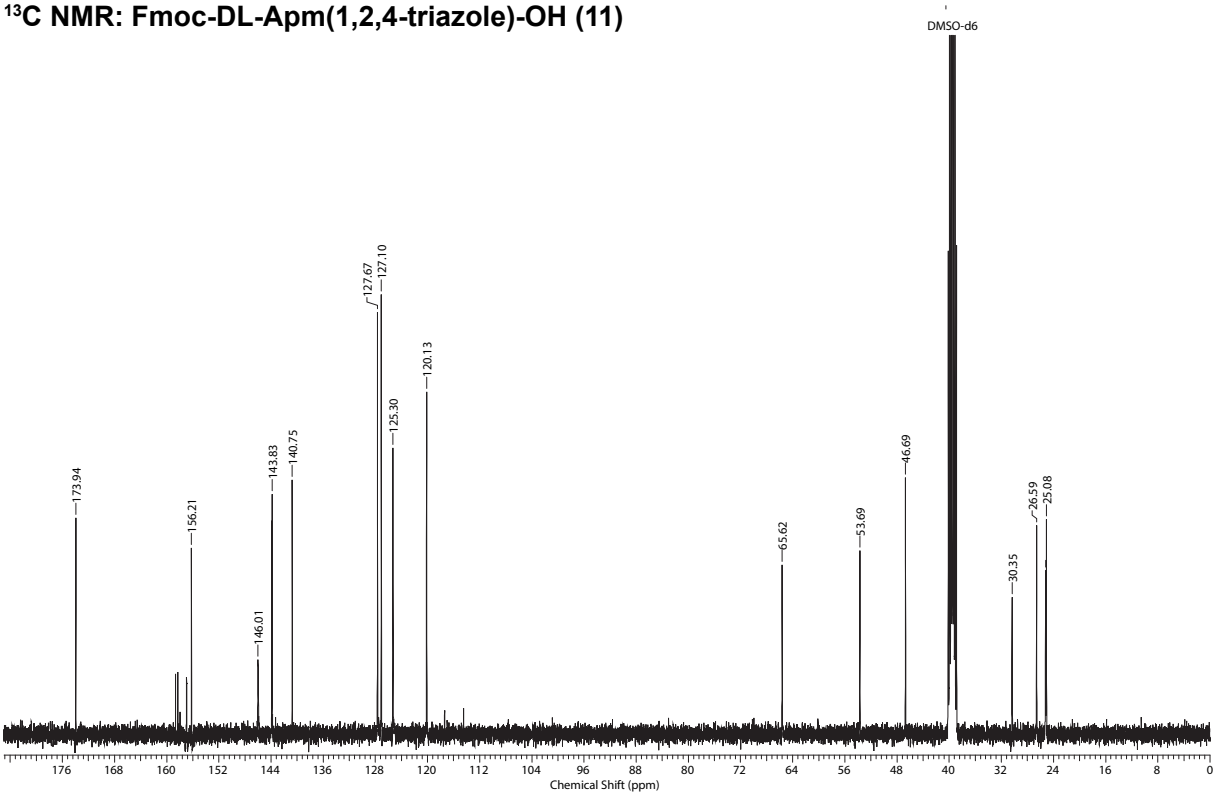
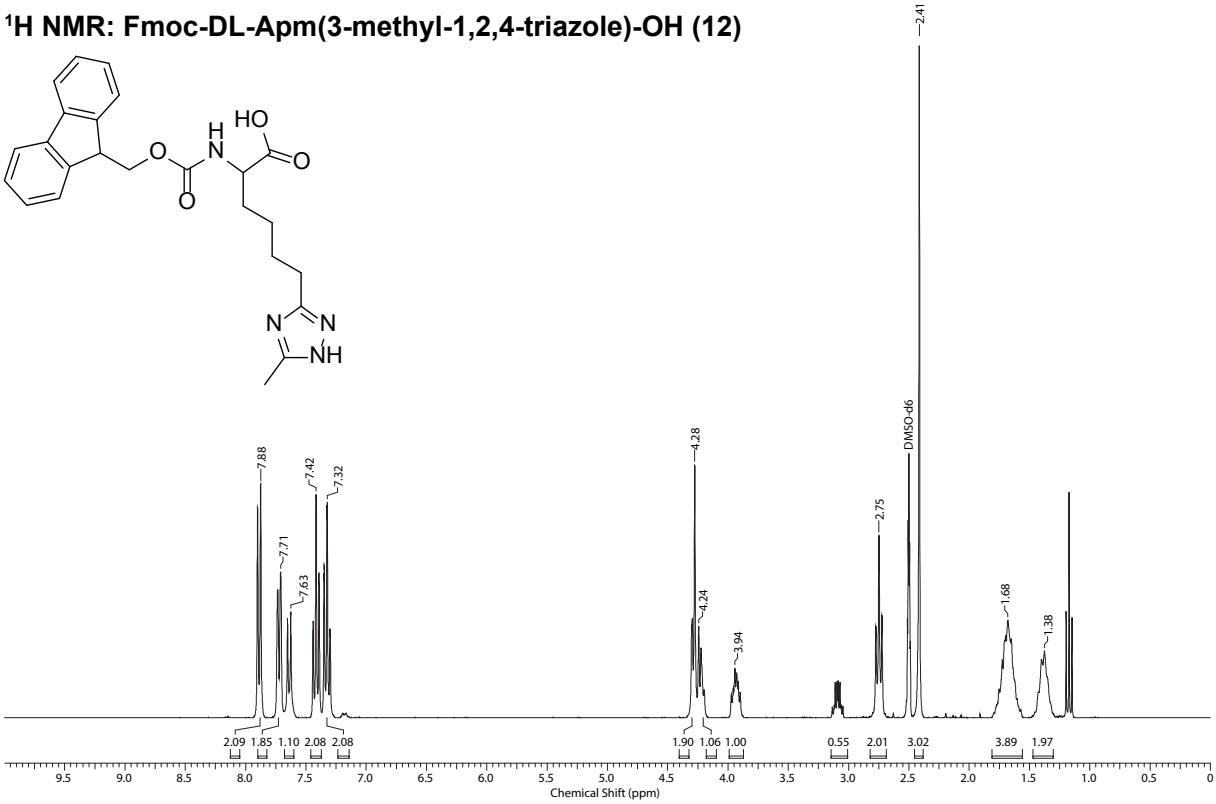


Figure C.18: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (101 MHz; DMSO-d_6) spectra of Fmoc-DL-Apm(1,2,4-triazole)-OH (Fmoc-ApmFTri) (11).

^{1}H NMR: Fmoc-DL-Apm(3-methyl-1,2,4-triazole)-OH (12)



^{13}C NMR: Fmoc-DL-Apm(3-methyl-1,2,4-triazole)-OH (12)

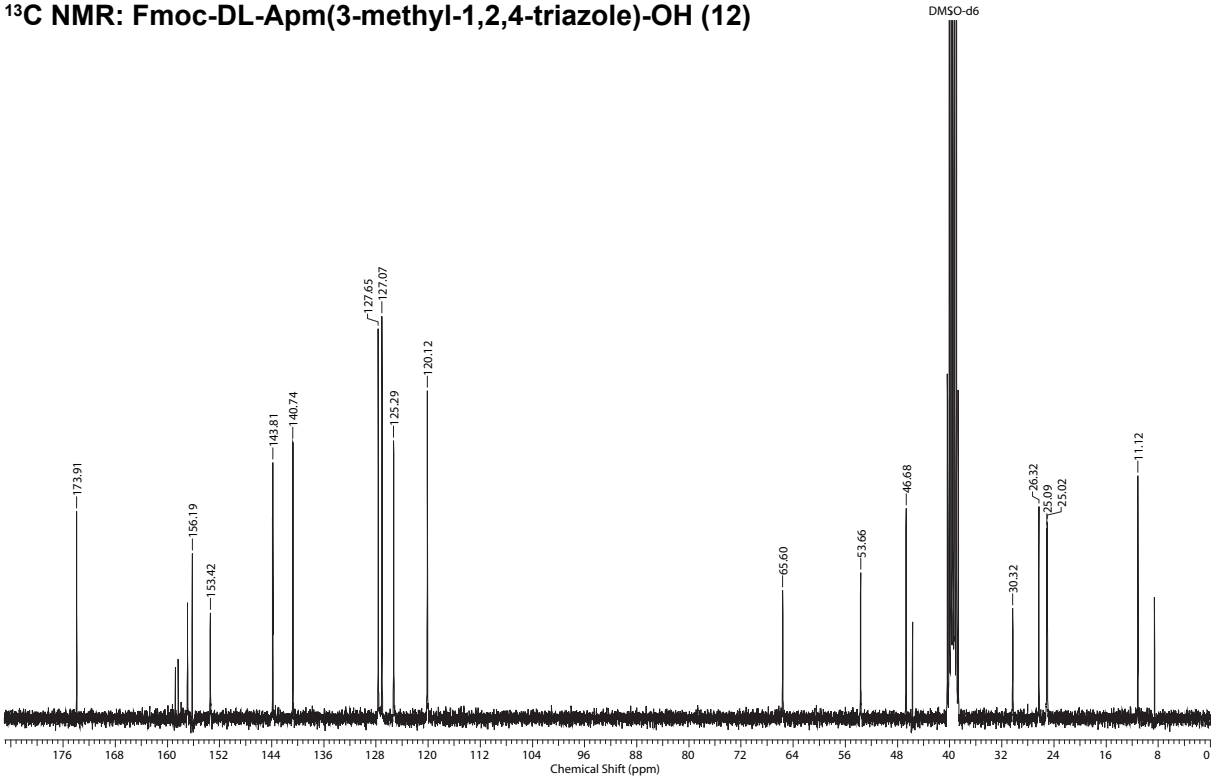
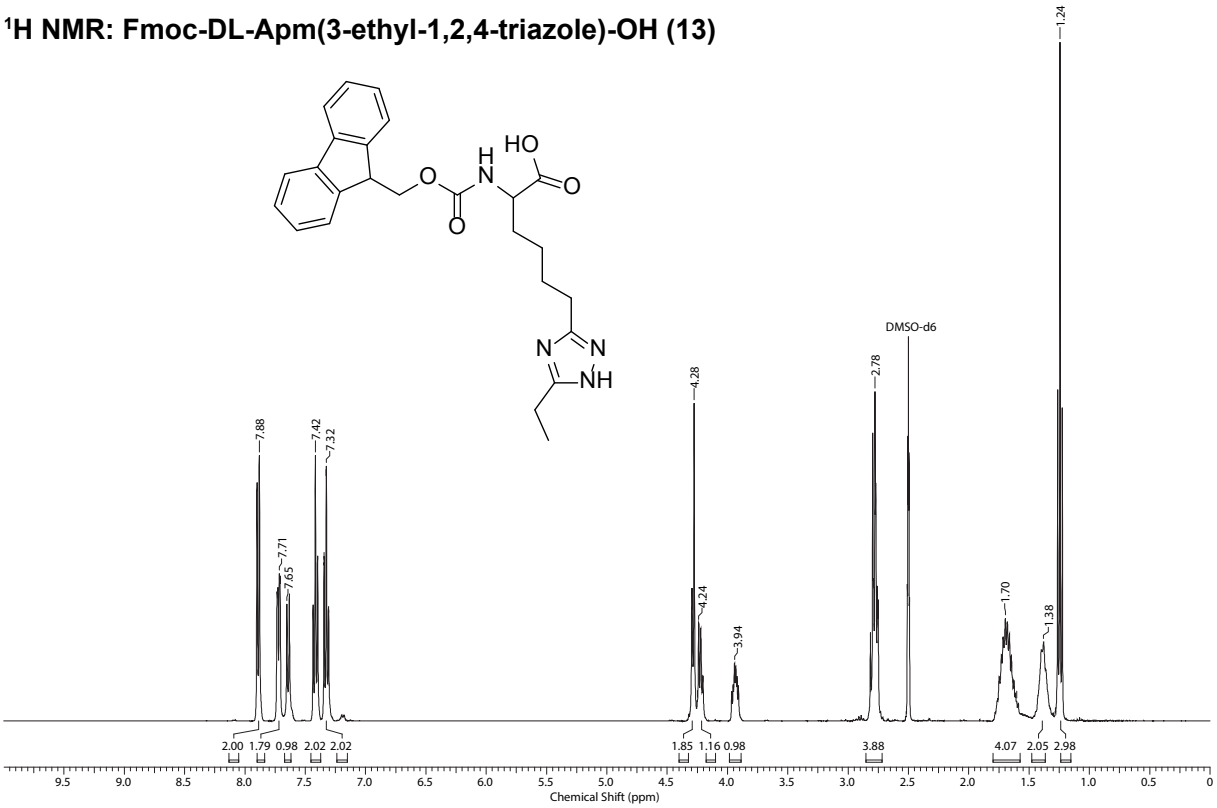


Figure C.19: ^{1}H NMR (300 MHz; DMSO-d_6) and ^{13}C NMR (75 MHz; DMSO-d_6) spectra of Fmoc-DL-Apm(3-methyl-1,2,4-triazole)-OH (Fmoc-ApmTri) (12).

^{1}H NMR: Fmoc-DL-Apm(3-ethyl-1,2,4-triazole)-OH (13)



^{13}C NMR: Fmoc-DL-Apm(3-ethyl-1,2,4-triazole)-OH (13)

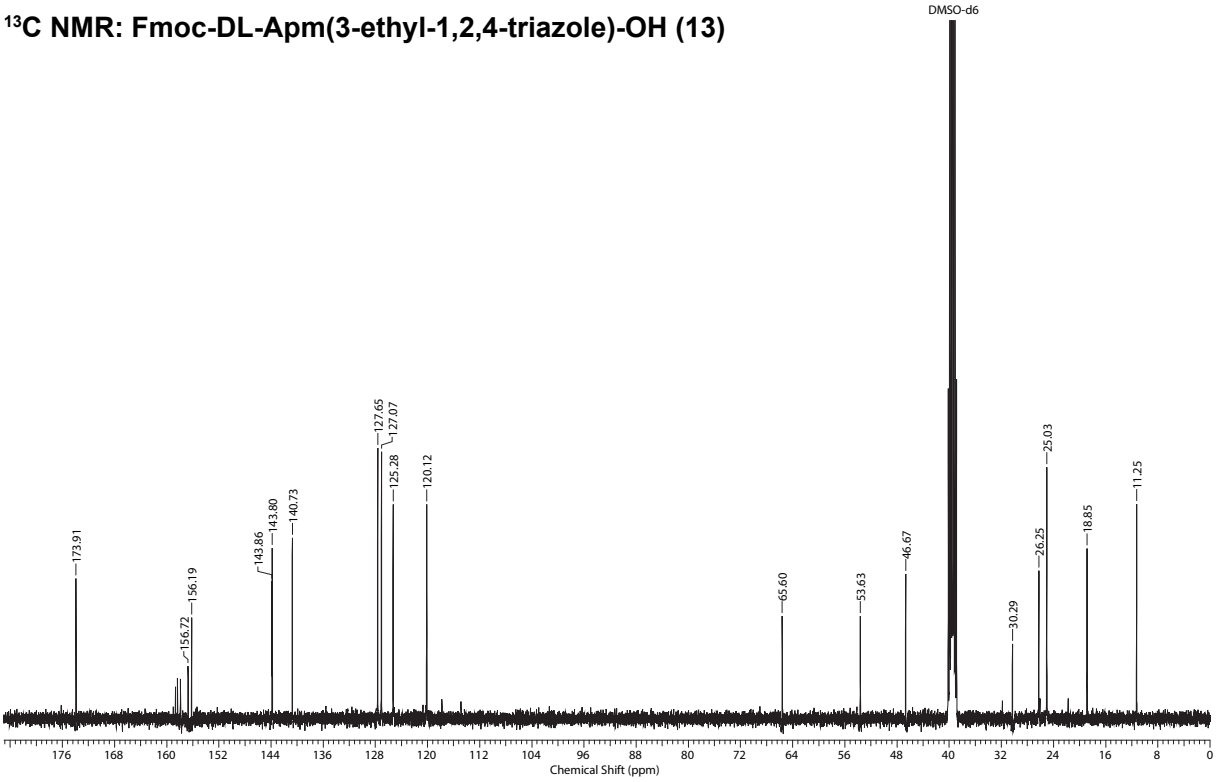
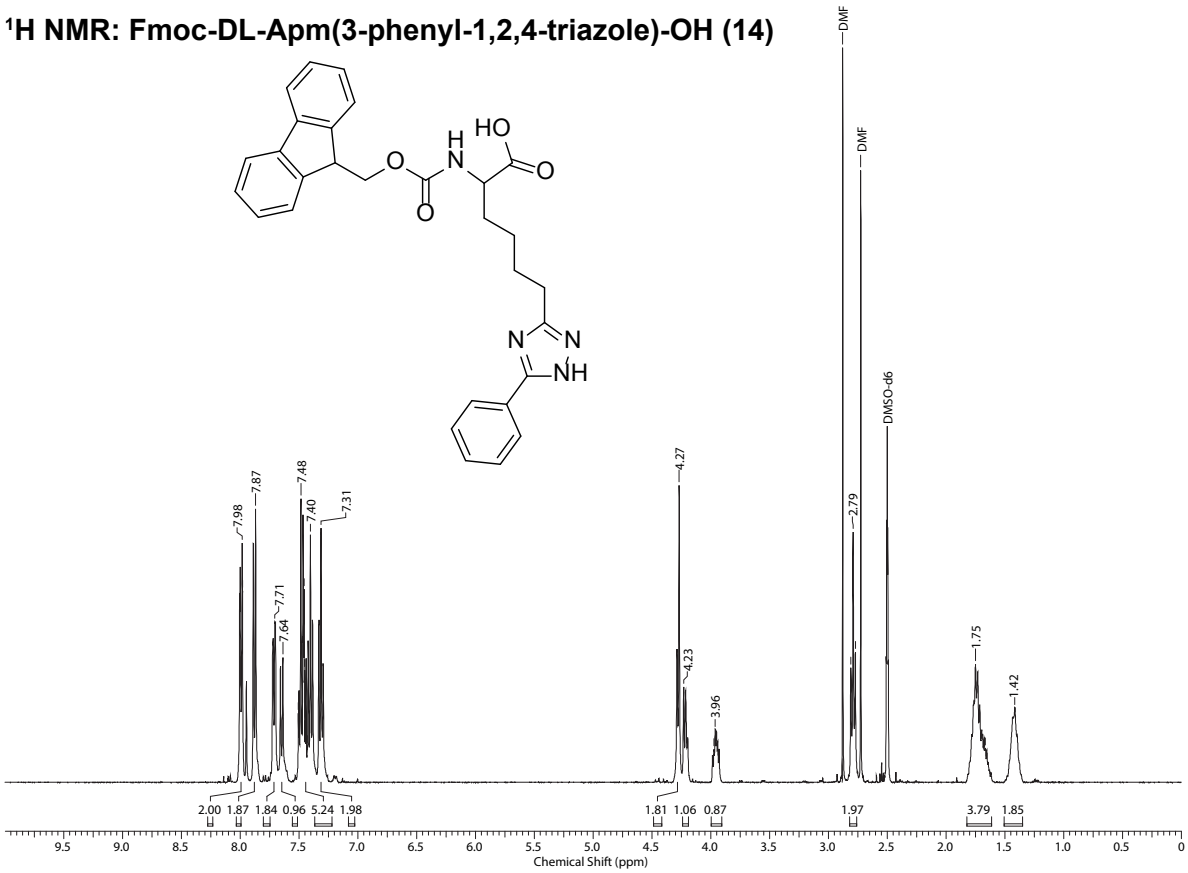


Figure C.20: ^{1}H NMR (300 MHz; DMSO-d_6) and ^{13}C NMR (75 MHz; DMSO-d_6) spectra of Fmoc-DL-Apm(3-ethyl-1,2,4-triazole)-OH (Fmoc-ApmPTri) (13).

