Development of peptide-based ligands for acetyllysine binding modules and proteomic profiling of N-terminal acetyltransferases

Dissertation

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

vorgelegt von
Stefan Schön
aus Freital

Tübingen
2023

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

Tag der mündlichen Qualifikation:	14.11.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Dirk Schwarzer
2. Berichterstatterin:	Prof. Dr. Gabriele Dodt

Eidestattliche Erklärung

Ich erkläre hiermit, dass ich die zur Promotion eingereichte Arbeit mit dem Titel: "Development of peptide-based ligands for acetyllysine binding modules and proteomic profiling of N-terminal acetyltransferases" selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich erkläre, dass die Richtlinien zur Sicherung guter wissenschaftlicher Praxis der Universität Tübingen (Beschluss des Senats vom 25.5.2000) beachtet wurden. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

Tübingen, 8. Dezember 2023 Stefan Schön

"I have loved the stars too fondly to be fearful of the night." - Sarah Williams

Abstract

Acetylation is one of the major co- and posttranslational protein modifications in the cell. Acetyl groups can be installed on the amino groups of the N-terminal (Nt) residues or lysine side chains of proteins by N-terminal acetyltransferases (NATs) and lysine acetyltransferases (KATs), respectively.

In chromatin, acetylated lysines of histones can convey epigenetic signals by recruiting bromodomain (Brd)-containing proteins (BRDs). BRDs can serve as transcriptional regulators and dysfunctions of these proteins is associated with the development of cancer. Consequently, small molecule ligands that bind BRDs have been developed for blocking oncogenic recruitment of chromatin factors. A recently invented non-natural amino acid 2-amino(3-methyl-1,2,4-triazole)pimelic acid (ApmTri) allows establishing a new class of peptide-based BRD ligands, broadening the biochemical toolbox and allowing to explore possible medical applications.

In this thesis, peptide-based ligands for each of the two Brds of BRD3 and BRD4 were developed by optimizing amino acid sequences of known Brd binding sites. Synthetic peptide libraries generated by the SPOT-approach in combination with ApmTri were used for this process. These experiments revealed optimized ligand sequences that exhibited superior Brd interactions in comparison to the native binding site in pulldown experiments. Stability optimization allowed for incorporation of N-methylated and D-amino acids into the sequence without major impact on Brd recruitment. In order to address both Brds of BRD3 and BRD4 simultaneously, optimized binding probes for both Brds were combined in a bivalent ligand construct that interacts efficiently with the full-length proteins. Finally, a set of five cell-penetrating peptides was provided to ultimately enable cellular uptake of the bivalent peptide ligands.

The second part of this thesis was focused on peptide probes for NATs. Malfunction of these enzymes is known to perturb protein homeostasis by altered N-terminal acetylation, resulting in hereditary diseases. A set of two CoA-peptide conjugates serving as bisubstrate NAT inhibitors were synthesized. High resolution LC-MS/MS analyses of interaction of said probes uncovered a distinct impact of the N-terminal residue on NAT recruitment from HeLa lysate resembling the specificity of these enzymes. The experiments further confirmed a switch in substrate specificity of NAA10 that depended on presence or absence of NAA10 binding partners.

Zusammenfassung

Acetylierung ist eine der häufigsten Proteinmodifikationen in der Zelle. Die Acetylgruppe kann durch N-terminale (Nt) Acetyltransferasen (NATs) an die N-α-Gruppe des Protein-N-Terminus oder durch Lysin Acetyltransferasen an die N-ε-Gruppe von Lysinen angelagert werden.
Im Chromatin können acetylierte Lysine in Histonen epigenetische Signale vermitteln, indem sie Bromodomain (Brd)-haltige Proteine (BRDs) rekrutieren. Diese BRDs können die Transkription regulieren, und Funktionsstörungen dieser Proteine werden mit der Entstehung von Krebs in Verbindung gebracht. Infolgedessen wurden kleine Molekülliganden entwickelt, die BRDs binden, um die onkogene Rekrutierung von Chromatinfaktoren zu blockieren. Die kürzliche Entwicklung der nicht-natürlichen Aminosäure 2-amino(3-methyl-1,2,4-triazole)pimelic acid (ApmTri) ermöglicht eine neue Klasse von peptidbasierten BRD Inhibitoren, welche den biochemischen Werkzeugkasten erweitern und den Weg für medizinische Anwendungen eröffnen.
In dieser Arbeit wurden peptidbasierte Inhibitoren für beide Brds von BRD3 und BRD4 entwickelt, indem die Aminosäuresequenzen bekannter Brd Bindungsstellen optimiert wurden. Für diesen Prozess wurden synthetische Peptidbibliotheken verwendet, die mit dem SPOT-Ansatz in Kombination mit ApmTri generiert wurden. Diese Experimente ergaben optimierte Sequenzen, die im Vergleich zu nativen Substraten in Pulldown-Experimenten eine bessere Interaktion zeigten. Zur Stabilitätserhöhung wurden N-methylierte und D-Aminosäuren in die Sequenz eingefügt, ohne die Affinität stark zu mindern. Um beide Brds von BRD3 und BRD4 gleichzeitig zu adressieren wurden optimierte Sonden für beide Brds zu einem bivalenten Liganden kombiniert, welcher effizient mit den Volllängenproteinen interagiert. Abschließend wurden fünf zellpenetrierende Peptide (CPPs) synthetisiert, um die Aufnahme der bivalenten Peptidliganden in die Zellen zu ermöglichen.
Der zweite Teil dieser Arbeit befasste sich mit Peptidsonden für NATs. Eine Funktionsstörung der NATs kann die Proteinhomöostase durch veränderte Nt-Acetylierung deregulieren, was zu Erbkrankheiten führen kann. Es wurden zwei CoA-Peptid-Konjugate synthetisiert, die als Bisubstrat-NAT-Inhibitoren dienen. Hochauflösende LC-MS/MS-Analysen der Interaktion dieser Sonden ergaben einen deutlichen Einfluss des N-terminalen Restes auf die NAT-Rekrutierung aus HeLa-Lysat entsprechend der Spezifität dieser Enzyme. Die Experimente bestätigten außerdem einen Wechsel in der Substratspezifität von NAA10, welcher von der Anwesenheit oder Abwesenheit von NAA10-Bindungspartnern abhängt.

Contents

List of Abbreviations

3-MPA	3-Mercaptopropionic acid
5-FAM	5-Carboxyfluorescein
9-BBN	9-Borabicyclo[3.3.1]nonane
ABC	Ammonium bicarbonate
ABPP	Activity-based protein profiling
ACN	Acetonitrile
ACOT7	Cytosolic acyl coenzyme A thioester hydrolase
ADP	Adenosine diphosphate
ApmTri	2-Amino(3-methyl-1,2,4-triazole)pimelic acid
BET	Bromo and extra terminal domain
BRD	Bromodomain-containing protein
Brd	Bromodomain
BSA	Bovine serum albumin
cCPP	Cyclic cell-penetrating peptide
CHCA	α-Cyano-4-hydroxycinnamic acid
CoA	Coenzyme A
COMU	(1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate
CPP	Cell-penetrating peptide
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition
DIC	N,N'-Diisopropylcarbodiimide
DIPEA	N,N-Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DTNB	5,5′-Dithiobis(2-nitrobenzoic acid)
DTT	Dithiothreitol

EDTA	Ethylenediaminetetraacetic acid
eq	Equivalent
ESI	Electron spray ionization
F_{AZ}	4-Azido-phenylalanine
FA	Formic acid
FASP	Filter aided sample preparation
Fmoc	Fluorenyl-9-methoxycarbonyl
Fmoc-OSu	*N*-(Fluorenyl-9-methoxycarbonyloxy)succinimide
GNAT	(General control non-derepressible 5)-related N-acetyltransferase
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
HDAC	Histone deacetylase
HFIP	Hexafluoroisopropanol
HPLC	High performance liquid chromatography
IMAC	Immobilized metal ion affinity chromatography
iMET	N-terminal initiator methionine
K_D	Dissociation constant
KAT	Lysine acetyltransferase
KDAC	Lysine deacetylase
LC-MS	Liquid chromatography mass spectrometry
LFQ	Label-free quantification
Lys(Dde)	*N*-ε-1-(4,4-Dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl-L-lysine-OH
Lys-C	Lysyl endopeptidase
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
NAD^+	Nicotinamide adenine dinucleotide

NAT	N-terminal acetyltransferase
NCP	Nucleosome core particle
NHS	*N*-Hydroxysuccinimide
NLS	Nuclear localization signal
NMC	NUT-midline carcinoma
NMM	*N*-Methylmorpholine
NMP	*N*-Methyl-2-pyrrolidone
NUT	Nuclear protein in testis
OD	Optical density
OGFOD1	2-Oxoglutarate and Fe(II)-dependent oxygenase domain-containing protein 1
Oxyma	Ethyl cyanohydroxyiminoacetate
PANK4	Pantothenate kinase-like protein
PBS	Phosphate-buffered saline
PD	Pull-down
PEG	Polyethylene glycol
PPI	Protein-protein interaction
Pra	Propargylglycine
PROTAC	Proteolysis targeting chimera
PS	Positional scanning
PTM	Post-translational modification
PVDF	Polyvinylidene fluoride
PyOxim	[Ethyl-cyano-(hydroxyimino)-acetato-O^2]-tri-1-pyrrolidinylphosphonium hexafluorophosphate
RaPID	Random nonstandard peptide integrated discovery
RP-HPLC	Reversed-phase high-performance liquid chromatography
rpm	Rounds per minute
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis

SM-A	Stripping mixture A
SM-B	Stripping mixture B
SPCL	Synthetic peptide combinatorial library
SPPS	Solid-phase peptide synthesis
StageTips	Stop-and-go-extraction tips
TBTA	Tris((1-benzyl-4-triazolyl)methyl)amine
TCP	Tritylchloride-polystyrene
TEAB	Triethylammonium bicarbonate buffer
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TFE	2,2,2-Trifluoroethanol
THF	Tetrahydrofuran
TIPS	Triisopropyl silane
TOC	Total organic carbon
Tris	Tris(hydroxymethyl)aminomethane
UV-Vis	Ultraviolet–visible
WCE	Whole cell extract

1 Introduction

1.1 Chromatin and histone proteins

The genetic information of eukaryotic cells is stored mainly in the nucleus and partly in the mitochondria in the form of a DNA double helix. Humans, as eukaryotic organisms, have a nuclear genome consisting of 3 billion base pairs which are distributed over 23 pairs of chromosomes [1]. With the single DNA strand measuring an impressive 2 m in length [2], the human genome must not only fit into the cell nucleus with a diameter of 6 µm [3], but must also be accessible for transcription, gene regulation and replication. This challenge is overcome by condensing the genetic information in a complex of DNA and scaffolding proteins, called chromatin.

The basic unit of chromatin is the nucleosome, which consists of a globular histone octamer around which the DNA strand is wrapped. The histone octamer is composed of the four core histone proteins H2A, H2B, H3 and H4. In detail, a heterotetramer consisting of two copies of H3 and two copies of H4 interacts with two H2A/H2B heterodimers to form the octamer (see figure 1.1). A DNA double strand of around 146-147 base pairs is wrapped around the histone octamer as a left-handed superhelix, constituting the nucleosome core particle (NCP) [4, 5]. The single NCPs are connected with a strand of linker DNA, forming the ”beads on a string” state of chromatin. Higher order condensation is achieved with the addition of the linker histone protein H1, through which the ”beads on a string” form is more tightly coiled together into a 30 nm wide chain known as the 30 nm chromatin fiber. While there are different theories regarding the arrangement of the NCPs within the 30 nm chromatin fiber, its structural details and prevalence in the cell remain poorly understood [6–8]. Eventually, additional non-histone scaffolding proteins enable the formation of the mitotic chromosome, the highest order of DNA condensation in the cell [9].

Chromosomal regions of eukaryotes can be grouped into two distinct classes depending on their condensation state: euchromatin and heterochromatin. Euchromatin is a loosely packed form of chromatin resembling beads on a string, which makes genes and regulatory areas of the DNA accessible. The DNA in heterochromatin, on the other hand, is tightly packed and therefore less transcribed [10]. Heterochromatin can be further divided into constitutive heterochromatin, which is mostly composed of repetitive DNA sequences and is found primarily in centromeres, telomeres and at the nuclear membrane [11], and facultative heterochromatin, which is transcriptionally inactive but can be activated due to certain developmental or environmental factors [12].

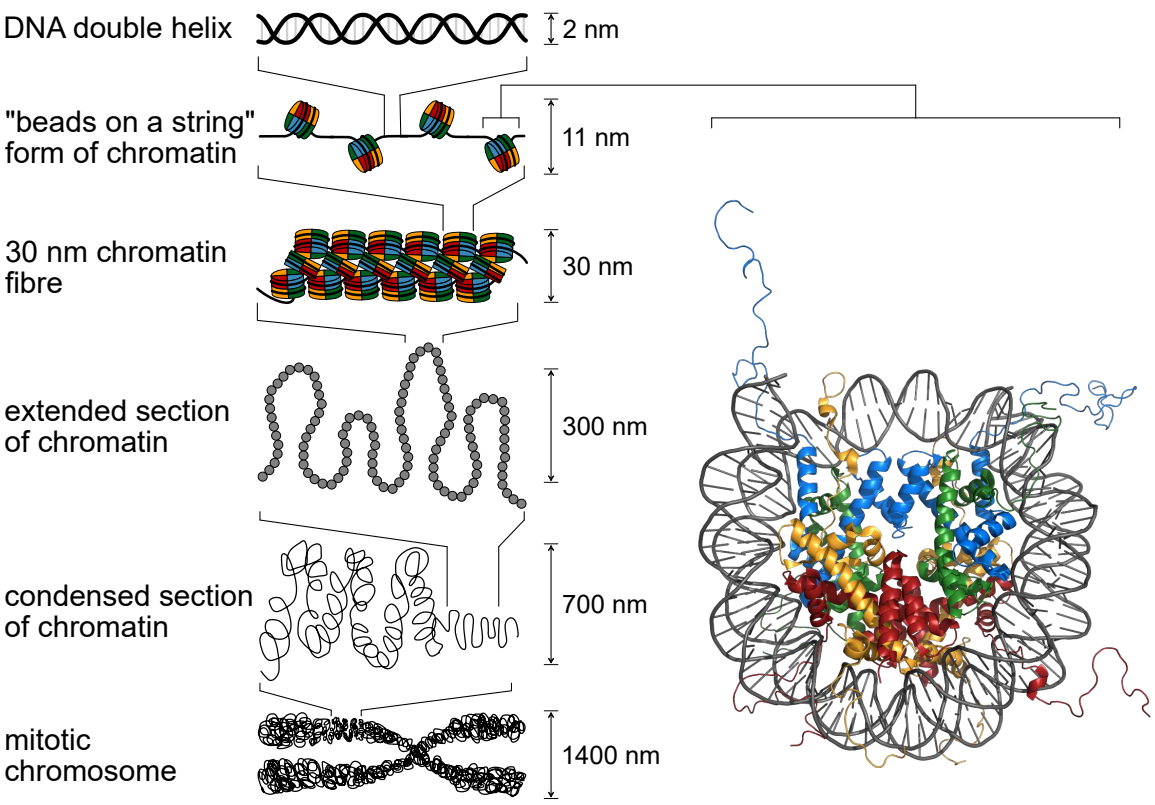


Figure 1.1: The hierarchical organization of DNA compaction. Left side: Major structure elements of chromatin with their approximate sizes, adapted from [13]. Right side: Crystal structure of the nucleosome core particle at a resolution of 1.9 Å (PDB-ID: 1KX5) [14].

In general, gene regulation is an important mechanism in all forms of life. Amongst others, the condensation state of chromatin, but also the recruitment of transcription factors are means to control gene output and can be regulated through the chemical modification of histone proteins and DNA [15, 16]. The four core histones that form the histone octamer each consist of a globular part and a so-called histone tail. While the negatively charged phosphate backbone of the DNA interacts with the positively charged surface of the histone core, the histone tails protrude to the outside, enabling interactions with the DNA as well as with other proteins and nucleosomes [15]. Post-translational modifications (PTMs) can be installed on the histone tails, thereby altering their interaction profile with other proteins as well as with the DNA [17]. Acetylation of lysine, phosphorylation of serine, tyrosine and threonine as well as methylation of lysine and arginine are some of the most prevalent PTMs found on the tails of histones while some sites were also found in the

globular part. Other modifications include citrullination of arginine, O-GlcNAcylation of serine and threonine, adenosine diphosphate (ADP) ribosylation of glutamate and arginine as well as sumoylation and ubiquitylation of lysine residues [15]. A selection of PTMs and their location on the N-terminal histone tails is displayed in figure 1.2

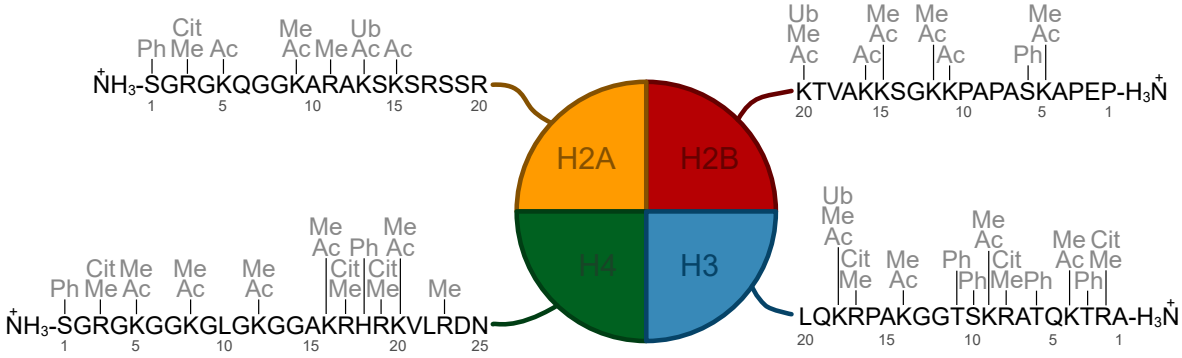


Figure 1.2: Selection of histone modifications found at the tails of the four histone core proteins. Modifications are abbreviated as follows: Ph = phosphorylation, Cit = citrullination, Me = methylation, Ac = acetylation, Ub = ubiquitylation [18]. Adapted from [19]. The DNA strand is not displayed.