^{1}H NMR: Fmoc-DL-Apm(3-phenyl-1,2,4-triazole)-OH (14)



^{13}C NMR: Fmoc-DL-Apm(3-phenyl-1,2,4-triazole)-OH (14)

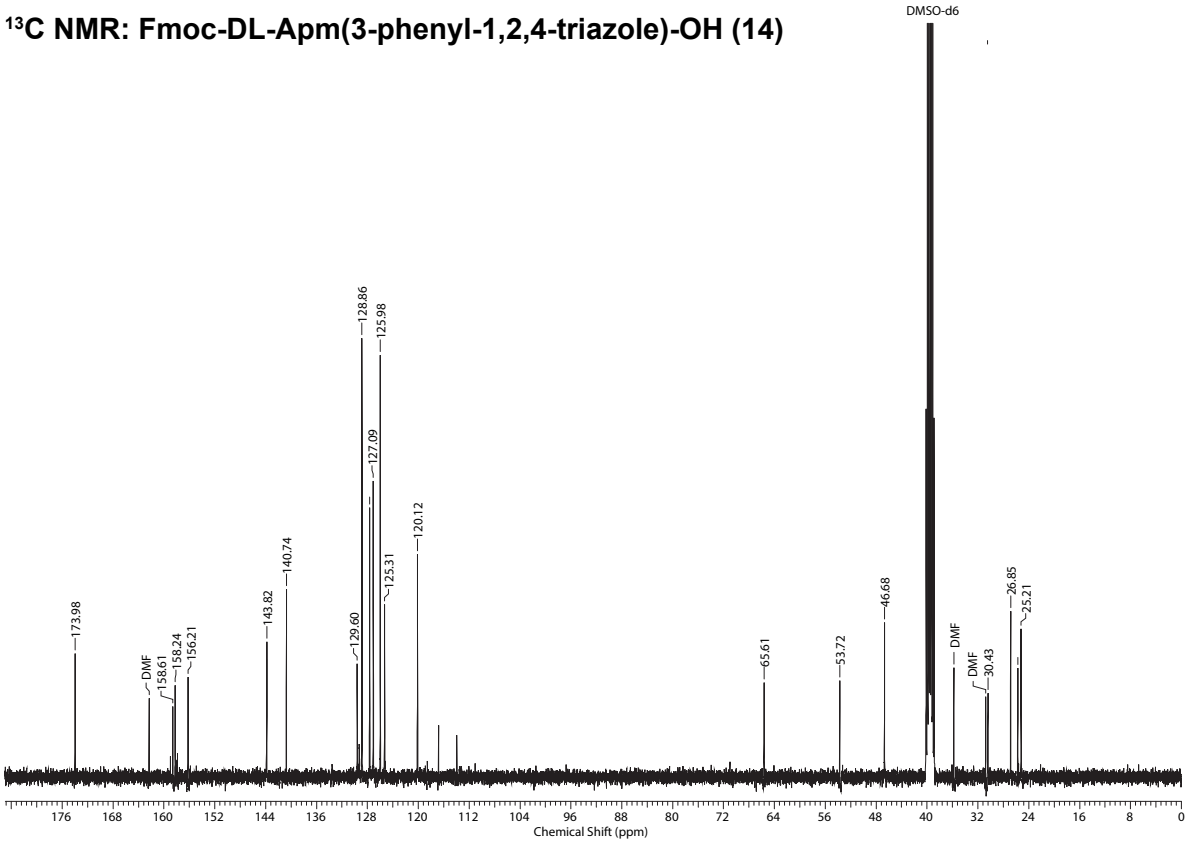
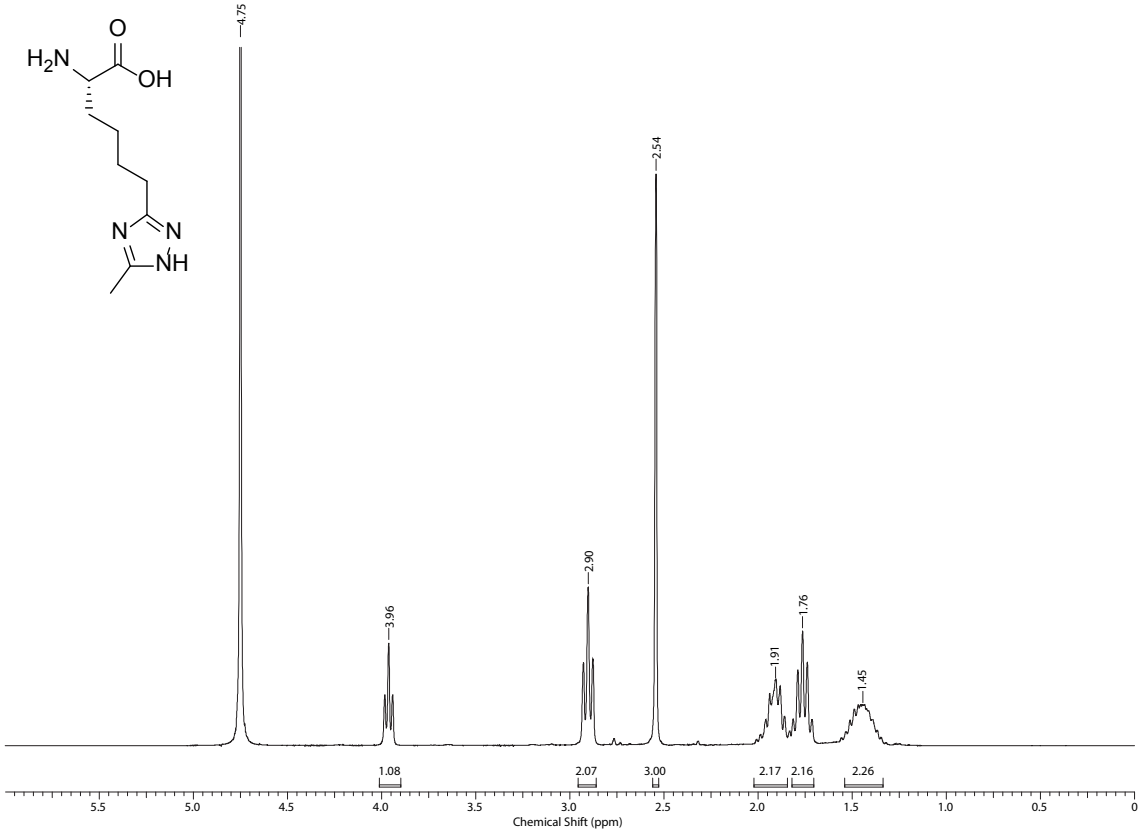


Figure C.21: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (101 MHz; DMSO-d_6) spectra of Fmoc-DL-Apm(3-phenyl-1,2,4-triazole)-OH (Fmoc-ApmBTri) (14).

^{1}H NMR: H-L-Apm(3-methyl-1,2,4-triazole)-OH (16)



^{13}C NMR: H-L-Apm(3-methyl-1,2,4-triazole)-OH (16)

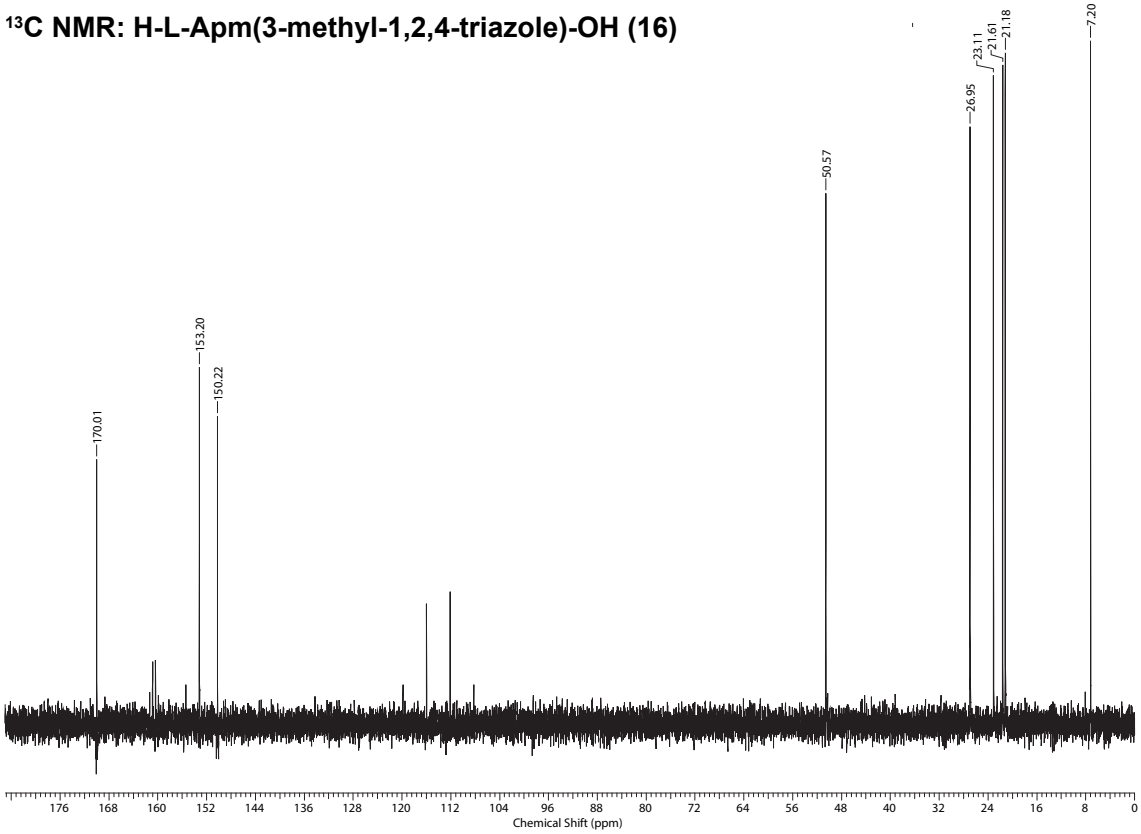
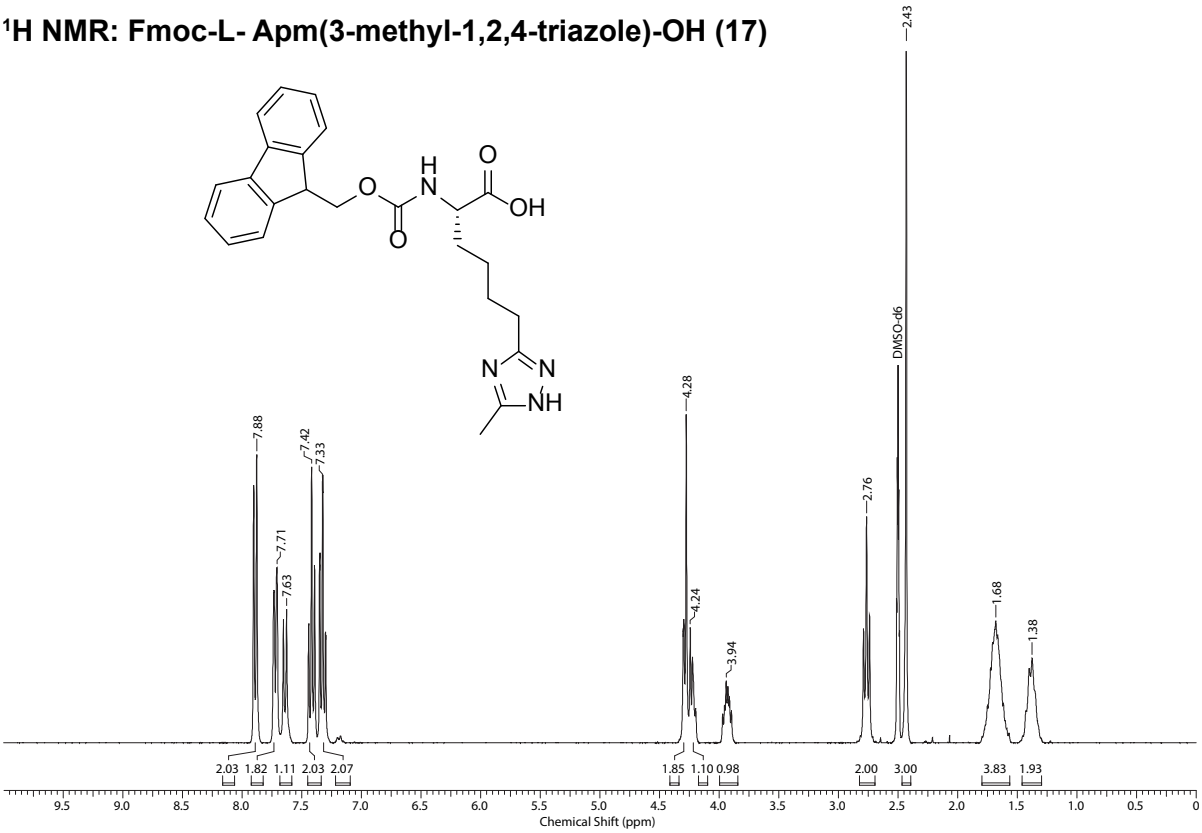


Figure C.22: ^{1}H NMR (300 MHz; D_2O) and ^{13}C NMR (75 MHz; D_2O) spectra of H-L-Apm(3-methyl-1,2,4-triazole)-OH (L-ApmTri) (16).

^{1}H NMR: Fmoc-L- Apm(3-methyl-1,2,4-triazole)-OH (17)



^{13}C NMR: Fmoc-L- Apm(3-methyl-1,2,4-triazole)-OH (17)

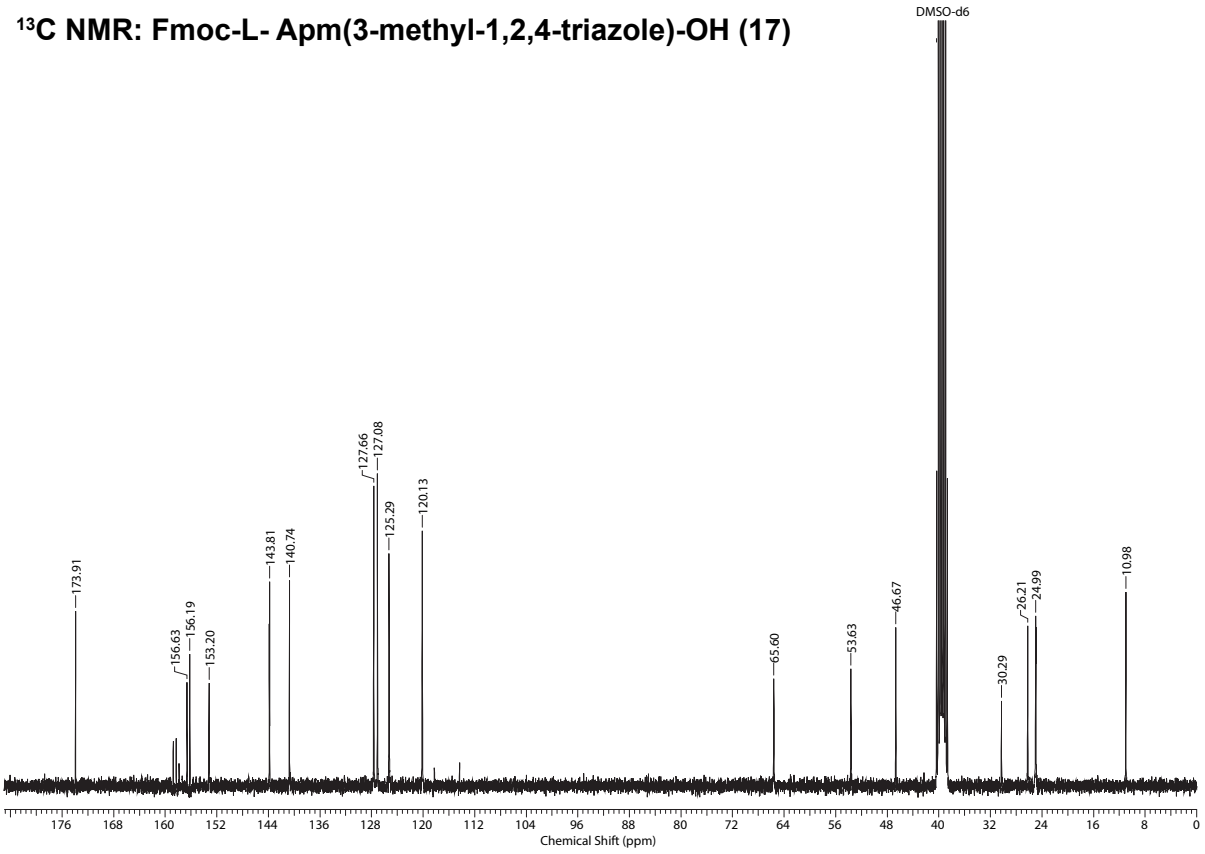
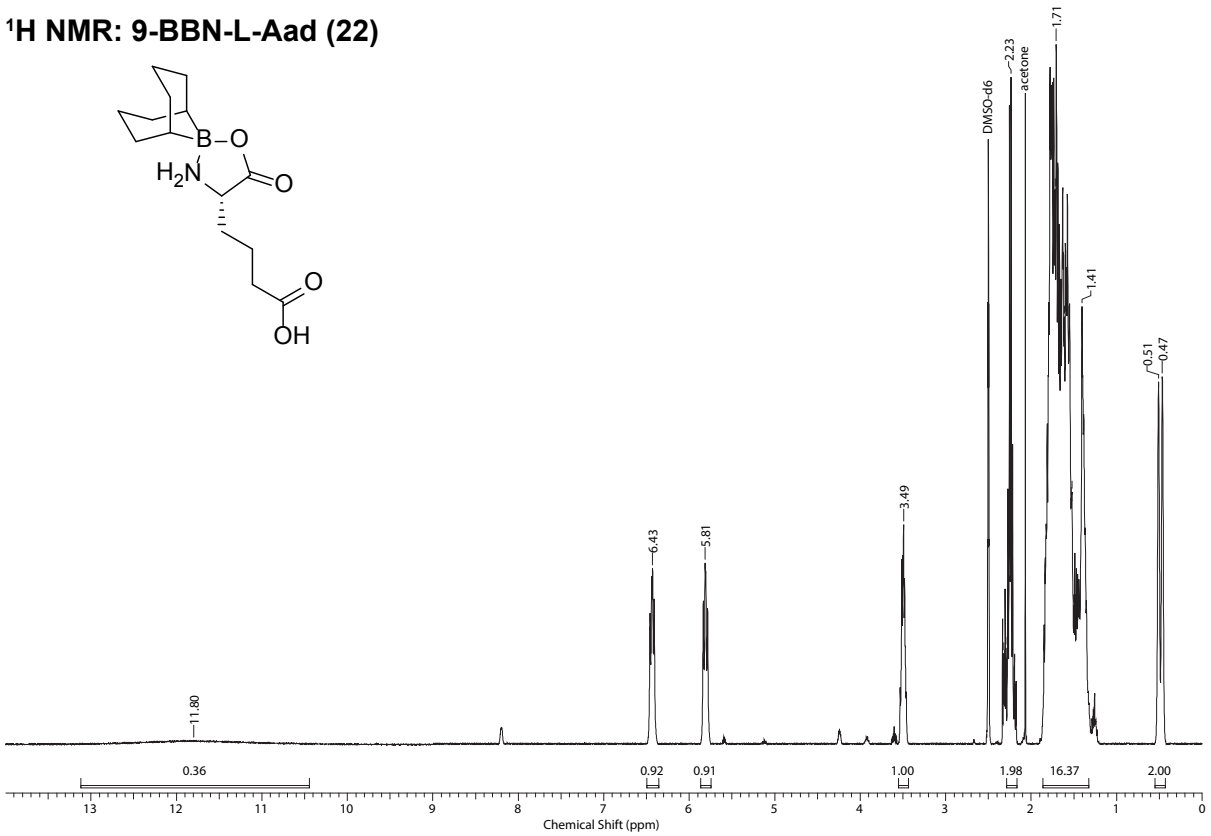


Figure C.23: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (101 MHz; DMSO-d_6) spectra of Fmoc-L-Apm(3-methyl-1,2,4-triazole)-OH (Fmoc-L-ApmTri) (17).

^{1}H NMR: 9-BBN-L-Aad (22)



^{13}C NMR: 9-BBN-L-Aad (22)

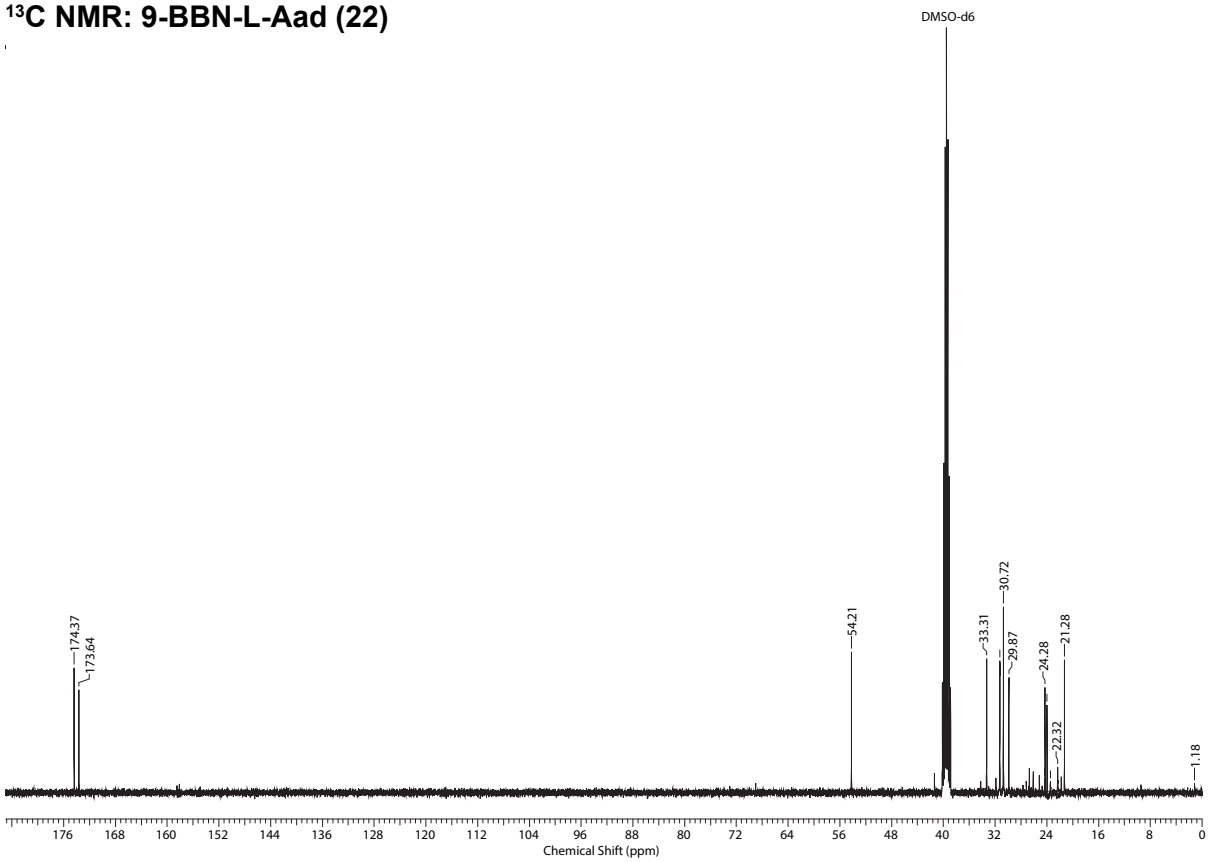
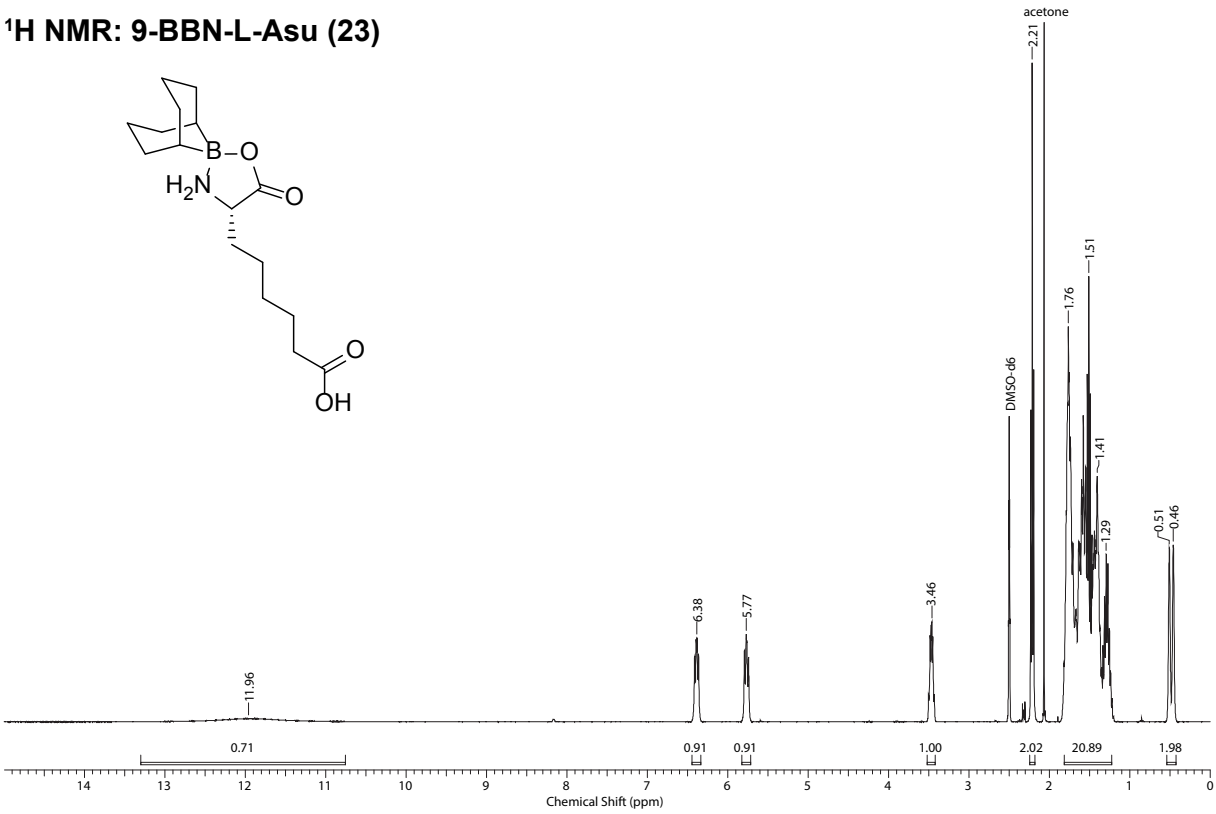


Figure C.24: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (101 MHz; DMSO-d_6) spectra of 9-BBN-L-Aad (22).

^{1}H NMR: 9-BBN-L-Asu (23)



^{13}C NMR: 9-BBN-L-Asu (23)

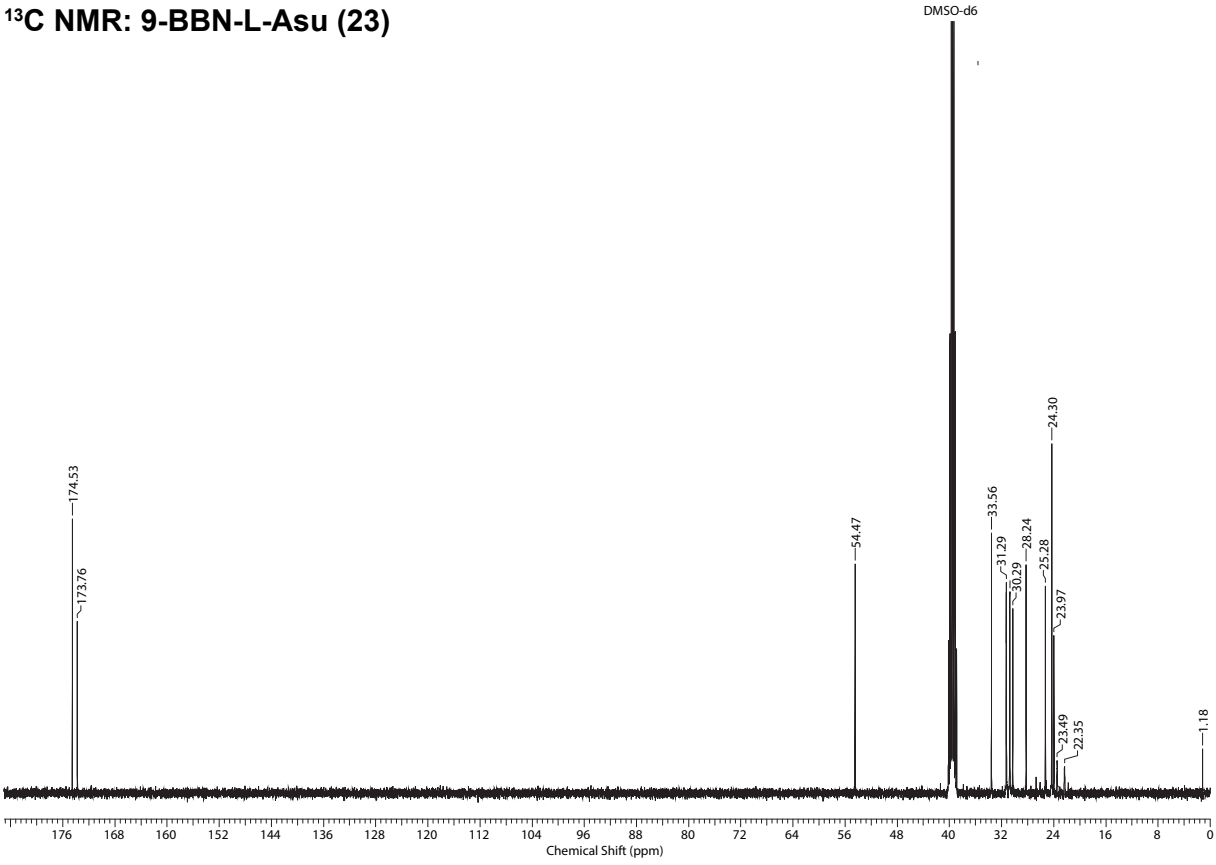
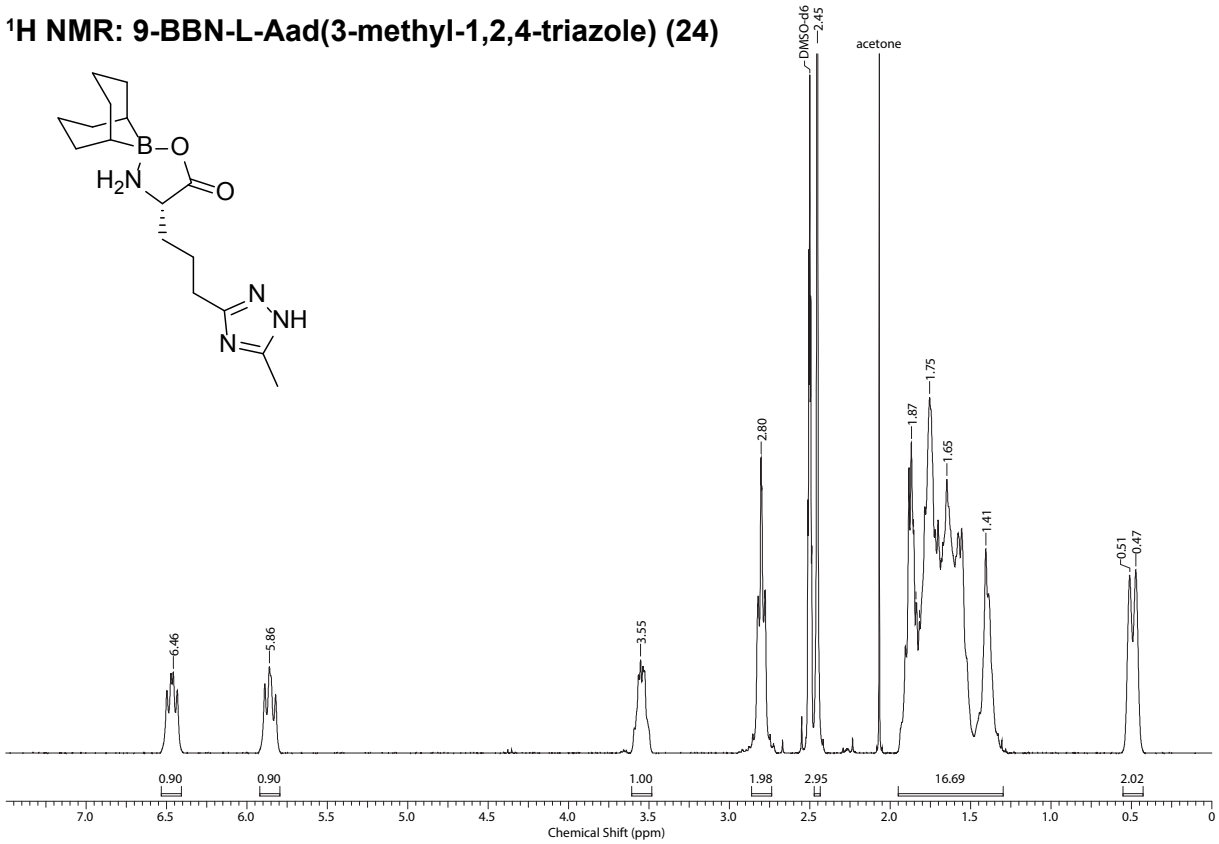


Figure C.25: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (101 MHz; DMSO-d_6) spectra of 9-BBN-L-Asu (23).

^{1}H NMR: 9-BBN-L-Aad(3-methyl-1,2,4-triazole) (24)



^{13}C NMR: 9-BBN-L-Aad(3-methyl-1,2,4-triazole) (24)

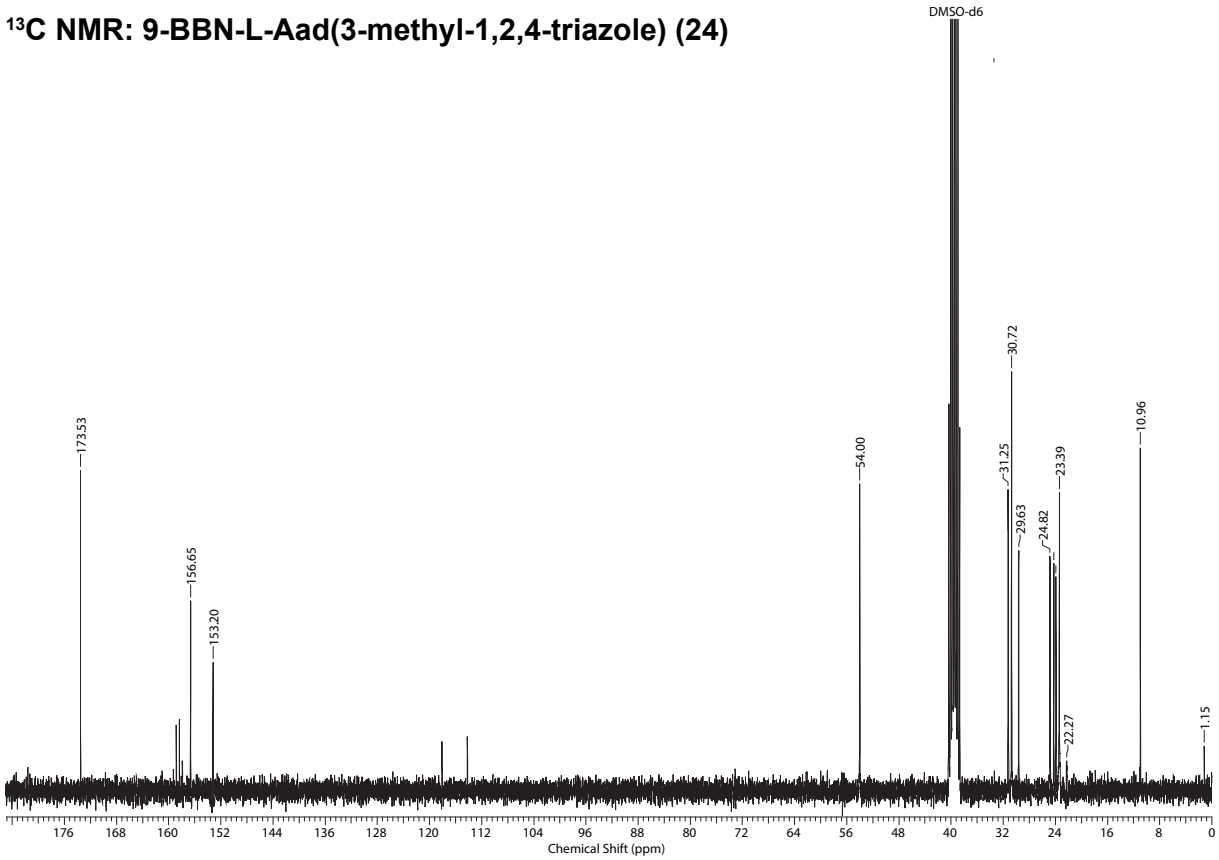
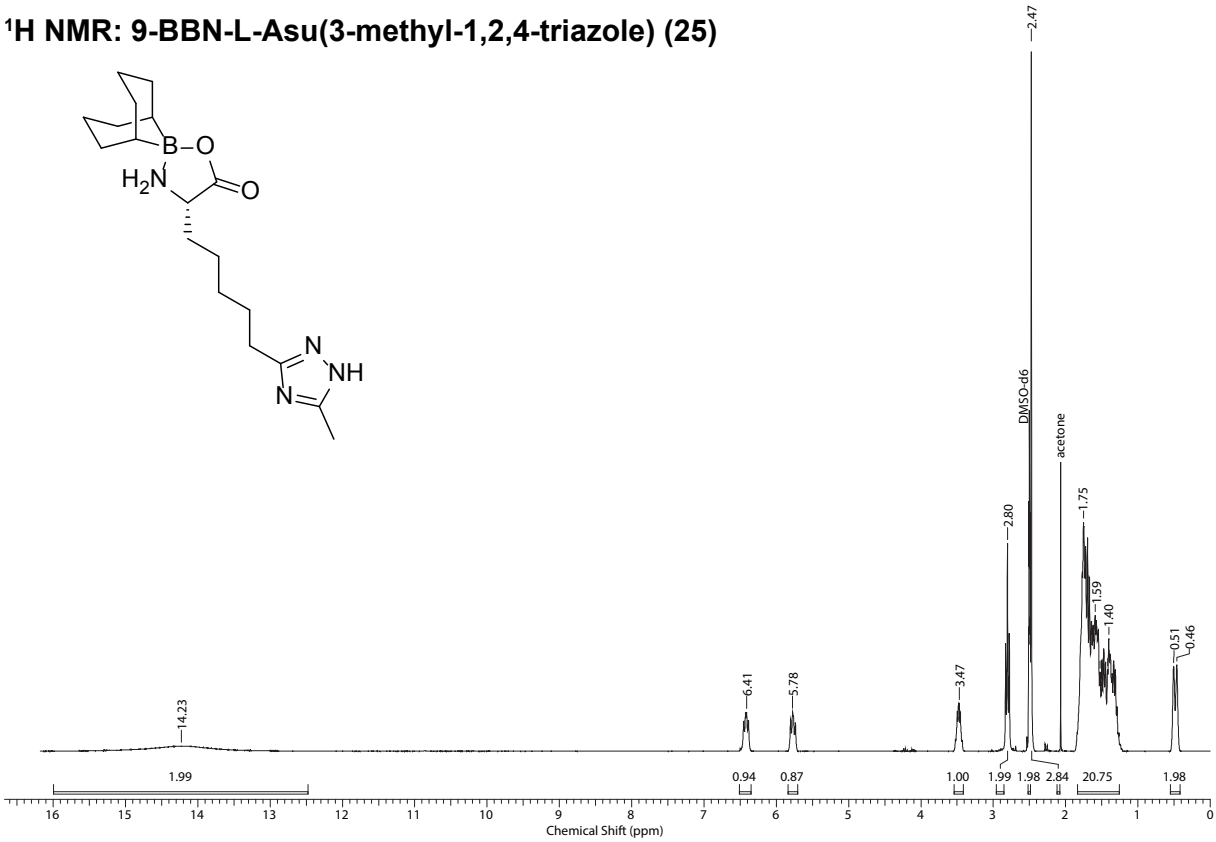


Figure C.26: ^{1}H NMR (300 MHz; DMSO-d_6) and ^{13}C NMR (75 MHz; DMSO-d_6) spectra of 9-BBN-L-Aad(3-methyl-1,2,4-triazole) (9-BBN-L-AadTri) (24).