Two models aim to explain the influence of histone modifications on the transcription of genes: the electrostatic model and the histone code hypothesis. The electrostatic model is based on the potential interaction of basic residues accommodated on the histone tails with the negative phosphate backbone of DNA. PTMs that neutralize the positive charge of the basic side-chains, like lysine acetylation or arginine ADP ribosylation, can thereby weaken the attraction between DNA and histones, leading to a less tightly packed chromatin state that results in elevated levels of gene expression [20–22]. While phosphorylation of hydroxyl side-chains introduces a negative charge to the histone, contributing to electrostatic repulsion and hence less densely packed chromatin, it was shown that phosphorylation can up- or down-regulate gene expression depending on its location on the histone tails [23, 24]. An even more complex PTM is the methylation of lysines, which can be mono-, di-, or tri-methylated. Depending on the degree of methylation and the position, lysine methylation can up-regulate transcription or correlates with gene repression [25].

Effects observed for PTMs like phosphorylation and methylation go beyond the explanations provided by the simple electrostatic model and require a more complex approach: the histone code hypothesis. The main concept of the histone code hypothesis is that each histone modification possesses designated binding proteins (chromatin readout) that

interact with this PTM in a sequence-specific manner [26]. Interaction between these "reader" proteins on one tail but also across different tails results in a complex interplay of downstream events. One example of these cross-talk interactions is the methylation of Arg at position 17 on histone 3 (H3R17) by the methyltransferase CARM1, which is promoted by the acetylation of H3K18 and H3K23 [27]. Besides influencing the modification state of neighboring residues, PTMs were shown to directly influence the gene activity through interaction of their readers imbedded into regulatory proteins of transcription [28].
On top of the reader proteins, there are two other classes of proteins involved in gene regulation through PTMs: the so-called "writer" proteins that introduce the PTM onto the histones and the "eraser" proteins that remove them. One set of writer proteins are the histone methyltransferases that are divided into three groups: the SET-domain containing, non-SET-domain containing and arginine methyltransferases which methylate lysine and arginine utilizing *S*-adenosyl methionine as a cofactor [29]. Initially, histone methylation was thought to be a stable and irreversible modification, but later findings uncovered two main classes of demethylases: flavin adenine dinucleotide (FAD)-dependent amine oxidases and α-ketoglutarate-dependent hydroxylases [30]. Both histone methyltransferase and demethylases exhibit relative high specificity towards target sites and the degree of methylation [31]. Phosphorylation of histones is catalyzed by protein kinases, while protein phosphatases can remove the phosphoryl group. Acetylation of lysines is another prevalent and well studied PTM of histones and involves lysine acetyltransferase (KATs) for introduction and lysine deacetylases (KDACs) for removal [32]. High specificity and interconnectivity of reader, writer and eraser proteins allow the cells to utilize the potential of billions of possible PTM patterns to fine-tune the transcription of individual genes and thereby enabling cellular homeostasis.

1.2 Protein Acetylation

In eukaryotes, protein acetylation is one of the major PTMs, affecting a variety of factors like protein stability, protein localization, protein-protein interactions (PPIs) and gene regulation, amongst others [33–35]. Acetyl groups can be introduced to the Nt-α-amino group of the N-terminal (Nt) amino acid of a protein, as well as to the ε-amino group of lysines. The modification is facilitated by Nt-acetyltransferases (NATs) or lysine acetyltransferases (KATs), respectively, utilizing Acetyl-Coenzyme A (CoA) as a cofactor (see figure 1.3). The added acetyl group neutralizes the charge and usually prevents further modification of this site [36].

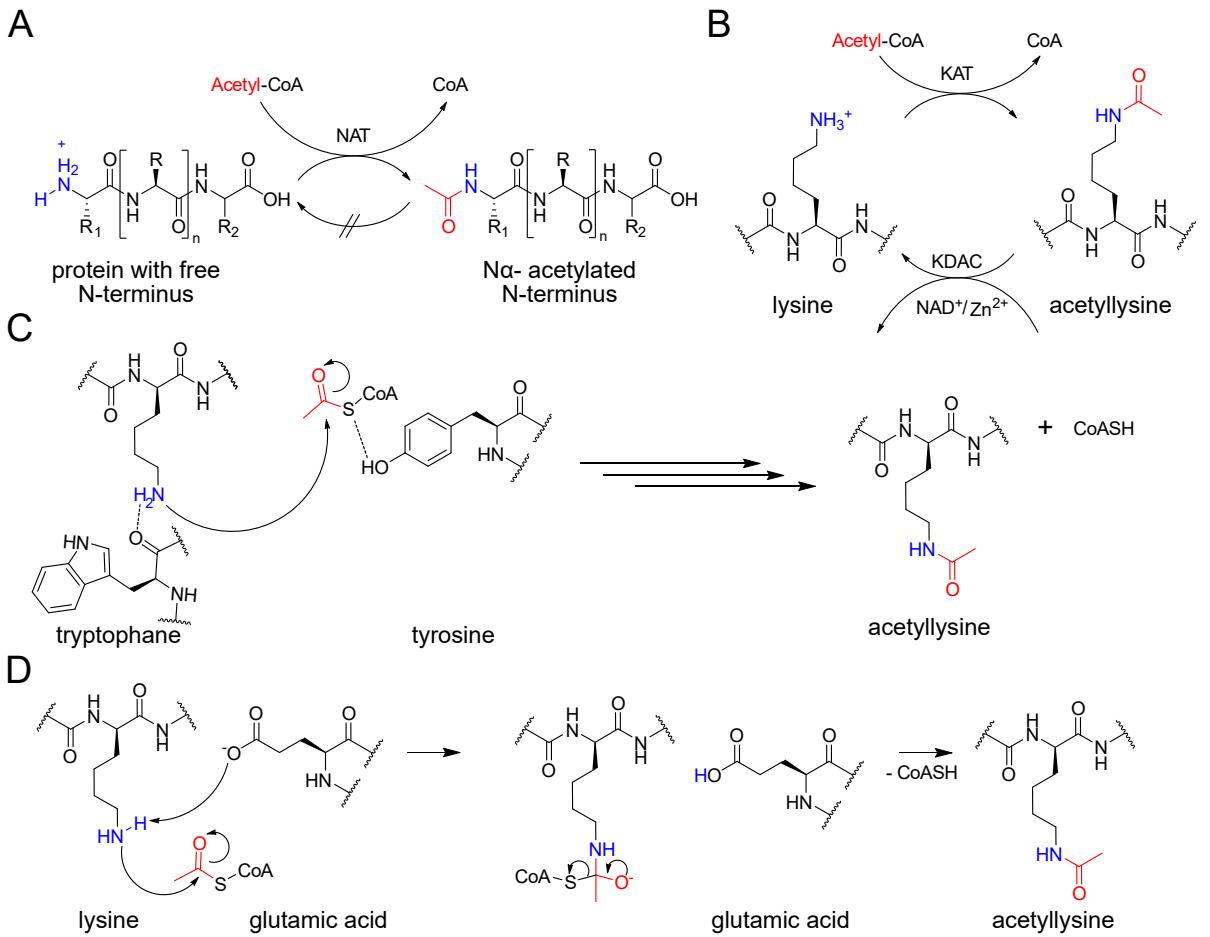


Figure 1.3: Protein acetylation. A) N-terminal acetylation of a protein, catalyzed by an Nt-acetyltransferase. B) Lysin acetylation is catalyzed by KATs utilizing acetyl-CoA as a cofactor. The reverse reaction is catalyzed by KDACs using either nicotinamide adenine dinucleotide (NAD^+) or Zn^{2+} as a cofactor. C) The catalysis of the p300/CBP family of HATs relies on a Theorell Chance mechanism of action. D) A different mechanism is observed in members of the GNAT and MYST families of HATs which involves a ternary complex with a glutamate and water (not displayed). Adapted from [36–38]

It was shown that 80-90 % of all human proteins are Nt-acetylated in a co-translational fashion [39]. While co-translational acetylation of the N-terminus is the more prevalent mode, evidence suggests that proteins can also undergo post-translational Nt-acetylation [40, 41]. Nt-acetylation can be incomplete, leading to the same protein being present in its unacetylated and its Nt-acetylated form concurrently in one cell [42]. So far, no Nt-deacetylase has been reported, thus Nt-acetylation is considered to be irreversible.

The other form of protein acetylation, lysine acetylation, was first found on histone proteins where it plays a role in gene regulation [43]. Therefore, the enzymes catalyzing

the acetylation and deacetylation of lysines were termed histone acetyltransferases (HATs) and histone deacetylases (HDACs), respectively [43, 44]. Later, acetylation of non-histone proteins was discovered and the writer and eraser enzymes were consequentially renamed to lysine acetyltransferases (KATs) and lysine deacetylases (KDACs) [45], although the abbreviations HAT and HDAC are still used in the literature. The well-regulated interplay between KATs and KDACs influences many cellular functions, including metabolic process and homeostasis. A malfunction of these delicate interactions can lead to severe diseases including cancer and neurodegenerative disorders [46, 47].

1.2.1 Lysine acetyltransferases (KATs)

KATs catalyze the acetylation of the ε-amino group of lysine, utilizing acetyl-CoA as cofactor. The overall number of human KATs is still being disputed, 17 proteins with exclusive lysine acetyltransferase activity have been identified so far while 5 further proteins display acetyltransferase activity in addition to their originally described enzymatic activity, adding up to a total of 22 KATs [45, 48]. KATs can be grouped into three major families: the Gcn5(general control non-derepressible 5)-related N-acetyltransferases (GNAT) family, the MYST family (MYST is an acronym of the first four identified members MOZ, Ybf2, Sas2, and Tip60) and the p300/CBP (CREB-binding protein) family. The histone acetyltransferase p300 (also known as EP300) is a paralog of CBP. Although they have a high sequence homology and structural similarity, they differ in their function and substrates [49]. In the members of the GNAT family, up to four conserved motifs A-D can be found in their catalytic KAT domains, which comprises around 160 amino acids. Amongst these, motif A (Arg/Gln-X-X-Gly-X-Gly/Ala) is the most highly conserved and is responsible for acetyl-CoA recognition and binding [50]. The GNAT family members acetylate lysines of histones H2B, H3 and H4 by first deprotonating the ε-amino group through a conserved glutamate residue and a water molecule in a ternary complex together with acetyl-CoA (shown in fig 1.3D). This is followed by the nucleophilic attack of the amine onto the carbonyl carbon in the spatially close acetyl-CoA [51]. KATs belonging to the MYST group acetylate lysines in a similar manner as members of the GNAT family, which was for example shown for the MYST HAT Esa1 as part of the piccolo NuA4 complex [52]. It was however also found that MYST family members can use a different mechanism: Here, the acetyl group is transferred to a conserved cysteine residue through the nucleophilic attack of the thiol onto the carbonyl carbon of acetyl CoA, forming a new thioester [51, 53]. The mechanism then follows that of the GNAT family. The substrates

of MYST family members are restricted to lysines found on histone H2A, H3 and H4. Their homologous regions stretch from a zinc-finger binding domain over a cysteine-rich region in their N-terminal part to a shared motif A that is also commonly found amongst members of the GNAT family [50]. In contrast to the other two families, members of the p300/CBP family lack the conserved glutamate residue for catalysis. Two residues, a conserved tyrosine and a tryptophan, were found to be important for the acetylation of the targets (shown in fig 1.3C). It was proposed that the tyrosine acts as a general acid while the tryptophan orientates the lysine into the active site, leading to a Theorell Chance mechanism of acetylation, in which no ternary complex is formed, which is in stark contrast to the mechanisms employed by KATs of the two formerly described families [51]. Members of the p300/CBP family have been observed to acetylate many non-histone targets like non-histone chromatin proteins, transcription factors and nuclear receptor coactivators, amongst others [54].

Most of the KATs associate with other subunits to form multiprotein complexes that refine their catalytic activity and substrate specificity, effectively being a means to regulate their functionality [55]. Some non-catalytic subunits in the complexes interact with other PTMs of histones to direct the acetylation machinery to certain loci on chromatin [56]. The impact on the specificity of the complex can go in both directions, either leading to a more broad or to a more restricted specificity than that of the single catalytic KAT subunits. Members of the MYST family for example have a restricted target specificity when they associate with their subunits into a complex, while members of the GNAT family can acetylate more sites on their target proteins while being present in their complex [50].

1.2.2 Lysine deacetylases (KDACs)

The counter-players of KATs are the KDACs, which remove the acetylation from ε-amino groups of lysines. Lysine deacetylases were first identified in 1995 in yeast, where they were found to exhibit deacetylase activity on histone proteins [57]. Overall, there have been 18 human KDACs identified so far, which can be divided into two groups: classical KDACs which are Zn^{2+} dependent and so-called sirtuins which utilize NAD^{+} as a cofactor. A further categorization divides them into the sub-class I Rpd3-like proteins, sub-class II Hda1-like proteins and sub-class IV with only one member, whereas sirtuins are categorized as class III Sir2-like proteins [37].

Several zinc dependent KDACs are located in the nucleus where they deacetylate chromatin while other HDACs have important functions in the cytoplasm, for example HDAC 6 is

known to deacetylate alpha tubulin. This also reflects on their substrates: class I and some members of class II KDACs generally deacetylate histone molecules and transcription factors, thereby playing a role in gene regulation. Class III KDACs localized in the nucleus play a role in heterochromatin formation. Sirtuins located in the mitochondria, mainly SIRT3, SIRT4 and SIRT5, influence mitochondrial metabolic processes [58]. The single member of class IV lysine deacetylases, HDAC11, seems to be a modulator for several immunological functions of the cell. It has been found to negatively regulate the expression of interleukin 10 in mouse and human macrophages, thereby effectively elevating the inflammatory response of the organism [59].

Fine-tuning of the acetylation state of chromatin is made possible by the interplay between KATs and KDACs. However, acetylation of lysines not only leads to a charge neutralization, it can also attract additional gene regulatory proteins and chromatin-modifying proteins containing a bromodomain (Brd), a protein domain that selectively binds acetyllysine.

1.2.3 Bromodomain-containing proteins (BRDs)

Brds are the largest group of known chromatin readers and comprise approximately 110 amino acids. The first Brd was identified in the drosophila gene *Brahma/brm*, encoding a protein similar to SNF2/SWI2, which plays a role in the transcriptional activation in yeast [60]. The ability of Brds to bind acetylated lysine was recognized later [61]. Noteworthy, other domains that recognize acetyllysine exist, namely tandem plant homeodomains (PHD) zinc fingers and tandem pleckstrin-homology (PH) domains have been shown to bind histones in an acetylation-sensitive manner, albeit their affinity to non-acetylated histones is also reported [62, 63]. Recently, the YEATS domain (named after the first discovered proteins containing this domain: Yaf9, ENL, AF9, Taf14, and Sas5) has been found to bind to several monoacetylated sequences derived from the histone H3 protein with low micromolar affinity [64].

The Brd is composed of a left-handled bundle of four alpha helices (αZ, αA, αB, αC) which are linked by loop regions called ZA, AB and BC loops that vary in length and charge, contributing to the overall low sequence homology of Brds [65] (fig 1.4). The loop regions determine the surface properties of the protein close to the bound peptide, thereby being responsible for the binding specificity of the Brds [66]. The bound acetyllysine is located in a central hydrophobic binding pocket, where it interacts with a conserved asparagine residue, a binding mode conserved in most but not all bromodomain-containing proteins (BRDs) [67].

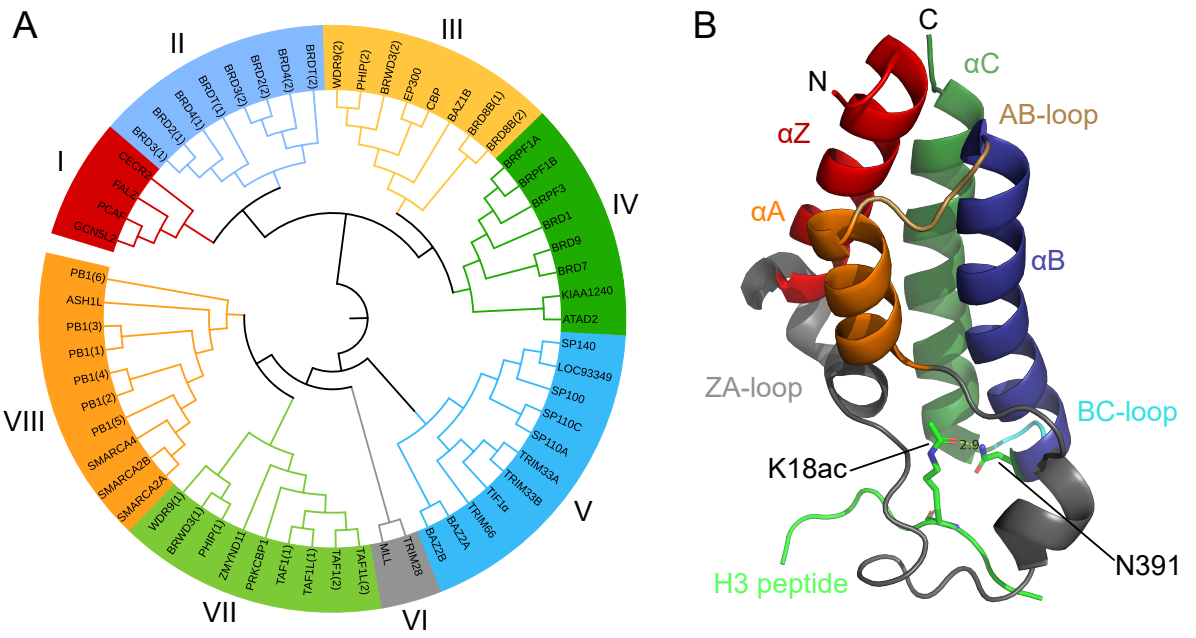


Figure 1.4: Phylogenetic tree of Brd families and general structure of the Brd. A) Structure based mid-point rooted phylogenetic tree of human Brd families. The subfamilies are indicated by Roman numbers I-VIII, adapted from [68]. B) The general structure and binding of Brds to acetylated lysine, exemplified with BRD3(2) in a complex with H3K18ac. Structural elements are labeled, the conserved asparagine as well as the acetylated lysines are indicated (PDB code: 5HJC).