^{1}H NMR: 9-BBN-L-Asu(3-methyl-1,2,4-triazole) (25)



^{13}C NMR: 9-BBN-L-Asu(3-methyl-1,2,4-triazole) (25)

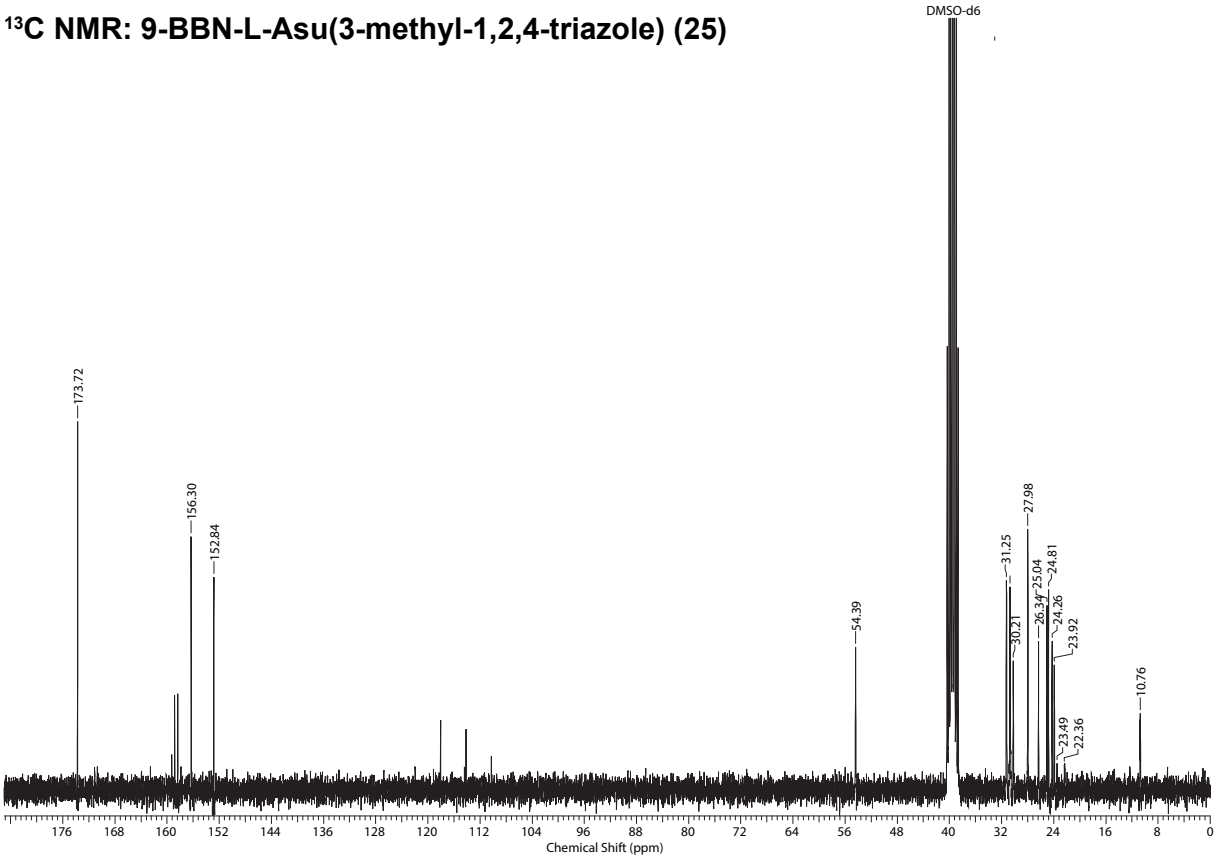
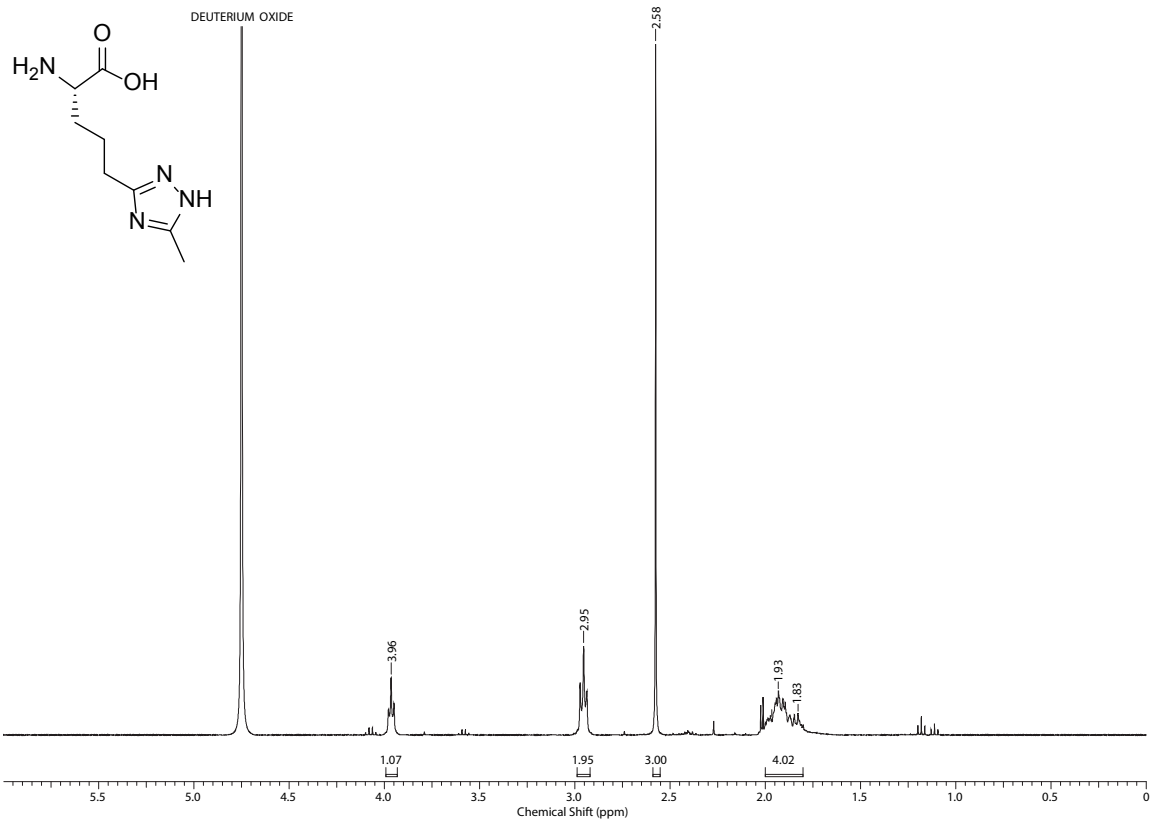


Figure C.27: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (101 MHz; DMSO-d_6) spectra of 9-BBN-L-Asu(3-methyl-1,2,4-triazole) (9-BBN-L-AsuTri) (25).

^{1}H NMR: H-L-Aad(3-methyl-1,2,4-triazole)-OH (26)



^{13}C NMR: H-L-Aad(3-methyl-1,2,4-triazole)-OH (26)

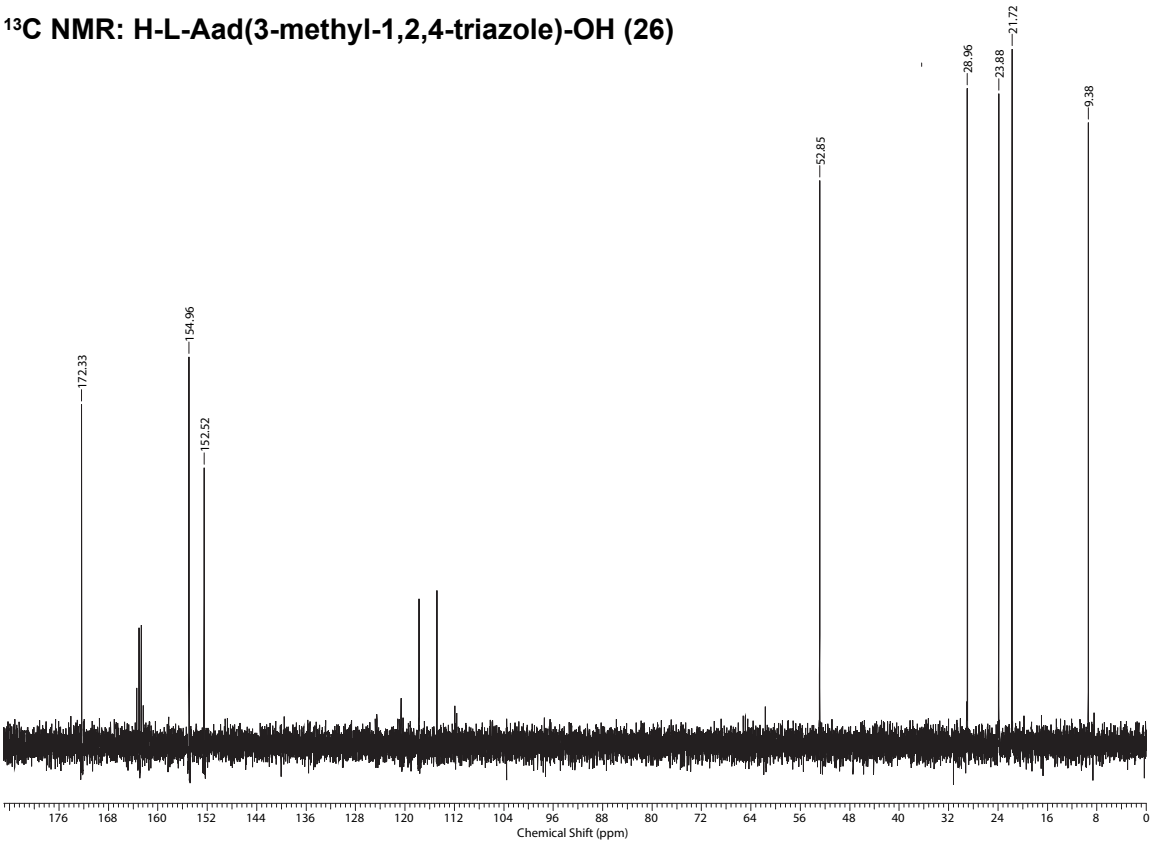
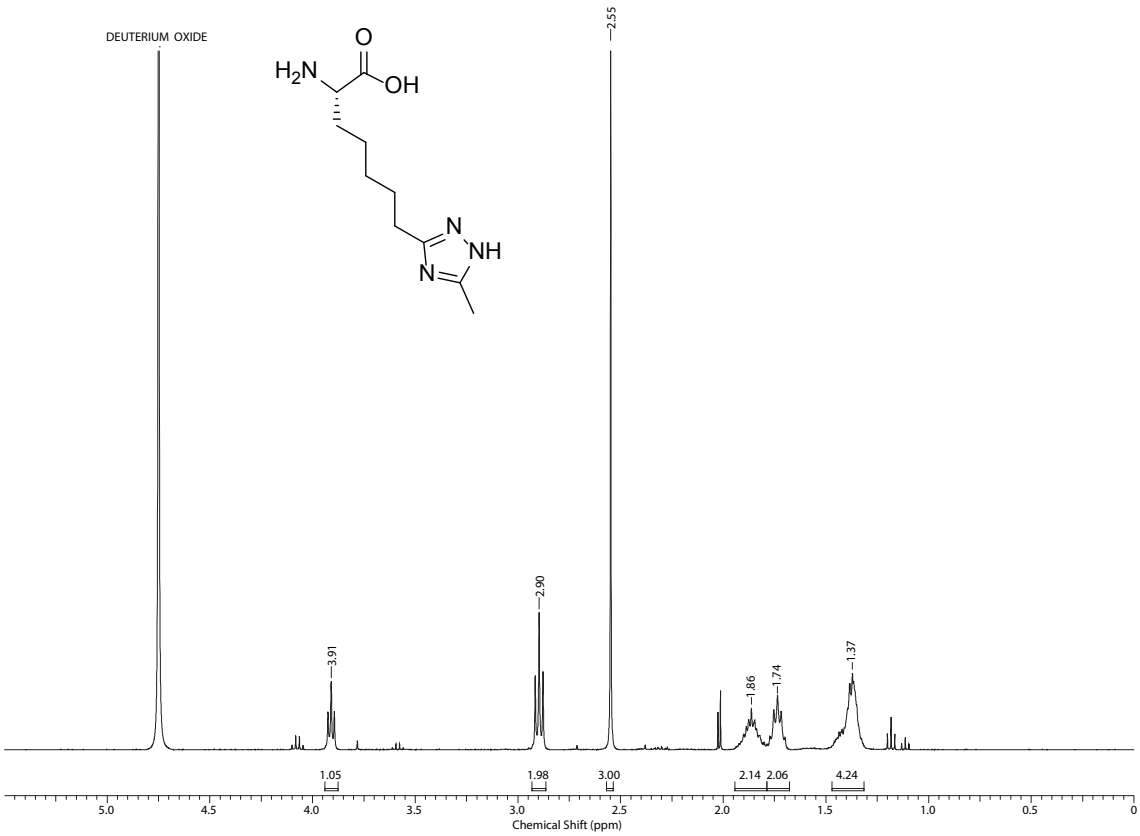


Figure C.28: ^{1}H NMR (400 MHz; D_2O) and ^{13}C NMR (101 MHz; D_2O) spectra of H-L-Aad(3-methyl-1,2,4-triazole)-OH (L-AadTri) (26).

^{1}H NMR: H-L-Asu(3-methyl-1,2,4-triazole)-OH (27)



^{13}C NMR: H-L-Asu(3-methyl-1,2,4-triazole)-OH (27)

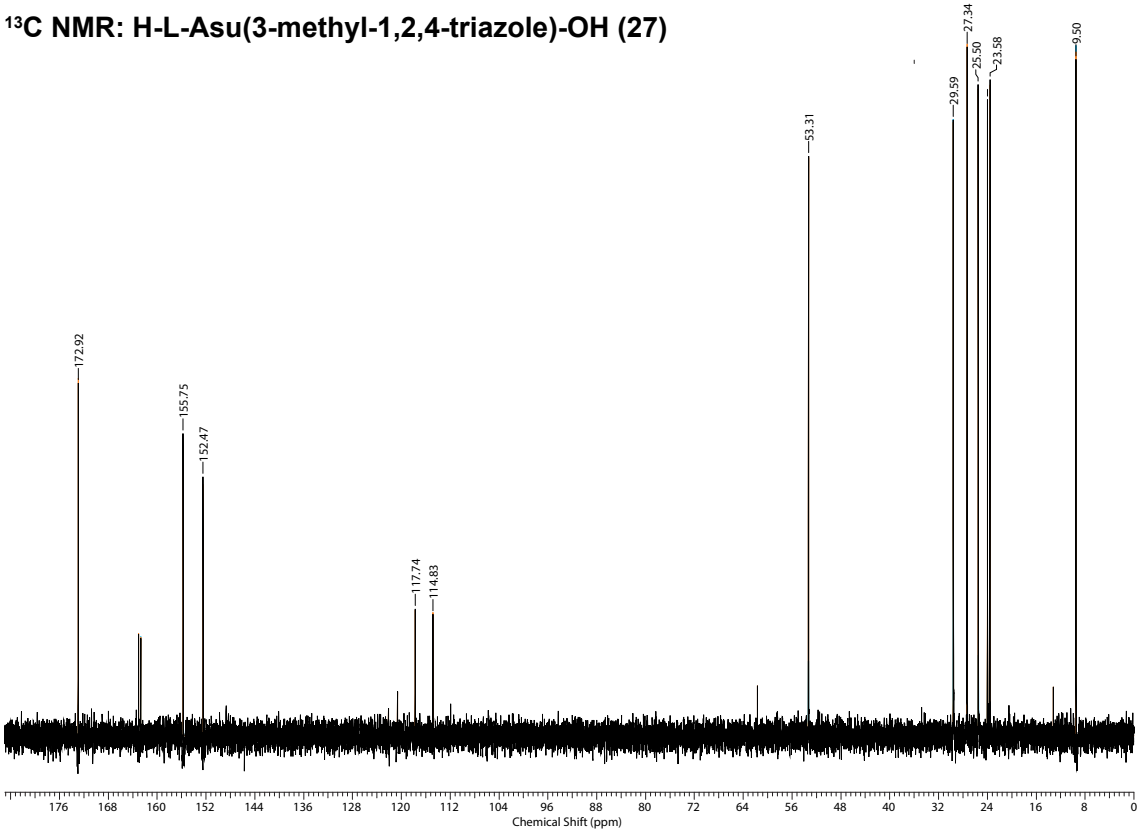
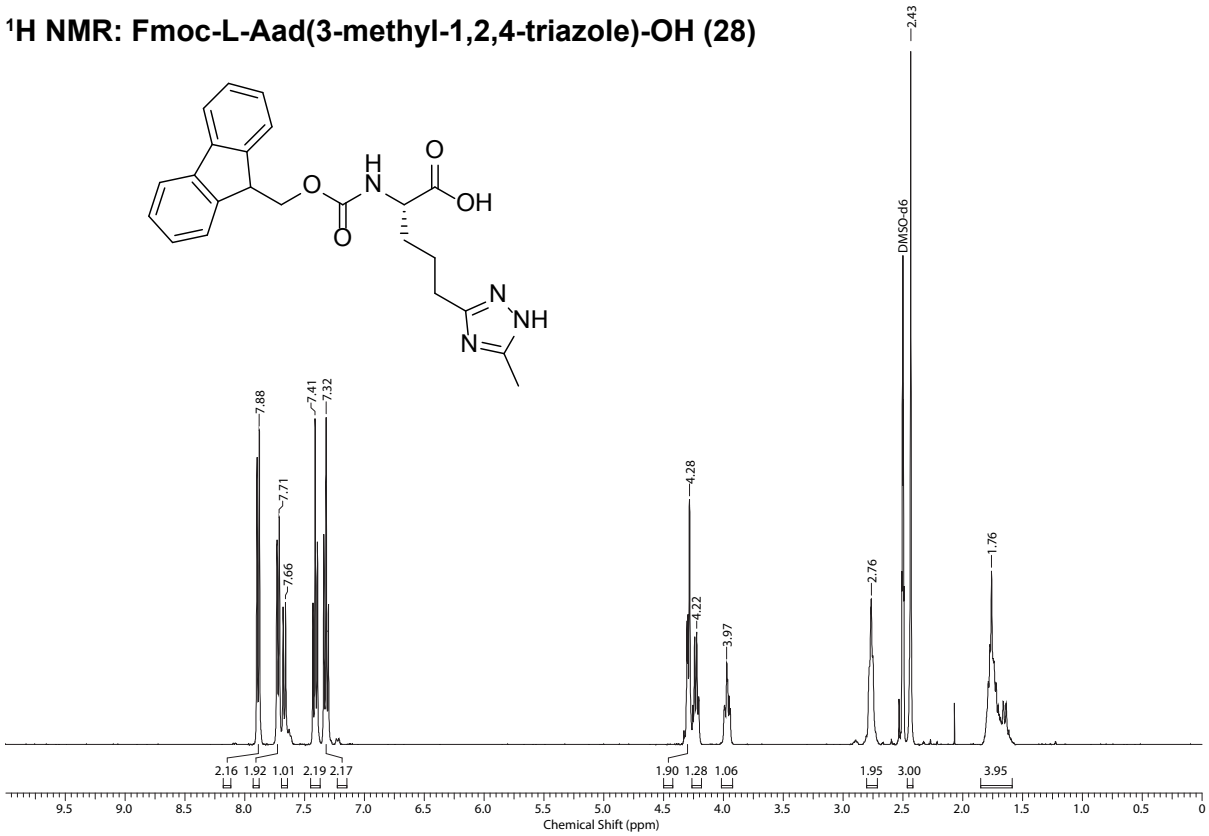


Figure C.29: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (101 MHz; DMSO-d_6) spectra of H-L-Asu(3-methyl-1,2,4-triazole)-OH (L-AsuTri) (27).