Brds are part of several nuclear proteins including KATs, chromatin-remodeling complexes, helicases, methyltransferases, nuclear-scaffolding proteins as well as transcriptional coactivators and mediators that mainly recognize acetylated histones [66, 67]. Bromodomain containing proteins can also hold multiple Brds as well as other functional domains like PHD domains, catalytic acetyltransferase domains, and domains that moderate protein-protein interaction like bromo-adjacent homology (BAH) domains and kinase-inducible domain-interacting (KIX) domains [69, 70]. So far, 61 human Brds have been identified which are part of 46 different proteins [68]. These 46 bromodomain-containing proteins can be clustered into 8 distinct families [66].

Members of the subfamily I are all localized in the nucleus and include the acetyl-transferase containing proteins p300/CBP-associated factor (PCAF), the general control of amino-acid synthesis 5-like 2 (GCN5L) acetyltransferase and the chromatin remodeling factor cat eye syndrome chromosome region 2 (CECR2) [61, 71, 72]. PCAF was the first BRD uncovered as an acetyllysine binder, which also supported the notion that Brds direct KATs towards specific locations in chromatin [61]. Binding specificity of subfamily I members

has been studied in detail, with PCAF binding acetylated lysines of histone H3 and H4 [66, 73], but it has also been shown to bind acetylated HIV-1 transcriptional trans-activator Tat, hinting towards a role in viral pathogenesis [73]. GCN5L was likewise found to bind acetyllysines located on H2A (K5ac) [74], H3 (K9ac, K14ac, K9ac/K14ac) and H4 (K5ac/K8ac/K12ac/K16ac) [75]. The chromatin remodeling factor CECR2 interacts with several acetylated lysines on histone H3 (K9ac, K14ac) [66].

Prominent members of subfamily III are the CREB binding protein (CBP) and E1A binding protein p300, both containing a lysine acetyltransferase domain [76]. These two proteins have been well studied to explore the navigation of KAT activity towards certain loci in the chromatin by recognizing acetylated lysine via the Brd. The BRD p300 was observed to bind H3 (K36ac, K56ac) and H4 (K12ac, K20ac, K44ac) [66], while CBP has been found to bind to acetyllysines on histone H2B (K85ac), histone H3 (K9ac/K14ac, K14ac, K36ac, K56ac) and histone H4 (K12ac, K20ac, K44ac) [66, 73]. One study found that binding of CBP to H3K14ac was enhanced 6-fold when the serine on position 10 was phosphorylated, demonstrating the regulatory effects of PTM interplay [66]. CBP exhibits affinity towards protein targets beyond histones and binds to acetylated p53 as part of the DNA damage response [73].

Subfamily IV is not well studied as of today, albeit some interactions with histones H3 (K9ac, K14ac) and H4 (K8acK12ac, K16ac) have been reported for the bromodomain-containing protein 7 (BRD7) [77]. Additional studies of the ATPase family AAA (ATPases Associated with diverse cellular Activities) domain-containing protein 2 (ATAD2), another member of this subfamily, have found to play a role in cell cycle progression and is overexpressed in several human tumors which leads to elevated transcriptional activity of the MYC oncogene [78]. Binding of ATAD2 to H3 (K14ac) [79] and of ATAD2B (KIAA1240) to H4 (K5ac) [66] have also been reported.

A common structural element of members of the subfamily V of BRDs is a PHD/BRD tandem module. Important members of this subfamily are the two proteins bromodomain adjacent to zinc finger domain 2A (BAZ2A) and bromodomain adjacent to zinc finger domain 2B (BAZ2B) [68]. BAZ2A is a regulatory subunit of the nucleolar remodeling complexes 1 and 5 (NoRC-1 and NoRC-5), both which regulate the spacing of nucleosomes in the chromatin fiber [80]. BAZ2B, a paralog of BAZ2A, is part of similar chromatin remodeling complexes, namely BAZ2B-associated remodeling factor 1 and -5 (BRF-1, BRF-5), which enable access to DNA for transcription, DNA repair and DNA replication [80]. Binding of H3K14ac has also been reported for BAZ2B [66].

With only two members, the histone methyl-transferase myeloid/lymphoid or mixed

lineage leukemia (MLL) and the transcriptional co-regulator tripartite motif-containing 28 (TRIM28), subfamily VI is the smallest class of Brds with little information about binding sites [68]. However, both Brds play a role in gene regulation, with MLL exhibiting histone methyltransferase activity as part of the MLL1/MLL complex [81] involved in chromatin remodeling [82]. TRIM28 recruits Mi-2α/CHD3, a subunit of the nucleosome remodeling and deacetylation (NuRD) complex, as well as SETDB1, which leads to methylation of promoter regions and silencing of the targeted genes [83, 84].

Subfamily VII includes transcriptional repressors like the zinc finger MYND domain containing protein 11 (ZMYND11) [85], transcriptional activators like TATA box binding protein (TBP)-associated factor (TAF1) [86] and TAF1-like TATA box binding protein (TBP)-associated factor (TAF1L) [87] as well as the chromatin remodeling factor WD repeat domain 9 (WDR9 domain 1) [88], amongst others. TAF1 is part of the multiprotein transcription factor IID (TFIID) complex, which facilitates the initiation of DNA transcription through RNA polymerase II [89]. This complex is composed of the TATA-binding protein (TBP) and multiple TAF subunits [89] in which TAF1 acts as the core scaffolding component important for the complex assembly [90]. It furthermore possesses acetyltransferase activity [91] as well as the ability to bind multiple acetylated lysines on histone H4 (K8ac/K16ac, K5ac/K12ac, K5ac/K8ac/K12ac/K16ac), which can stabilize the TBP-TATA box complex [92, 93]. Affinity of TAF1 to di-acetylated p53 (K373ac/K382ac) has also been shown, hinting towards targets and functions beyond chromatin and gene regulation [94].

Members of the last BRD subfamily VIII are the methyl transferase ash1 (absent, small, or homeotic)-like (ASH1L), the chromatin remodeling factors SWI/SNF-related matrix associated actin-dependent regulator of chromatin a2 and a4 (SMARCA2 and SMARCA4) and the polybromo 1 (PB1) protein. SMARCA2 and SMARCA4 have been reported to bind H3 (K9ac, K14ac and K9ac/K14ac) and H4 (K8ac, K12ac, K16ac and K5ac/K8ac/K12ac/K16ac) [66, 75] as well as H2B (K5ac), H3 (K14ac, K9ac/K14ac) and H4 (K8ac and K12ac, K16ac), respectively [95, 96]. As part of the SWItch/Sucrose non-Fermentable (SWI/SNF) complex these two proteins play a role in gene silencing and activation by remodeling the structure of chromatin [96].

Most publications so far focused on the interaction between BRDs and acetylated lysines in histones and with that, their functions in gene enhancing and silencing. However, proteomics data suggests that lysine acetylation is found in a variety of proteins of the cell [97]. BRDs have shown to bind targets beyond histones like p53 or MyoD [94, 98], but more studies are needed to uncover the physiological functions of these interactions.

1.2.3.1 Subfamily II: Bromo and extra terminal (BET) proteins

Bromo and extra terminal domain (BET) proteins are part of the subfamily II of BRDs. The four mammalian members of this family are BRD2, BRD3, BRD4 and BRDT [99]. Whereas the expression of the *BRDT* gene is limited to testis, BRD2, BRD3 and BRD4 are universally present in the nucleus of different cell types [100]. The members share three structural features: two tandem N-terminal Brds with high sequence homology, an extra-terminal (ET) domain and a C-terminal recruitment domain with a less conserved sequence [68]. The C-terminal recruitment domain is only found in BRD4 and BRDT where it enables interaction with the positive transcription elongation factor b (P-TEFb), which in turn phosphorylates serine residues on the RNA polymerase II, thereby promoting transcriptional elongation [101, 102]. Additional protein-protein interactions are mediated through the ET-domain, which interacts with effector proteins forming multiprotein complexes. These complexes perform several regulatory functions of transcription in addition to P-TEFb transcriptional activation [103]. The single or the tandem Brds of each of the four proteins have been shown to bind to acetylated H4 peptides with high affinity. Binding of the N-terminal Brd to acetylated H4 is stronger when compared to the C-terminal Brd. However, the C-terminal Brd also binds acetylated H3. This selectivity suggests two potential operation modes of BET proteins: either they bind to an H3 and an H4 tail on the same nucleosome, or they are connecting two adjacent nucleosomes by binding to the H3 tail of one and the H4 tail of the other [68].

1.2.4 N-terminal protein acetylation

Nt-acetylation of proteins is facilitated by NATs utilizing acetyl-CoA as a cofactor. Currently, eight eukaryotic NAT complexes are known of which seven are found in mammals: NatA, NatB, NatC, NatD, NatE, NatF and NatH, all displaying a distinct substrate specificity [104]. In this work, the abbreviation NAT is used for describing catalytic subunits (NAA10-NAA60 + NAA80) of the NAT complexes NatA-NatF and NatH which are comprised of catalytic and auxiliary subunits. Auxiliary subunits found in NAT complexes convey specificity but do not have catalytic properties.
All NATs are part of the GNAT superfamily, members of which can be found across prokaryotes and eukaryotes, including the above-mentioned human MYST subfamily of KATs [36, 105]. NATs are composed of at least one catalytic subunit and up to two auxiliary subunits, forming mono- or multiunit enzymes. The auxiliary subunits determine the activity and substrate specificity of the catalytic subunit [106].

NATs can exhibit a very broad substrate specificity like for NatA, which acetylates about 40 % of the human proteome, or a very specific set of substrates like NatH, acetylating only β- and γ-actin [107–109]. The first two to three residues of a protein determine if the N-terminus will be acetylated and which NAT complex will catalyze this modification. The N-terminal methionine (iMet) of proteins followed by a relatively small second amino acid (G, A, V, S, T, C, or P) is often cleaved off by methionine aminopeptidases [110]. Following that, the proteins are acetylated according to the new N-terminus by NATs like NatA or NatD. Proteins that retain their iMet can be modified by different NATs like NatB, NatC, NatF and NatE (NAT complex selectivity is summarized in fig 1.5) [107, 111].

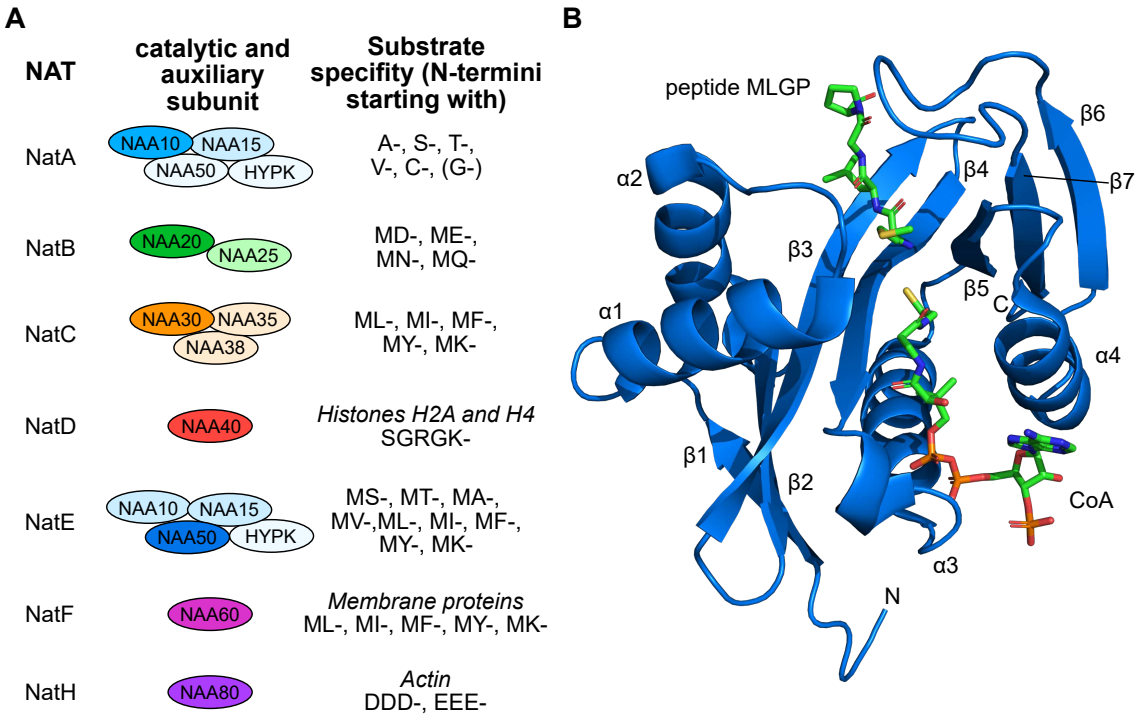


Figure 1.5: The seven mammalian NAT complexes and general composition of catalytic subunits. A) Overview of the composition and substrate specificity of the seven mammalian NATs. The catalytic subunit of each complex is indicated with the brightest color, auxiliary units are indicated with fainter colors. B) General structure of the catalytic subunit of NAT complexes, exemplified in the crystal structure of human Naa50p in complex with CoA and the peptide MLGP. Structural elements as well as both substrates are labeled, adapted from [112]. PDB code: 3TFY

The seven mammalian NATs can be divided into two groups: the five ribosome-associated NATs NatA, NatB, NatC, NatD and NatE which act co-translationally and the two more

recently discovered NATs NatF and NatH that acetylate proteins in a post-translational manner [108, 109, 113]. NatG is a plant-specific NAT complex that contains NAA70 as the catalytic subunit [114]. The ribosomal NATs contain a specific ribosomal anchor unit that interacts with the ribosomal expansion segments, thereby placing the NAT in close proximity to the nascent polypeptide chain [115]. NatD, in contrast to the other four ribosomal NATs, does not contain a ribosomal anchor subunit [116].

As mentioned before, NatA is responsible for the acetylation of the majority of proteins, covering 40 % of all Nt-acetylated proteins in humans [107]. It is composed of the catalytic subunit NAA10 and the auxiliary subunit NAA15, which defines the substrate specificity and associates with the ribosome, effectively anchoring the NAT complex into place [106, 117]. Two additional subunits of this complex are the Huntingtin interactor and chaperone-like protein (HYPK), which has been found to be important for correct Nt-acetylation through NatA [118], and NAA50, which is part of this complex but also exhibits catalytic activity as a subunit of the NatE complex [119]. NatA recognizes and acetylates protein N-termini starting with A, S, T, V, C and G [108]. The NatB complex consists of two subunits: the catalytic NAA20 and the auxiliary NAA25 [120]. Its substrate specificity focuses on iMet-retaining proteins bearing Asx and Glx residues at the second position, namely MD, ME, MN and MQ [107]. NatC is composed similar to NatA and NatB, with NAA30 as catalytic subunit and NAA35 as auxiliary subunit. The function of the second auxiliary unit, NAA38 remains elusive [34, 121]. The substrates of NatC are protein N-termini starting with ML, MI, MF, MY and MK [108]. In contrast to the other ribosome-associated NATs, NatD contains no auxiliary subunit functioning as ribosomal anchor [116]. NAA40 contains no further subunits and possesses narrow substrate specificity, acetylating only the histones H2A and H4 after iMet cleavage [122]. The catalytic subunit of the last ribosomal Nt-acetyltransferase NatE, NAA50, associates with NAA10, NAA15 and HYPK and acetylates N-termini of nascent proteins starting with MS, MT, MA, MV, ML, MI, MF, MY and MK. In this enzymatic complex, HYPK and NAA50 inhibit NAA10 by altering its substrate binding pocket, thereby shaping the specificity of NatE [123]. NatF constitutes only of the NAA60 acetyltransferase and acetylates cytosolic exposed N-termini of transmembrane proteins starting with the amino acid pairs ML, MI, MF, MY and MK. It contains a C-terminal sequence not present in other NATs that associates with the Golgi membrane, suggesting post-translational activity [113]. The newest member of the NAT family is NatH, which consists of the single unit NAA80 [109]. The enzyme is localized solely in the cytosol, where it was shown to acetylate β- and γ-actin after iMet removal [104, 109, 124]. All NATs combined recognize and acetylate a great fraction

of the proteome. Especially for the NATs that acetylate thousands of proteins, it can be difficult to pinpoint cellular phenotypes to a singular NAT. However, overall effects of Nt-acetylation on protein stability, protein-protein interactions and their pathological alterations, amongst others, have been studied.

PPIs can be modulated by Nt-acetylation, inherently changing the polarity of the N-terminus and increasing the affinity between binding partners, as shown for the binding between E2 enzyme UBC12 and E3 ligase DCN1 [35]. Another interaction influenced by Nt-acetylation is the binding of membranes by α-synuclein (α-Syn), a protein involved in Parkinson's disease. Its N-terminal α-helix is stabilized by Nt-acetylation, leading to stronger binding of lipid membranes in *in vitro* studies [125]. Nt-acetylation has also been found to effect the aggregation rate of α-Syn, which is a hallmark for Parkinson's disease [126]. Unacetylated α-Syn was shown to have a much higher rate of aggregation *in vitro* than the acetylated form, which is prevalent *in vivo* [127].