^{1}H NMR: Fmoc-L-Aad(3-methyl-1,2,4-triazole)-OH (28)



^{13}C NMR: Fmoc-L-Aad(3-methyl-1,2,4-triazole)-OH (28)

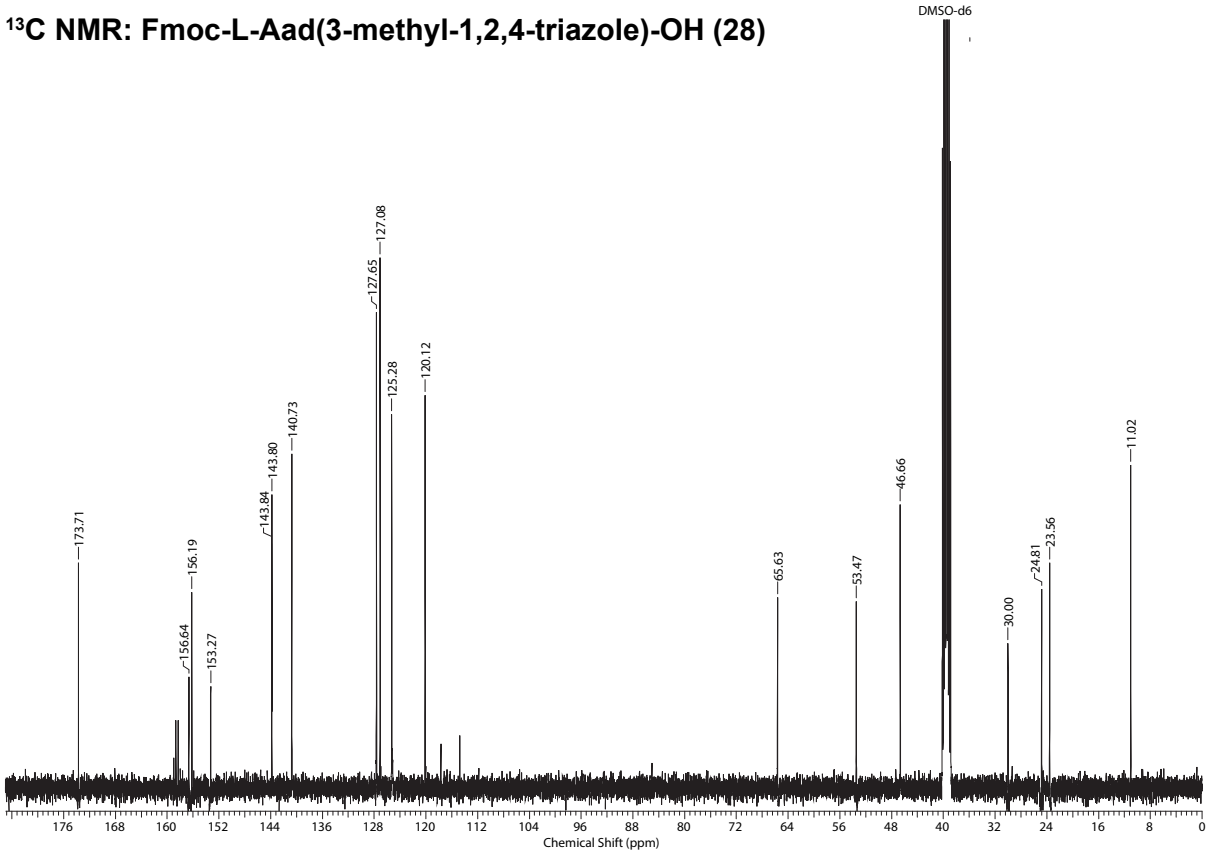
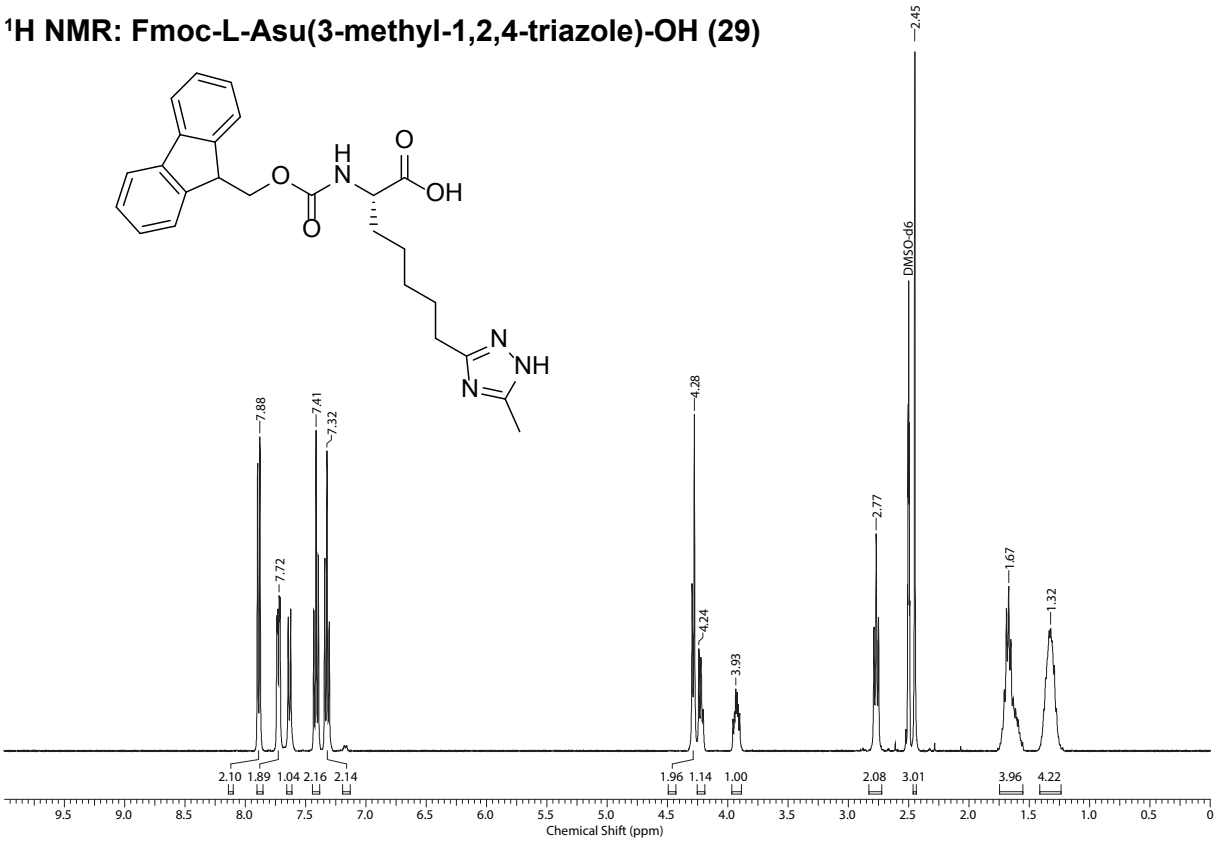


Figure C.30: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (101 MHz; DMSO-d_6) spectra of Fmoc-L-Aad(3-methyl-1,2,4-triazole)-OH (Fmoc-L-AadTri) (28).

^{1}H NMR: Fmoc-L-Asu(3-methyl-1,2,4-triazole)-OH (29)



^{13}C NMR: Fmoc-L-Asu(3-methyl-1,2,4-triazole)-OH (29)

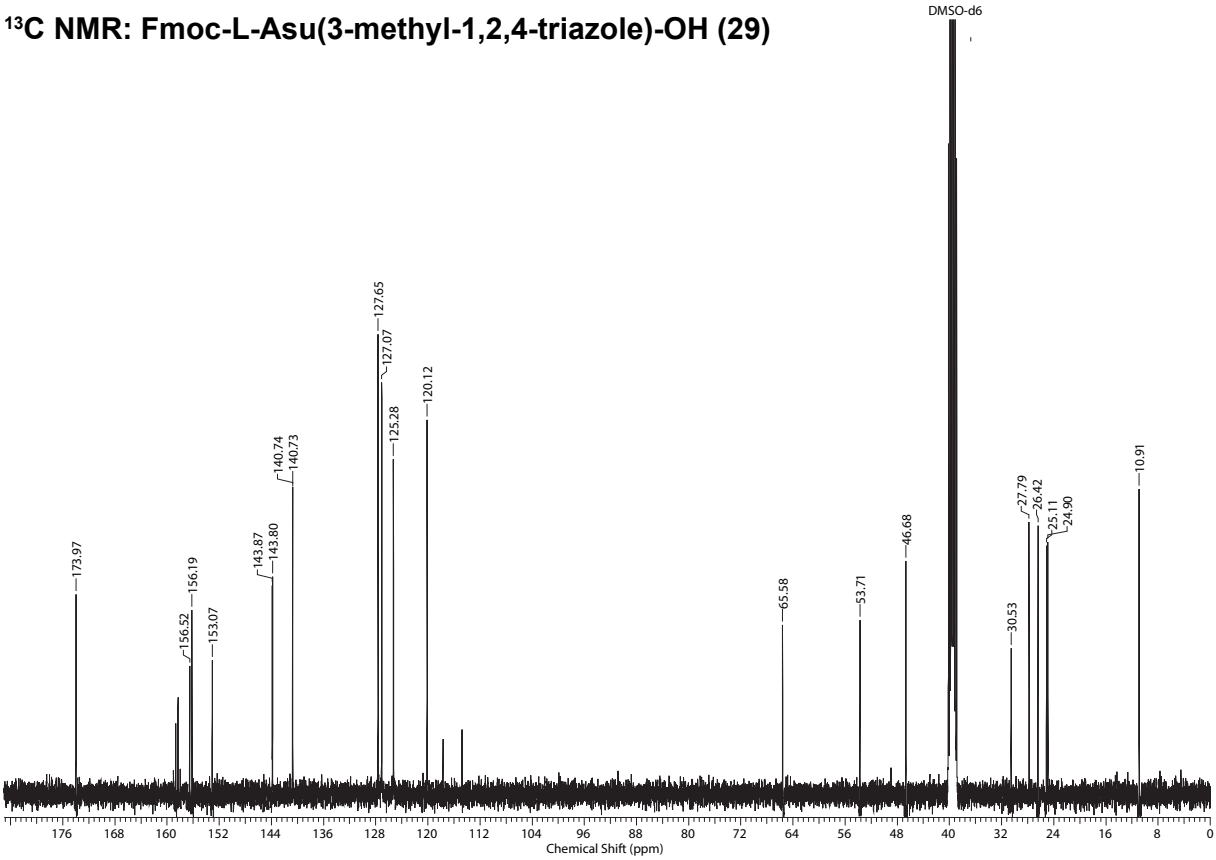
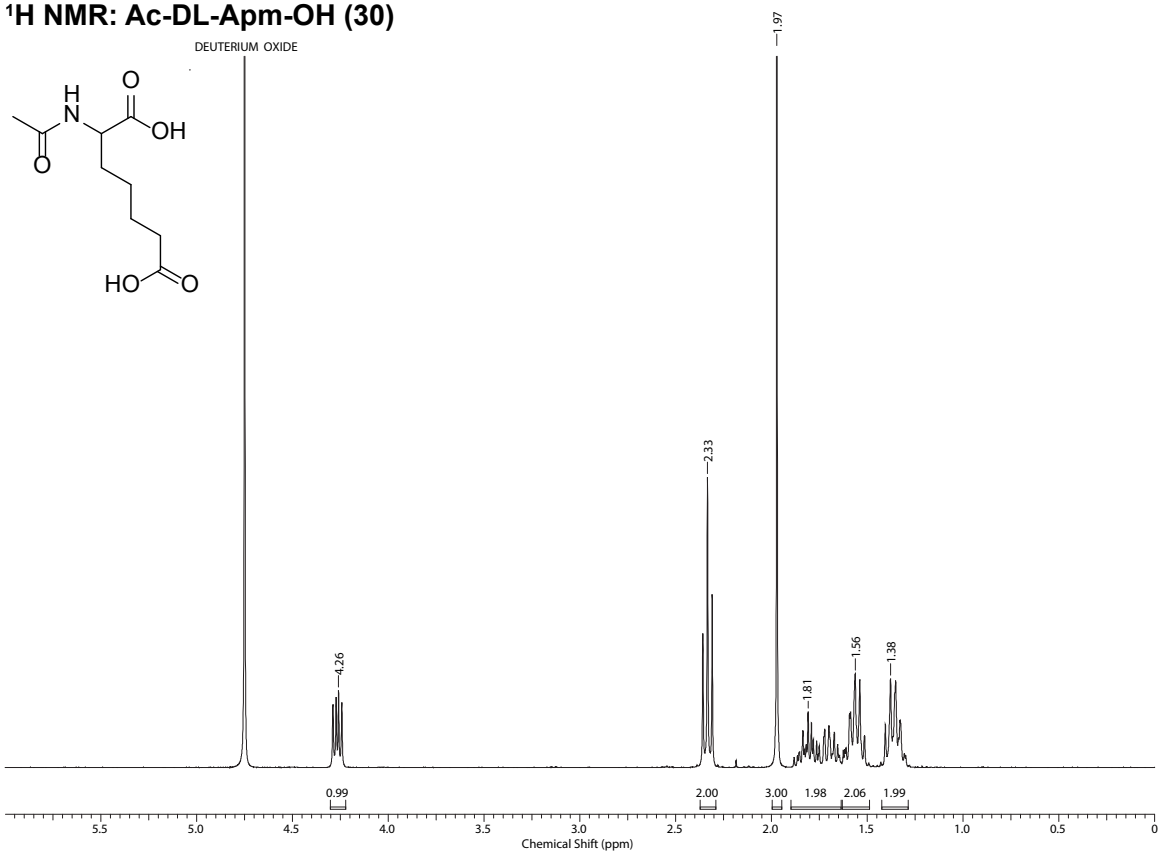


Figure C.31: ^{1}H NMR (400 MHz; DMSO-d_6) and ^{13}C NMR (101 MHz; DMSO-d_6) spectra of Fmoc-L-Asu(3-methyl-1,2,4-triazole)-OH (Fmoc-L-AsuTri) (29).

^{1}H NMR: Ac-DL-Apm-OH (30)



^{13}C NMR: Ac-DL-Apm-OH (30)

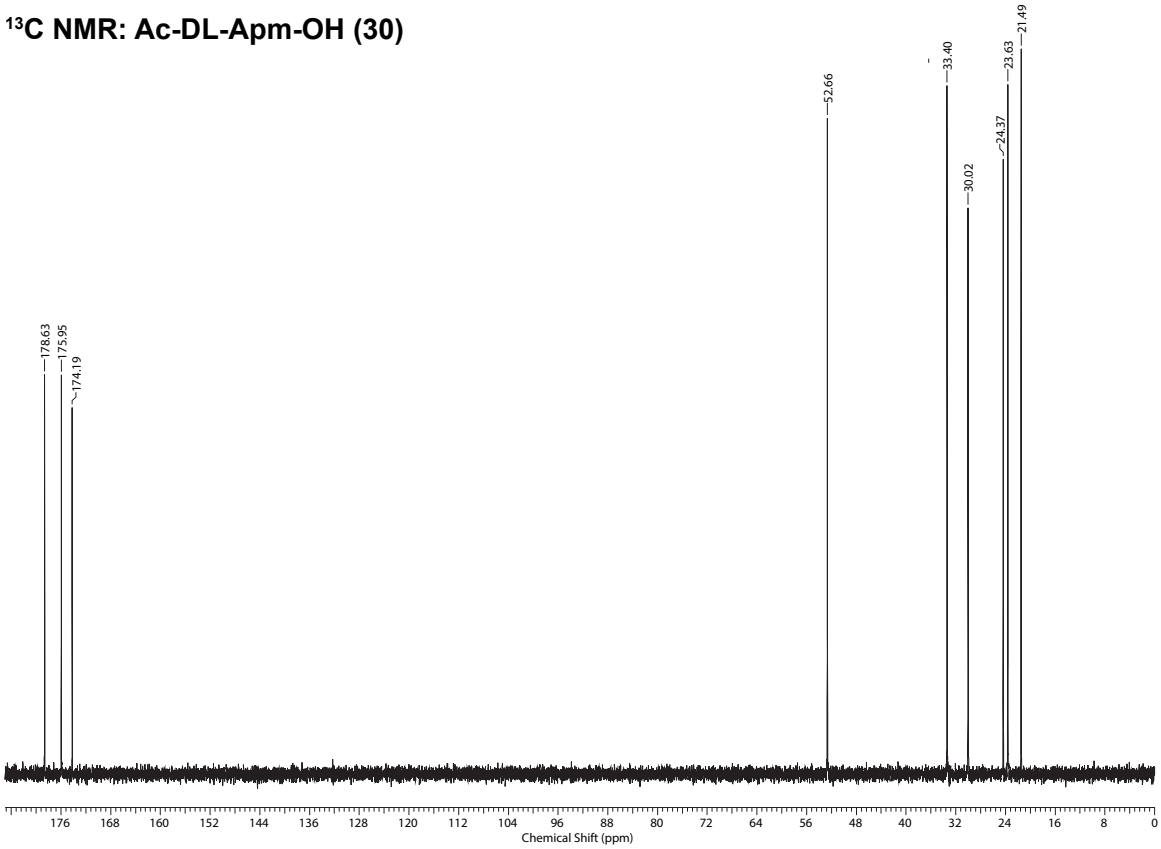
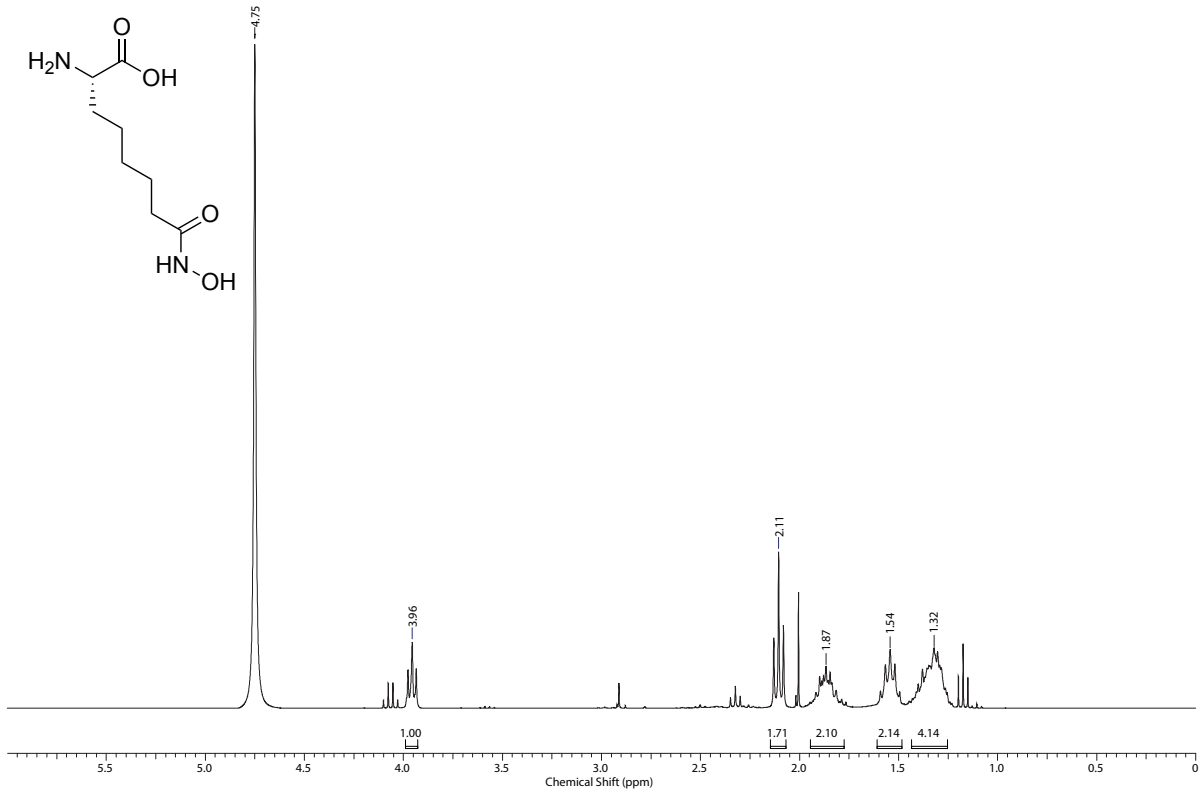


Figure C.32: ^{1}H NMR (300 MHz; D_2O) and ^{13}C NMR (75 MHz; D_2O) spectra of Ac-DL-Apm-OH (30).

^{1}H NMR: H-L-Asu(NH-OH)-OH (37)



^{13}C NMR: H-L-Asu(NH-OH)-OH (37)

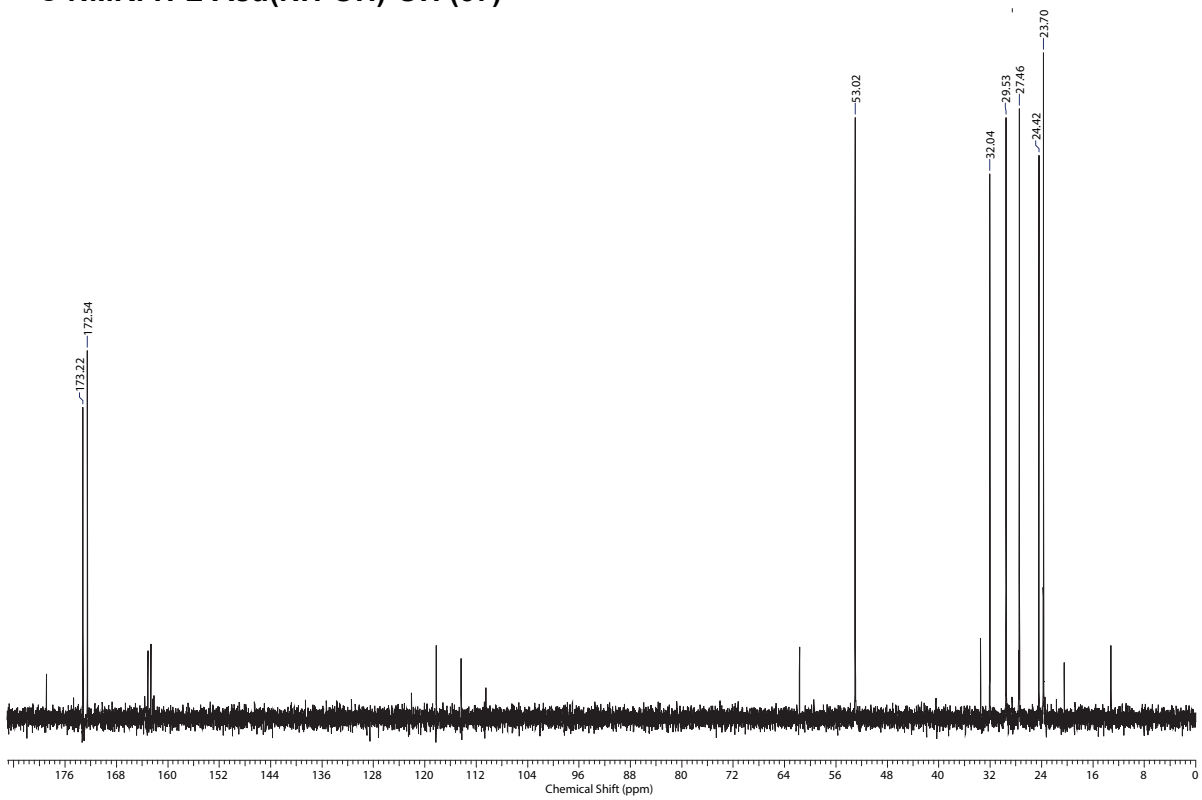
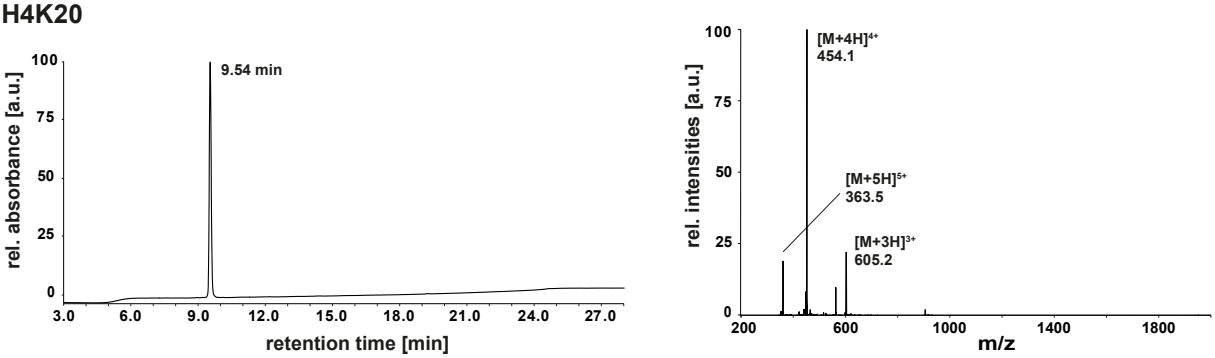


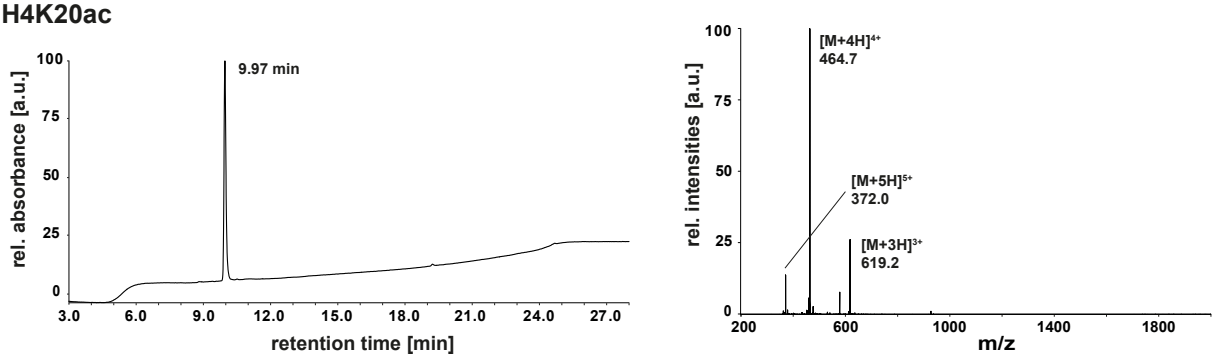
Figure C.33: ^{1}H NMR (300 MHz; D_2O) and ^{13}C NMR (75 MHz; D_2O) spectra of H-Asu(NH-OH)-OH (AsuHd) (37).