Apart from the aggregation rate, it was assumed that Nt-acetylation generally leads to increased protein half-life *in vivo* [128] until recently, when the Ac/N-end rule was discovered [129]. The Ac/N-end rule is part of the N-end rule, which was proposed in 1986 by Varshavsky and co-workers [130]. The N-end rule states that the half-life of a protein *in vivo* depends on the nature of its N-terminal amino acid, leading to faster or slower protein degradation through the ubiquitin system in eukaryotes. The Ac/N-end rule takes Nt-acetylated proteins into account, which compose the majority of the proteome in a cell [131]. For example, acetylated cytosolic proteins starting with M, S, A, T or V appear to function as Ac/N-degrons, amino acid sequences that are recognized by the ubiquitin system, leading to faster degradation of the protein [132].

Correct functionality of Nt-acetylation also seems to be important for cell homeostasis, as shown for NAD^+-levels in yeast. Here, NatB-deficient yeast exhibited cell-wide decreased NAD^+ concentrations in comparison to wild-type cells [133]. NAD^+-deficiency was likely caused by low levels of the NatB substrates nicotinamide mononucleotide adenylyltransferases Nma1 and Nma2, showing that their Nt-acetylation state is important for NAD^+-homeostasis.

Overall, many cellular functions are influenced by Nt-acetylation of proteins. It appears to be not only crucial for cell homeostasis but also for protein half-life and PPIs. Consequently, malfunctions of NATs can have a huge impact on the organism and play a role in different diseases and maldevelopments like neurodegenerative diseases, development abnormalities and cancer.

1.3 Disease relevance of protein acetylation

Aberrant Nt- and lysine acetylation patterns can lead to severe diseases and malfunctions. Abnormalities in Nt-acetylation patterns can be caused by mutations in the genes coding for catalytic and auxiliary subunits of the NAT complexes. Depending on the location of the mutated amino acid, the catalytic activity can be diminished or lost [134], correct folding can be impaired which can potentially influence catalytic activity [135] or the mutation can be localized in domains responsible for PPIs, disrupting complex formation of multi-subunit NATs [104, 134]. In the case of NAA10, these mutations can lead to cardiac and skeletal abnormalities, growth failure, but most often developmental delay and intellectual disability [104]. Effects of mutations in genes coding for auxiliary subunits overlap with the effects of mutations in catalytic subunits they form complexes with. This was shown for NAA15: 38 patients with a total of 25 different NAA15 variants exhibited neurological effects like intellectual disability, autism spectrum disorder and impaired motor functions, amongst others [136]. Together with the clinical manifestation of NAA10 mutations, these results clearly suggest a vital role of NatA in brain development. Overexpression of NAT complexes is linked to cancer cell proliferation and cancer cell survival as shown for NatB [137], NAA30 as a subunit of NatC [138] as well as NatD [139]. However, the effect of aberrant NAA10 on cancer progression is not yet clear. On the one hand, elevated expression levels of NAA10 have been found in different kinds of cancer, sometimes linked to increased aggressiveness and low survival rate [140, 141]. On the other hand, overexpressed NAA10 corresponded with lower recurrence rates in patients [142], it was found necessary for the induction of apoptosis [143] and it reduced cell growth in breast cancer [144, 145]. With its broad substrate pool that includes protein N-termini, lysines [141] as well as non-catalytic targets [146], it seems that NAA10 can act both as an oncoprotein and a tumor suppressor. This apparent dichotomy will remain a challenge for future studies.

While abnormal Nt-acetylation can effect different developmental pathways and cell proliferation, lysine acetylation was found to play a major role in tumor biology, which is not surprising regarding its function in the context of histone proteins and gene regulation [36]. Generally, abnormal expression levels as well as mutations on the protein level have been found for Brd-containing proteins in a variety of carcinomas [70]. Additionally, translocation events involving genes encoding for BRDs result in fusion genes which translate into new proteins that are linked to very aggressive cancer forms. One notable example are the BET family members *BRD3* and *BRD4*, which can form fusion products

with the *nuclear protein in testis* (*NUT*) gene, a gene that is almost exclusively expressed in testes and ovaries [147]. In particular, the *NUT* gene is translocated on chromosome 15q14, with BRD4 on the short arm of chromosome 19 (t(15;19)(q14;p13.1)) or with BRD3 on the long arm of chromosome 9 (t(15;9)(q14;q34.2)), leading to BRD4-NUT or BRD3-NUT fusion proteins, respectively [148]. These fusion products cytogenetically characterize the so-called NUT-midline carcinoma (NMC), a type of cancer occurring in epithelial tissues along the midline of the body [149]. NMC patients show a median overall survival of 6.7 months, making it one of the most aggressive solid tumors known [150]. An effect of the BRD4-NUT fusion protein observed was the recruitment of proteins like p300 that hyperacetylate certain regions of the chromatin [151], resulting in elevated transcription levels of oncogenes like *MYC*, *TP63*, *SOX2* and *MYB* [152, 153]. It is thought that these growth factors keep the cancer cells in a proliferative and undifferentiated stem cell-like state, contributing to the poor prognosis of this disease [152].

1.4 Chemical and biochemical tools

Studying the interaction between an enzyme and its substrate is a challenging task throughout all fields of biochemistry. Elucidation of binding affinities and -mechanisms not only helps to uncover physiological pathways, it is also crucial for the development of inhibitors of proteins involved in diseases like cancer. Inhibitors are molecules that can bind to enzymes, thereby blocking their catalytic activity. They can bind directly to the catalytic site of the enzyme or modulate the enzyme in an allosteric fashion. Ligands can bind to interaction surfaces for PPIs and thus regulate these interactions. Especially more recent approaches tried to address PPIs to modulate protein function and activity [154]. A specific form of PPIs is regulated via PTMs, mainly via the interplay of the modified amino acid with certain binding modules [155]. Here, binding takes place in special binding pockets confined to a much smaller area than other PPIs and are thus better addressable with ligands.

Many drugs are enzymatic inhibitors used in a disease context to impair aberrant enzyme activity. Pharmaceutical drugs in general can be divided into two classes, depending on their molecular weight: small molecules with a molecular weight of <500 Da and much larger biomolecules also referred to as biologics with molecular weights above 5000 Da, each class possessing specific advantages and disadvantages [156]. This leaves a major gap of molecules with a weight between 500 Da and 5000 Da that could potentially be filled by therapeutic peptides. An important tool in peptide drug discovery are display techniques

like phage- and mRNA-display, that in theory can produce libraries of up to 10^{15} unique peptides which can be tested for their binding of a particular protein target [157].
Another challenge in the broad field of enzyme-substrate relationships is the determination of the substrate specificity of an enzyme or an enzymatic complex. This can be addressed by combining immobilized specific inhibitors and chemoproteomic profiling techniques utilizing mass spectrometry, yielding information on specificity as well as complex composition of potential binding partners.

1.4.1 Small molecule inhibitors

With a market share of 75 %, small molecules make up the vast majority of the global marketed drugs [157]. Regarding diseases related to protein acetylation, KDACs/HDACs have been the main target of drug research. Consequently, several small molecule HDAC inhibitors (HDIs) have been discovered over the years, with Trichostatin A identified in 1990 [158] being the first one including a hydroxamate moiety 1.6. The hydroxamate binds to the Zn^{2+}-ion of the active site of the HDAC while the aliphatic linker reaches into the binding pocket. Additional interaction with the enzyme are mediated through a cap part, thereby exhibiting a strong affinity to Zn^{2+}-dependent HDACs [159]. The first FDA approved hydroxamate-based HDI was suberoylanilide hydroxamic acid (SAHA) in 2006 for the treatment of cutaneous T cell lymphoma, where it alters gene expression and promotes growth arrest and apoptosis in tumor cells [160] 1.6. Up until 2020, a total of five HDIs have been approved as cancer therapeutics [159]. HDIs are investigated for further applications in treatment of Alzeimer's and Huntington's disease [161], depression [162] as well as inflammatory diseases [163].
Small molecule inhibitors have also been developed for several BRDs, with the BET-family being in the spotlight of drug research. They do not act as an inhibitor in a classical understanding but serve as ligands that tightly bind to the KAc binding pocket. Selective inhibitors for members of the BRD families III (SGC-CBP30 for CBP/p300) [164], IV (BI-7273 and BI-9564 for BRD9) [165], V (BAZ2-ICR for BAZ2a/BAZ2B) [166] and VIII (PFI-3 for SMARCA2/4 and PB1(5)) [167] have been described so far and mostly used to investigate functions of these BRDs *in vitro* and *in vivo*. Inhibitors of the BRD family II have progressed to potential application as drugs with 13 BET inhibitors in early clinical trials up to 2017 [168]. The loop regions of Brds exhibit sequence variation, harbouring the potential for highly selective inhibitors. This stands in contrast to the alpha helix and beta fold regions of the Brds belonging to BET proteins, which are highly conserved.