C.3 Analytical results of peptide synthesis

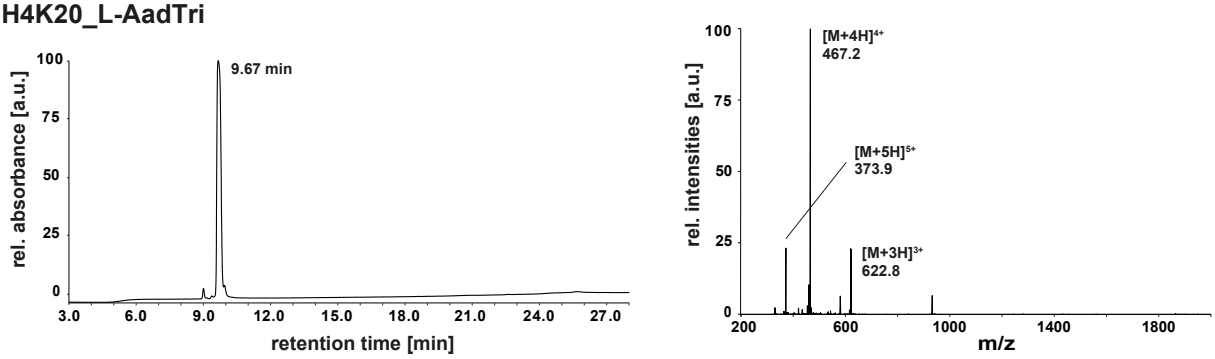
H4K20



H4K20ac



H4K20_L-AadTri



H4K20_L-ApmTri

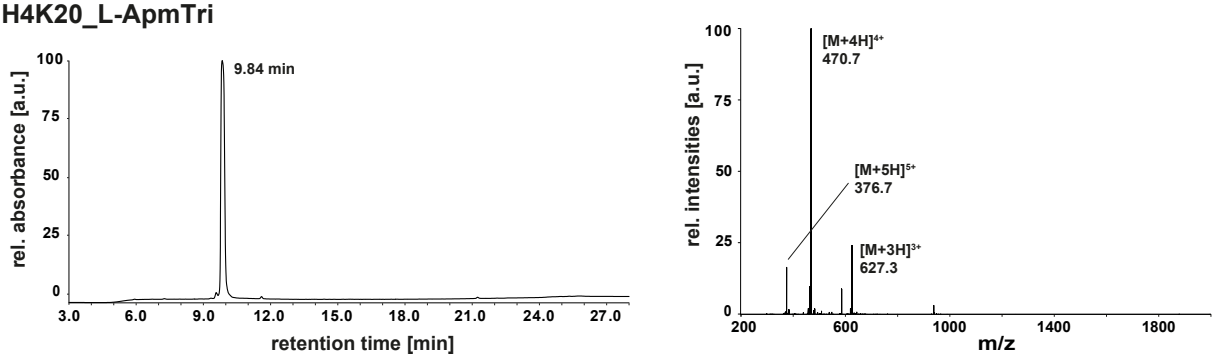
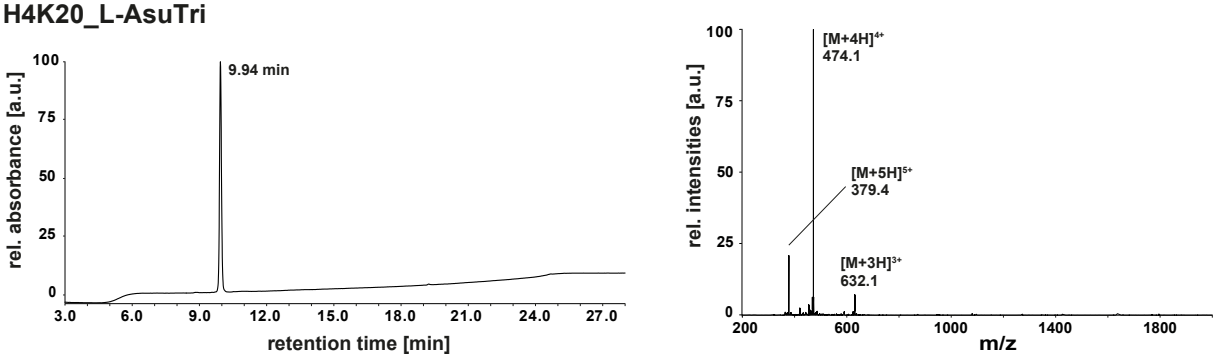
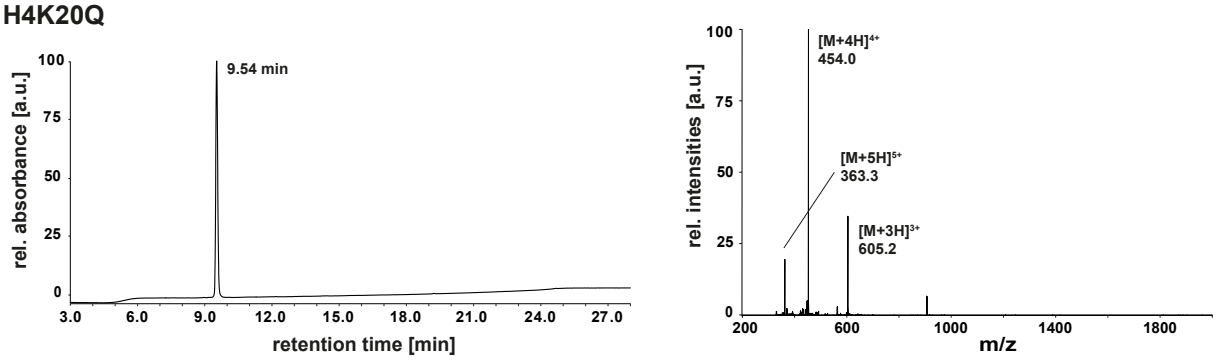


Figure C.34: LC-MS analysis of synthesized peptides H4K20, H4K20ac, H4K20_L-AadTri and H4K20_L-ApmTri.

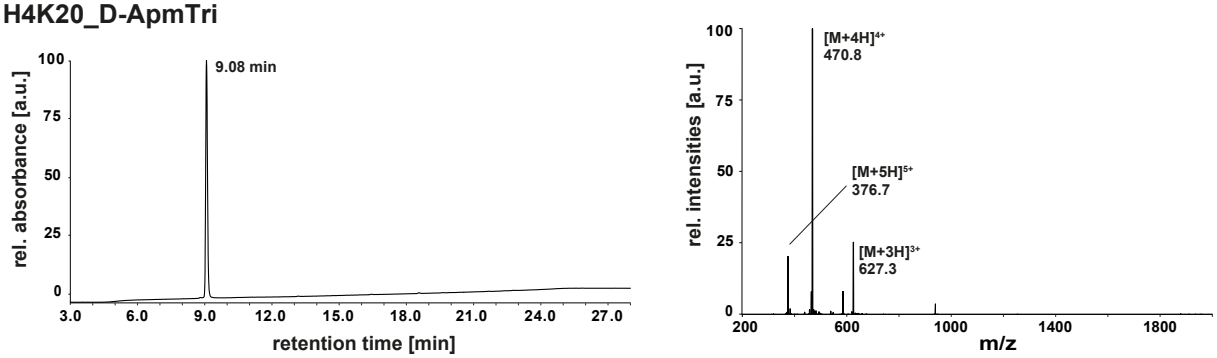
H4K20_L-AsuTri



H4K20Q



H4K20_D-ApmTri



H4K20_DL-ApmTri

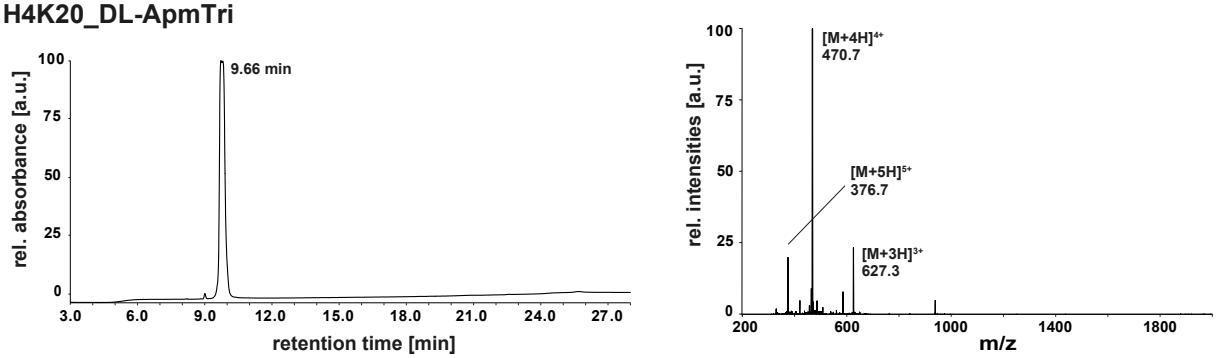
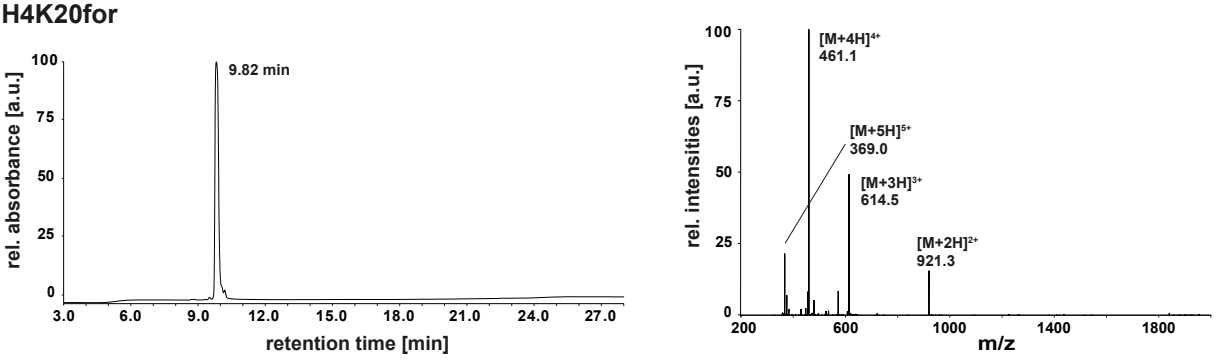
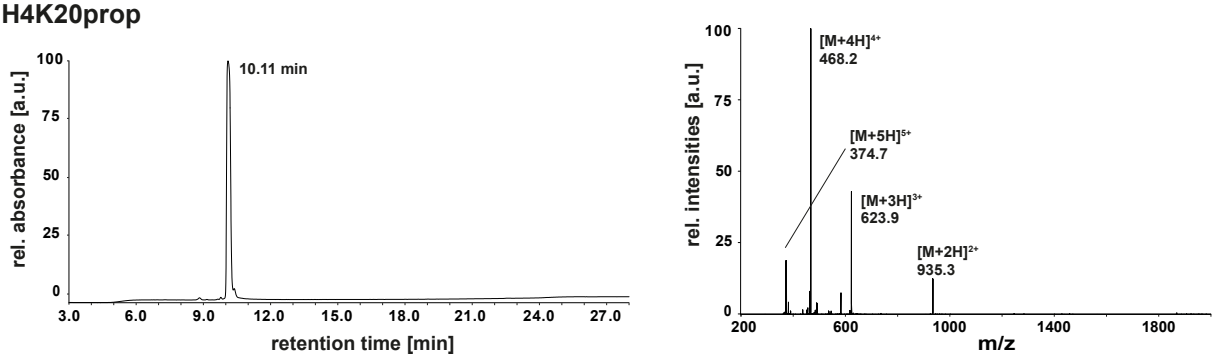


Figure C.35: LC-MS analysis of synthesized peptides H4K20_L-AsuTri, H4K20Q, H4K20_D-ApmTri and H4K20_DL-ApmTri.

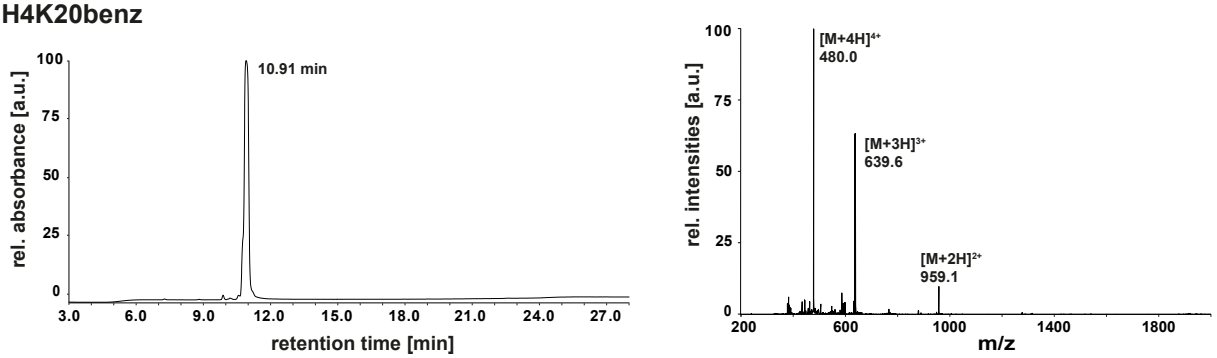
H4K20for



H4K20prop



H4K20benz



H4K20_DL-ApmFTri

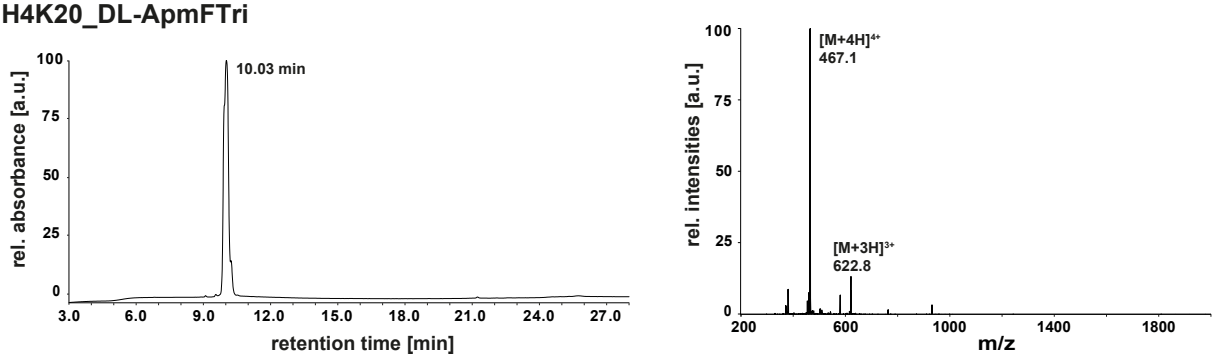
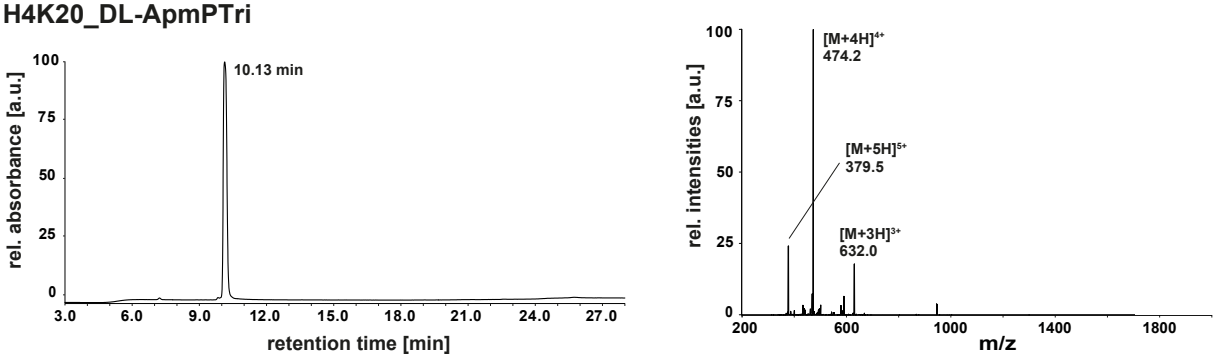
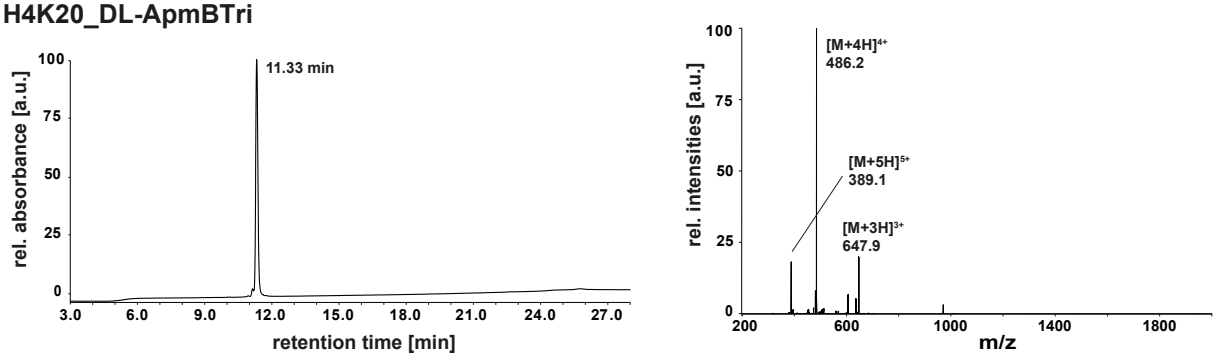


Figure C.36: LC-MS analysis of synthesized peptides H4K20for, H4K20prop, H4K20benz and H4K20_DL-ApmFTri.

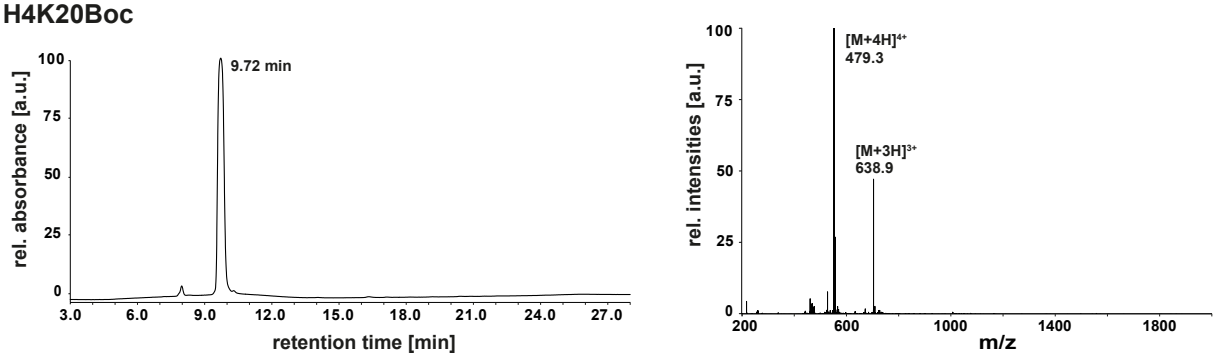
H4K20_DL-ApmPTri



H4K20_DL-ApmBTri



H4K20Boc



H4K20_L-ApmTri short

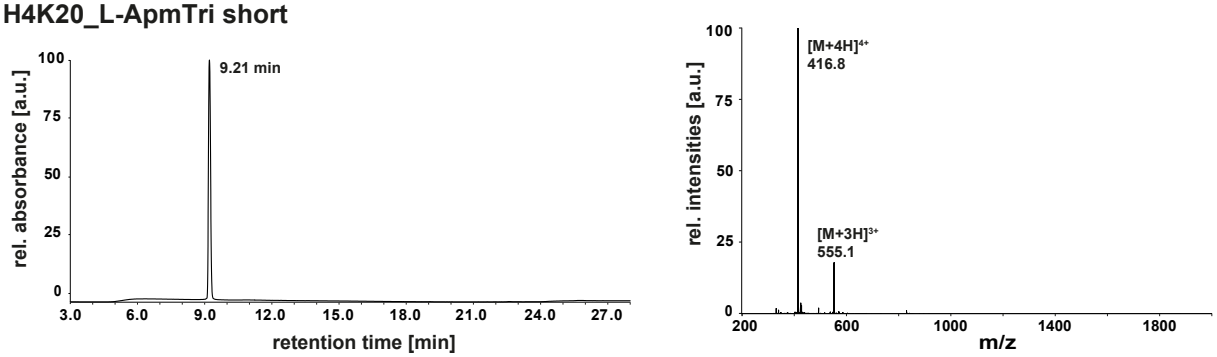
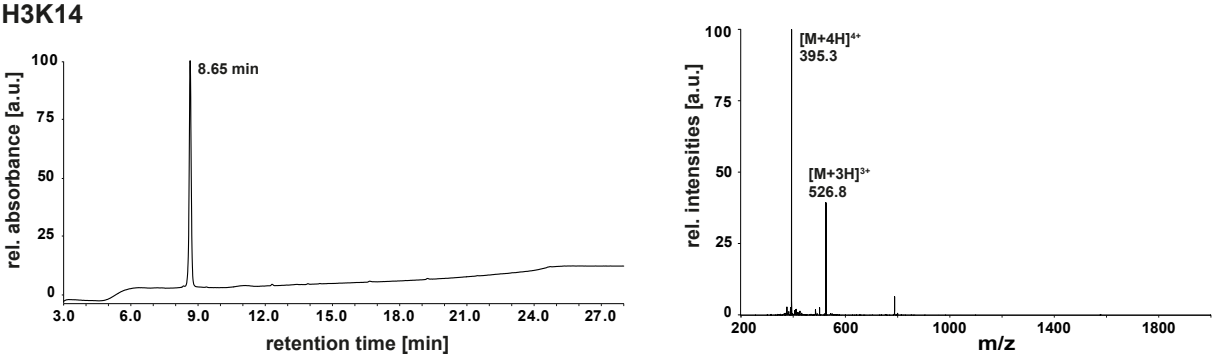
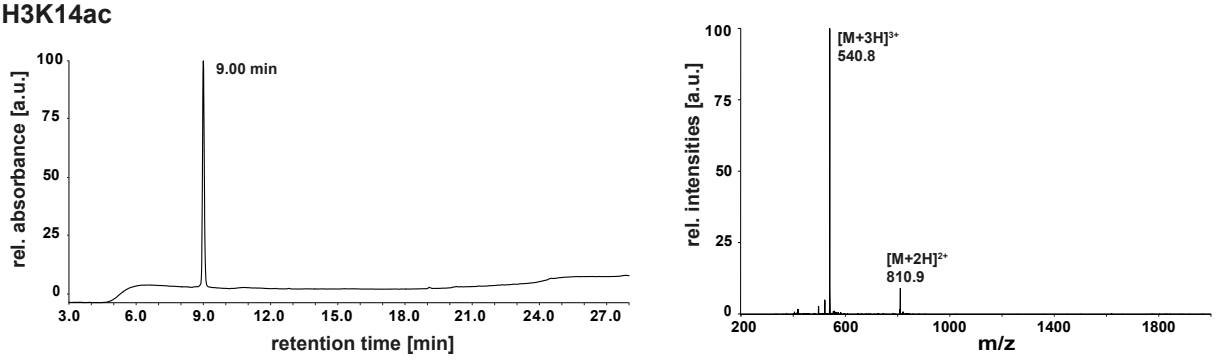


Figure C.37: LC-MS analysis of synthesized peptides H4K20_DL-ApmPTri, H4K20_DL-ApmBTri, H4K20Boc and H4K20_L-ApmTri short.

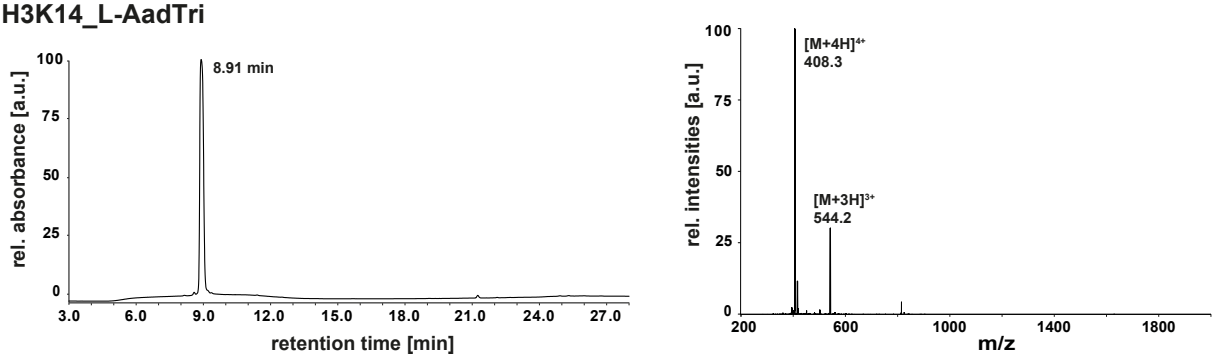
H3K14



H3K14ac



H3K14_L-AadTri



H3K14_DL-ApmTri

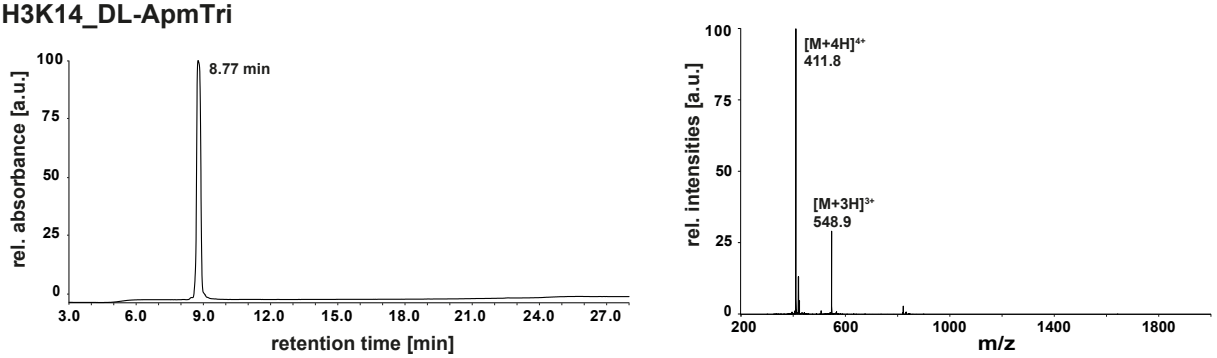


Figure C.38: LC-MS analysis of synthesized peptides H3K14, H3K14ac, H3K14_L-AadTri and H3K14_DL-ApmTri.

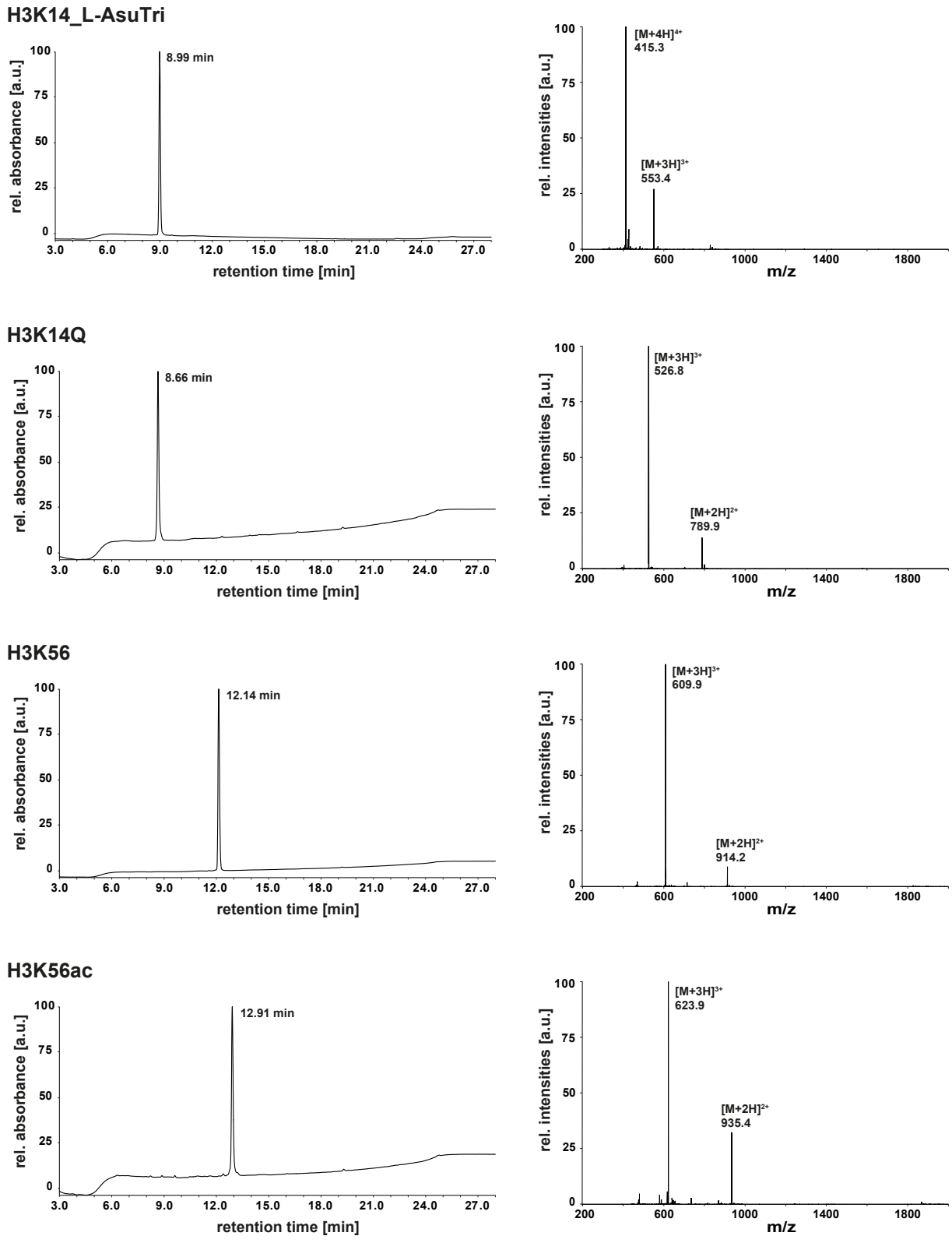
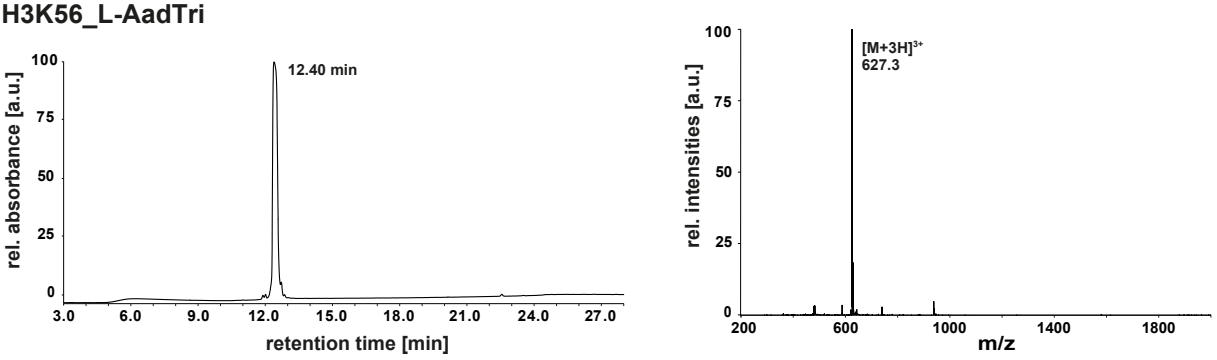
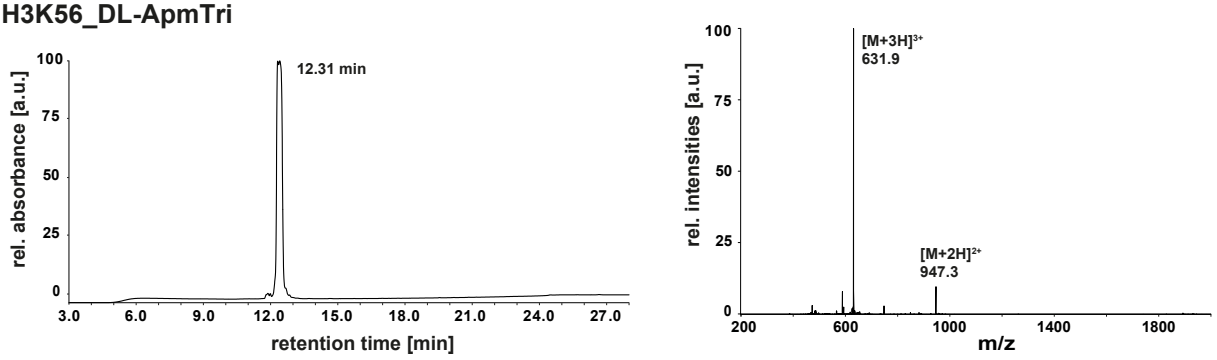


Figure C.39: LC-MS analysis of synthesized peptides H3K14_L-AsuTri, H3K14Q, H3K56 and H3K56ac.

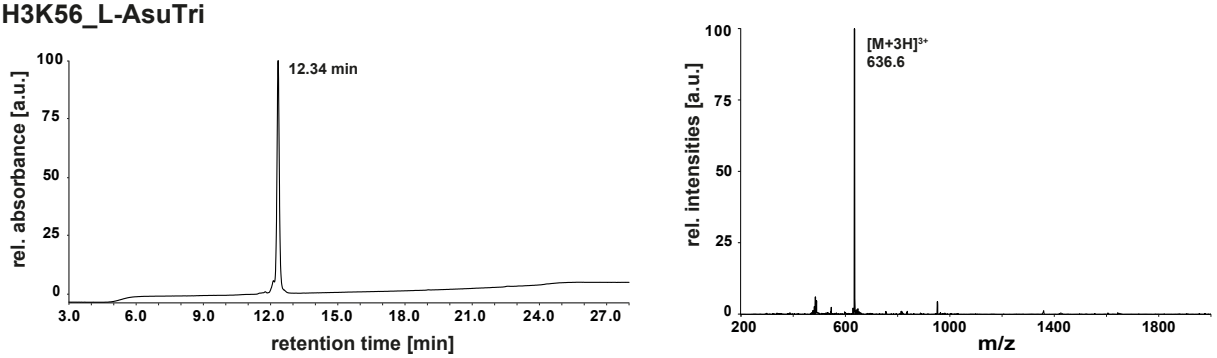
H3K56_L-AadTri



H3K56_DL-ApmTri



H3K56_L-AsuTri



H3K56Q

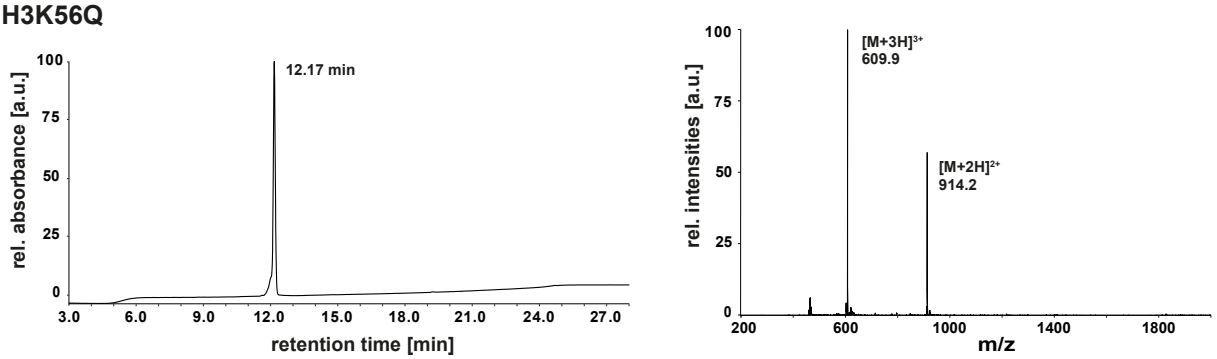
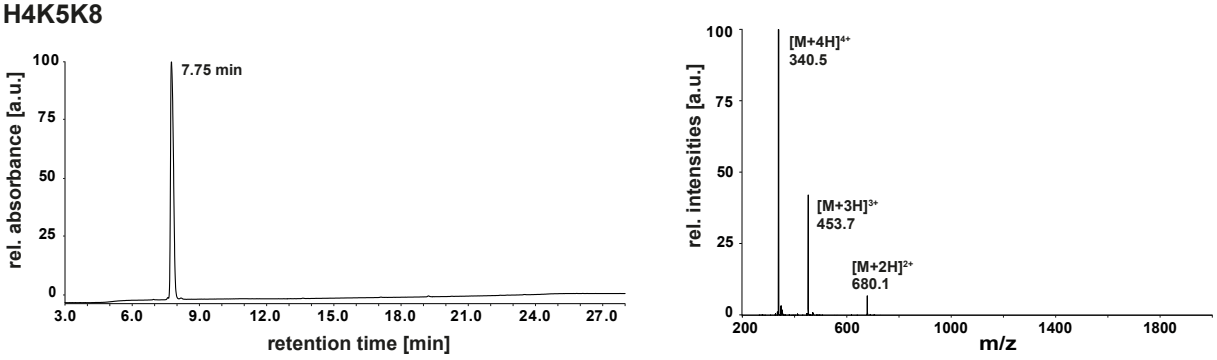
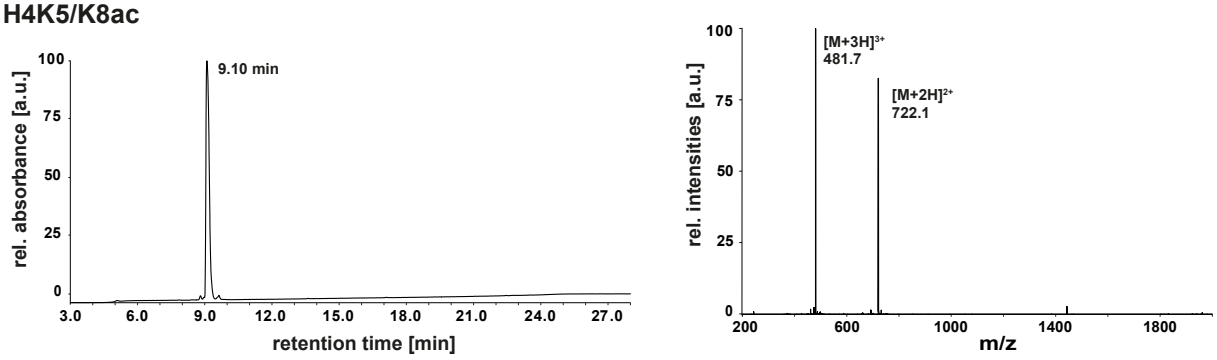


Figure C.40: LC-MS analysis of synthesized peptides H3K56_L-AadTri, H3K56_DL-ApmTri, H3K56_-L-AsuTri and H3K56Q.

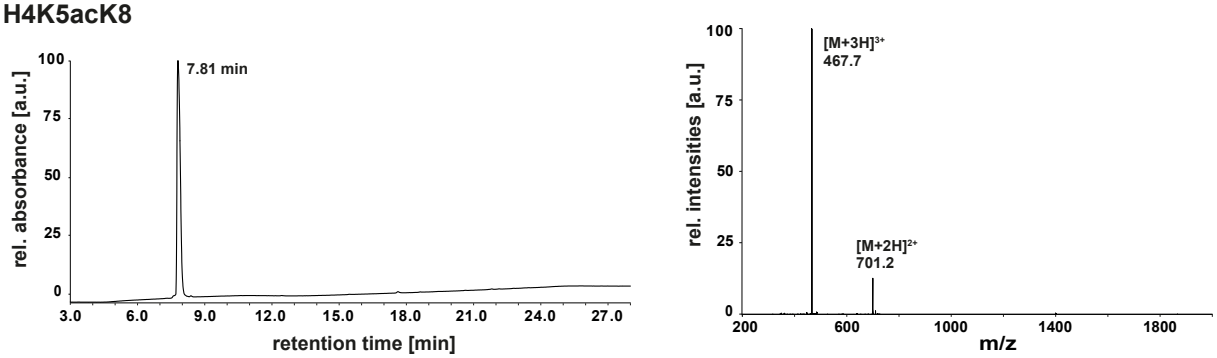
H4K5K8



H4K5/K8ac



H4K5acK8



H4K5K8ac

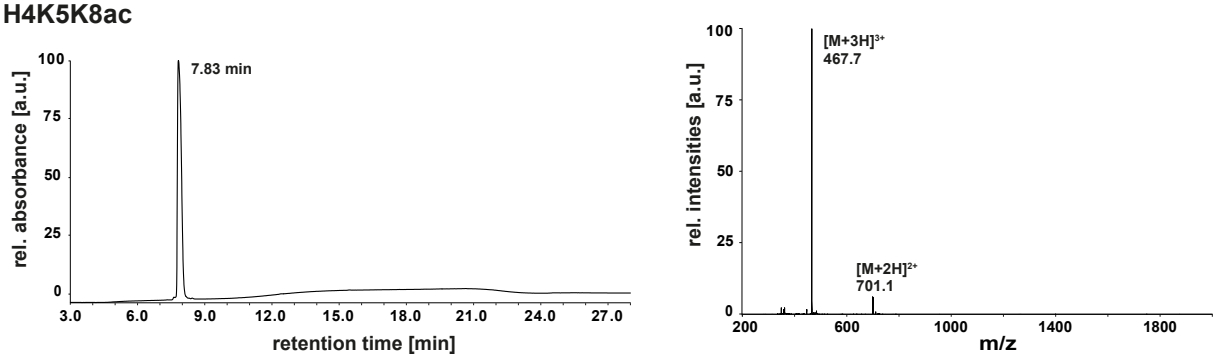
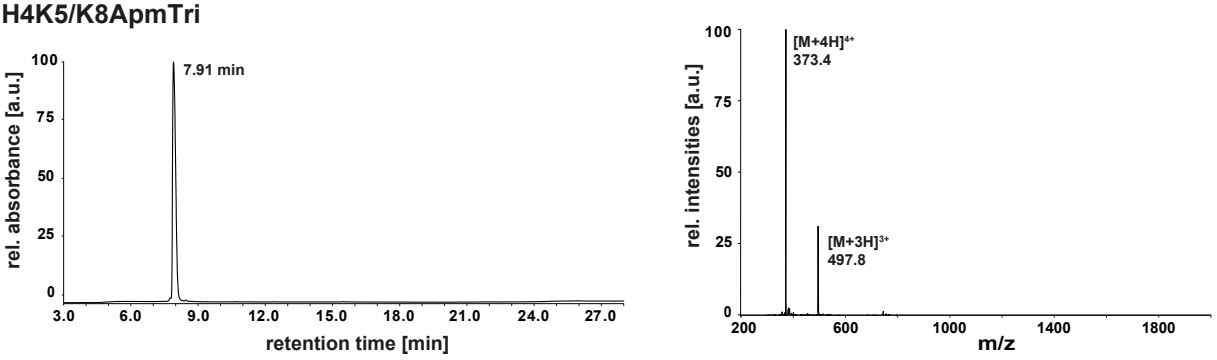
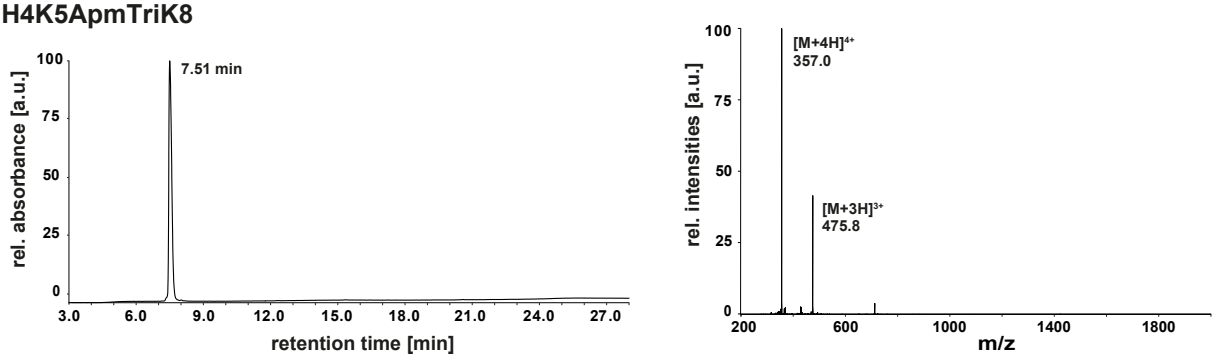


Figure C.41: LC-MS analysis of synthesized peptides H4K5K8, H4K5/K8ac, H4K5acK8 and H4K5K8ac.

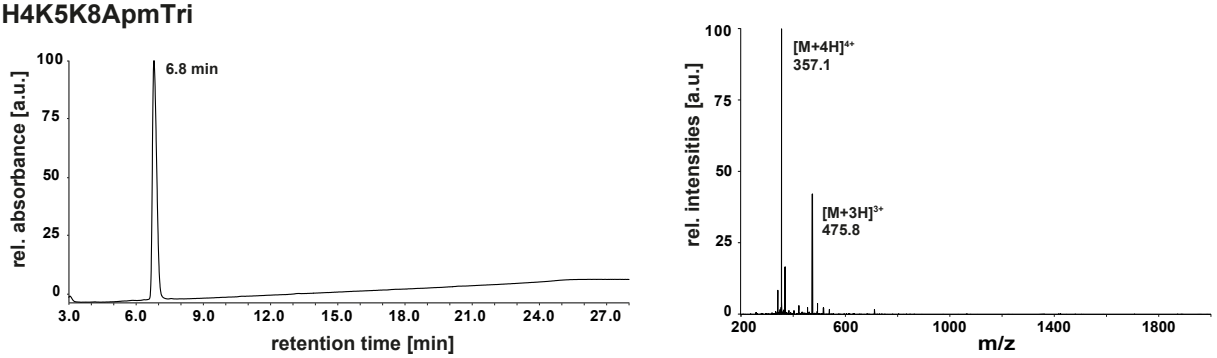
H4K5/K8ApmTri



H4K5ApmTriK8



H4K5K8ApmTri



H4K5ApmTriK8ac

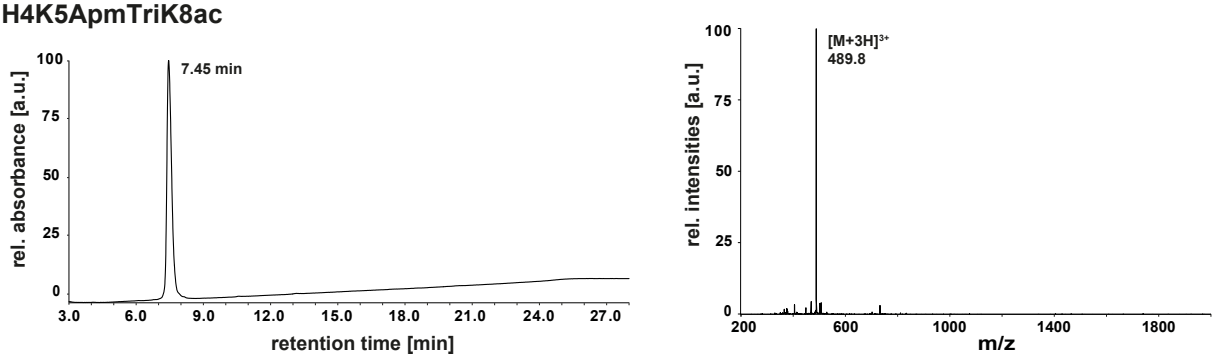
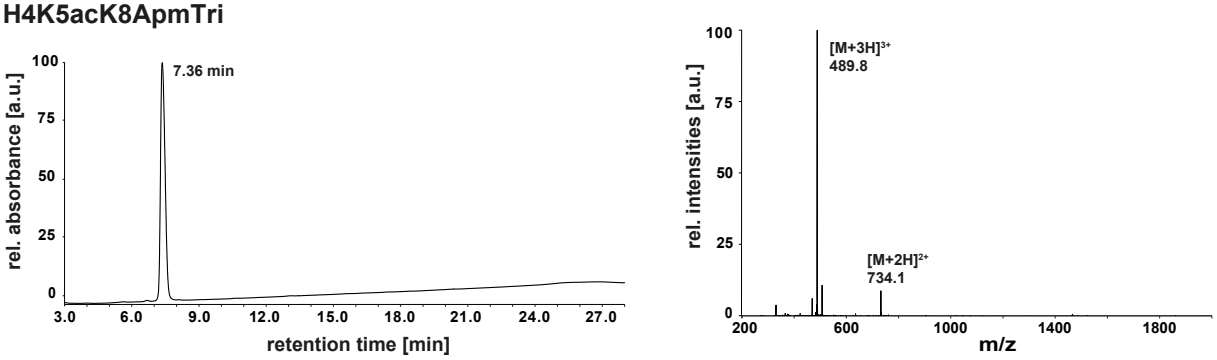
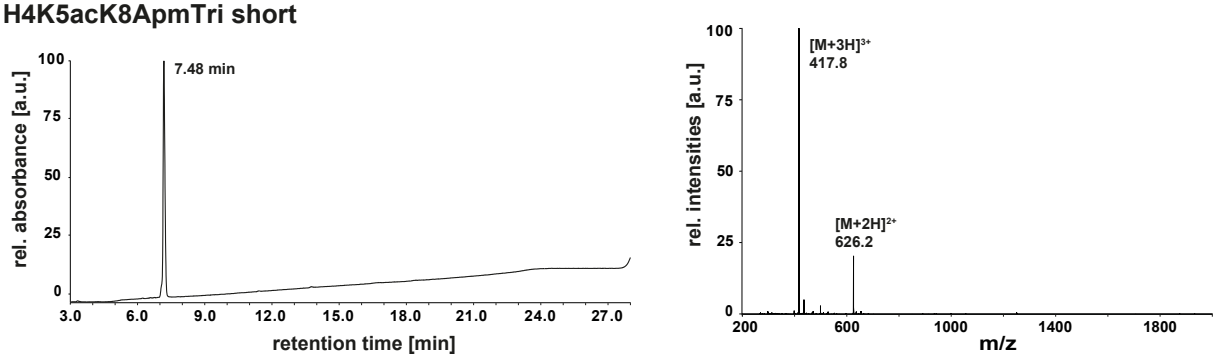


Figure C.42: LC-MS analysis of synthesized peptides H4K5/K8ApmTri, H4K5ApmTriK8, H4K5K8ApmTri and H4K5ApmTriK8ac.

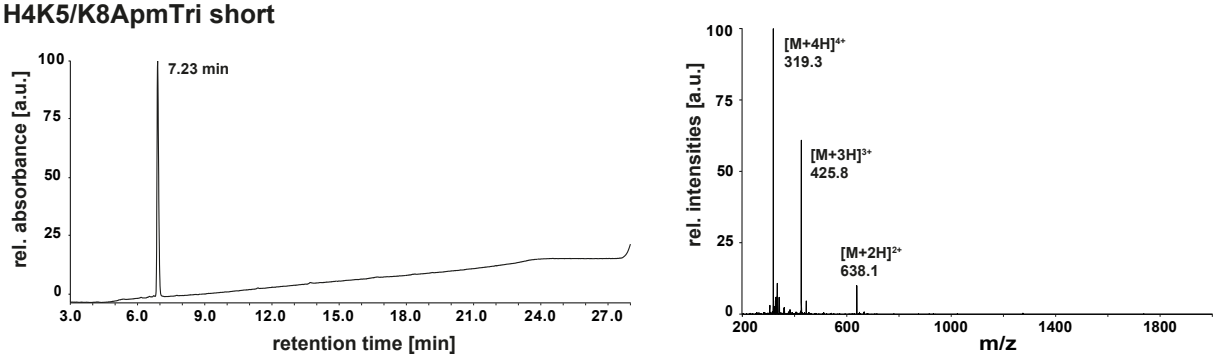
H4K5acK8ApmTri



H4K5acK8ApmTri short



H4K5/K8ApmTri short



Biotin-linker peptide

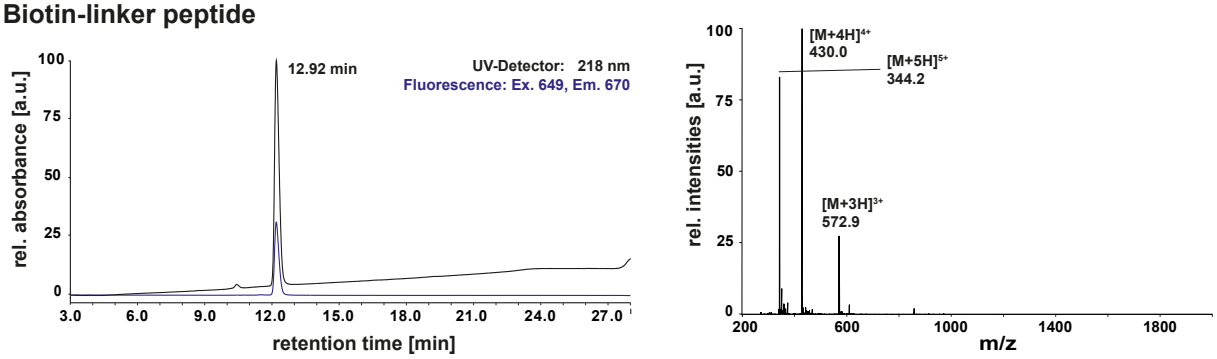
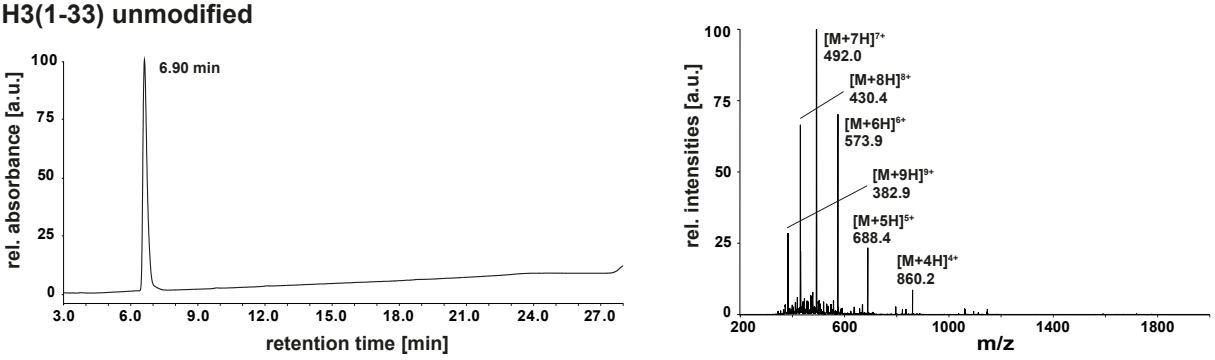
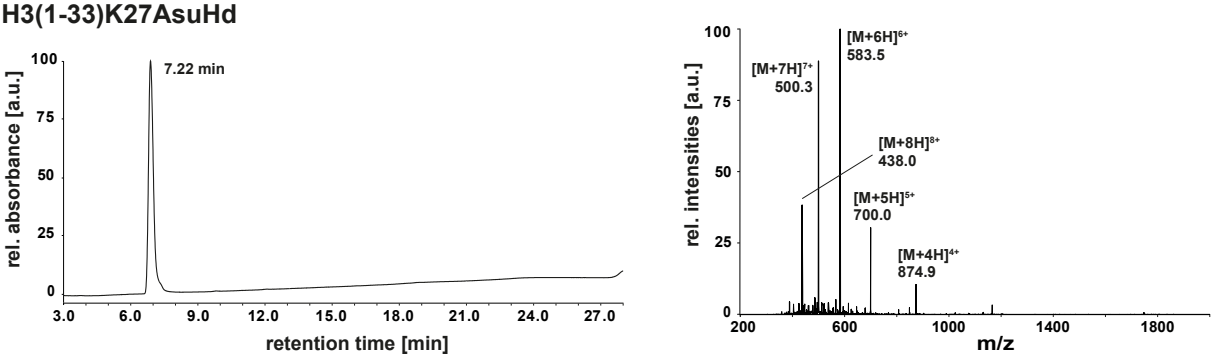


Figure C.43: LC-MS analysis of synthesized peptides H4K5acK8ApmTri, H4K5acK8ApmTri short, H4K5/K8ApmTri short and the biotin-linker peptide.

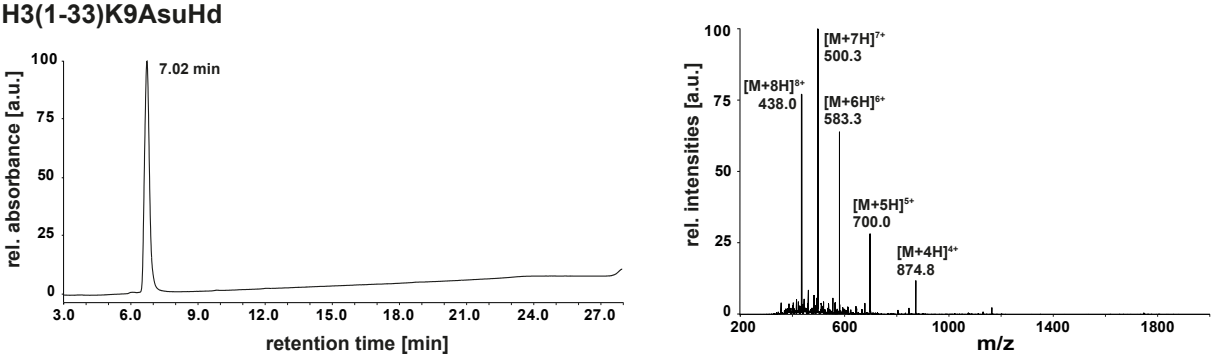
H3(1-33) unmodified



H3(1-33)K27AsuHd



H3(1-33)K9AsuHd



H3(1-33)K4AsuHd

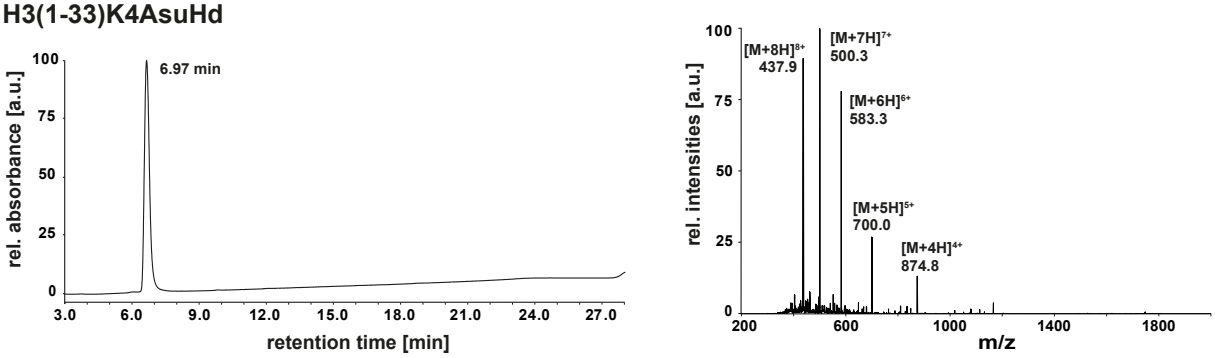


Figure C.44: LC-MS analysis of synthesized peptides H3(1-33) unmodified, H3(1-33)K27AsuHd, H3(1-33)K9AsuHd and H3(1-33)K4AsuHd.