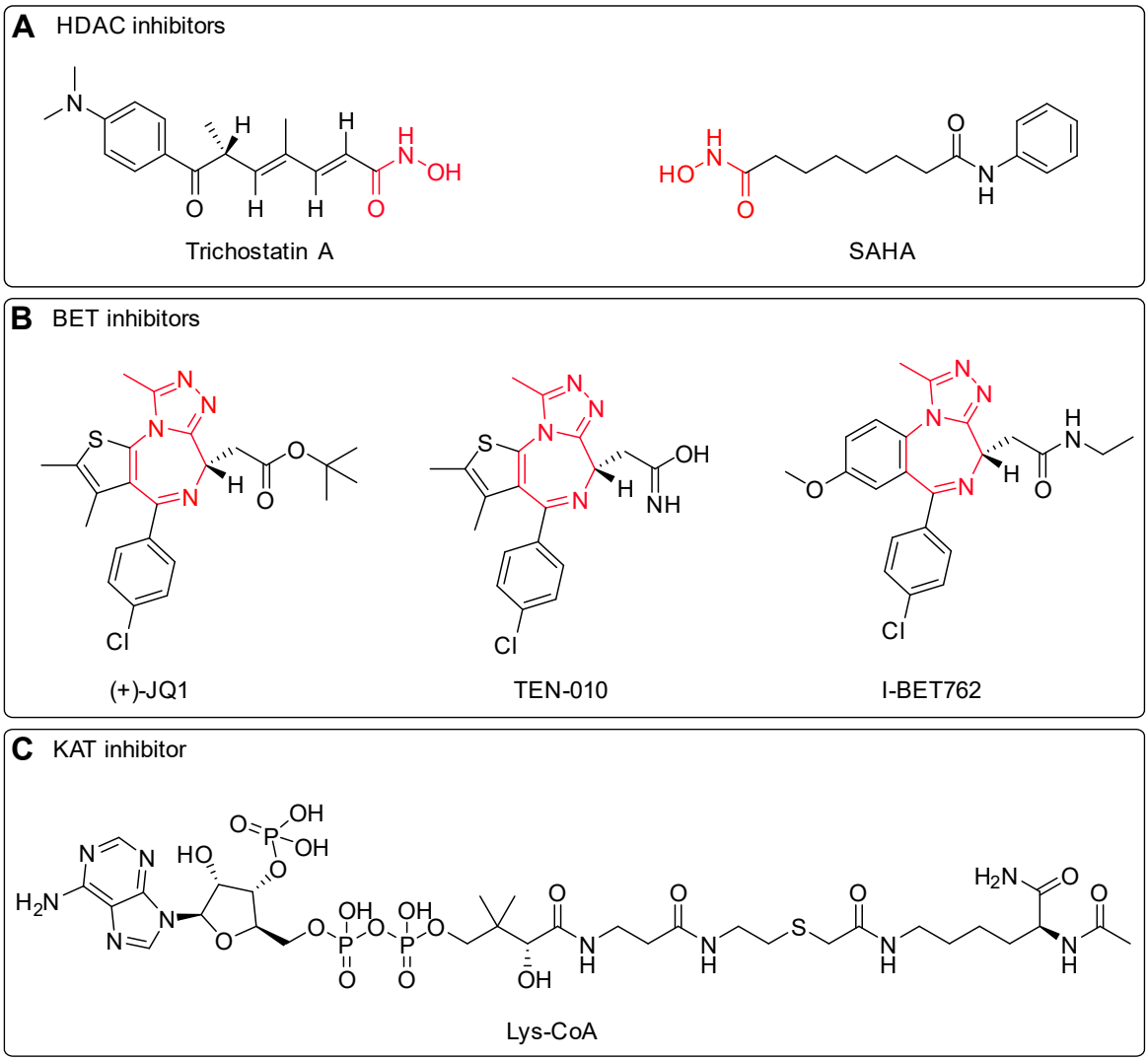


Figure 1.6: Chemical structures of a selection of HDAC, BET and KAT inhibitors. A) Two HDIs including a hydroxamate moiety indicated in red. B) The three BET inhibitors JQ1, TEN-010 and I-BET762 with the uniting structural motif of a methyl-triazolo-diazepine marked in red. C) Chemical structure of Lys-CoA, which is recognized and bound by KATs.

The first two BET inhibitors (+)-JQ1 (hereafter referred to as JQ1) and I-BET762 have been reported in 2010 [169, 170] 1.6. JQ1 binds to BRD4(1) and BRD4(2) with K_Ds of 50 nM and 90 nM, respectively, and comparable binding was observed for the two Brds of BRD3 [169]. Crystal structures revealed a crucial hydrogen bond between the triazole moiety in JQ1 and the conserved asparagine within the Brd resembling the hydrogen bond between KAc and asparagine found in the native binding context. The

overall interaction was stabilized via hydrophobic contacts with conserved BET residues in the ZA- and BC-loops [169]. While BET inhibitors generally discriminate between the different BRD subfamilies, the selectivity within the BET family is low. So far, no BET inhibitor has been reported that distinguishes between BRD2, BRD3, BRD4 and BRDT [171]. Promising pre-clinical investigations with JQ1 were performed in NMC cell lines where BRD4-NUT was effectively hindered from binding chromatin, which was followed by differentiation and apoptosis in NUT carcinoma cells [168]. However, short half-life and a poor pharmacokinetic profile limit the translation of JQ1 into a clinical success [172]. Adding to this, clinical trials of several other BET inhibitors revealed weak responses in patients with lymphoma, multiple myeloma and solid tumors [173] while patients with NMC treated with the JQ1-related TEN-010 showed fast initial response but experienced relapse within 2–8 months [174]. While the results of the clinical trials of the monotherapies have not managed to measure up to expectations, combination therapy appears promising. In cancer cell lines and xenograft models, BET inhibitors exhibited synergistic effects if administered with kinase inhibitors or HDIs, even reducing the required dose and thus lowering toxicity [175, 176]. However, as of April 2022, no BRD inhibitor has received FDA approval [177]. KATs as enhancers of various oncogenes coding for transcriptional factors and kinases also pose as promising drug targets [178]. The first reported KAT inhibitor Lys-CoA acts as a bisubstrate inhibitor by addressing the CoA- as well as the lysine binding pocket simultaneously [179] 1.6. NATs not only utilize the same co-substrate as KATs, they are also part of the GNAT superfamily. Thus, it is not surprising that the bisubstrate principle of Lys-CoA was applied to peptides, conjugating the N-terminal amino acid with CoA and thereby creating a selective inhibitor for recombinant and endogenous NATs [180].

1.4.2 Peptide-based inhibitors

Since the development of insulin in 1922, 80 peptide drugs have been approved for medical use, with 150 more in clinical trials [157]. Important discoveries that paved the way to the commercial application of peptides in medicine were the invention of the solid-phase peptide synthesis (SPPS) by Merrifield in 1963 [181] as well as the development and the refining of high performance liquid chromatography (HPLC) starting in the 1960s and the advent of recombinant expression systems in the 1970s. Peptide-based inhibitors are a sub-class of peptide drugs that can inhibit enzymatic activity or PPIs by competing with substrates and contact surfaces of proteins. In contrast to small molecules, most

peptides suffer from low oral bioavailability, short half-life, rapid clearance, poor membrane permeability and potential immunogenicity [182]. On the other hand, they are remarkably potent and selective, possess low toxicity, their synthetic protocols are standardized and they can address PPIs that are otherwise considered "undruggable" by most small molecule drugs [182, 183]. To overcome these limitations, many optimization steps are required until a candidate inhibitor can reach pre-clinical trials, a general discovery strategy is given in 1.7.

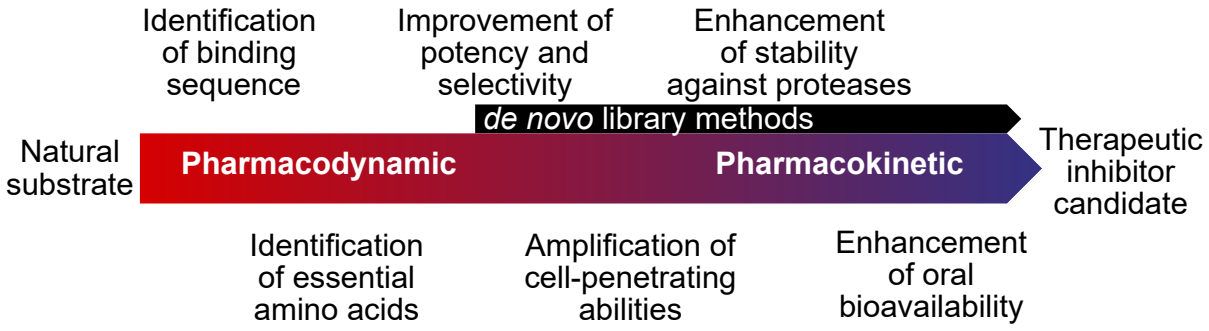


Figure 1.7: General discovery strategy for peptide-based inhibitors of a protein starting from a natural substrate. Starting from natural substrates, the binding towards the cellular target is optimized which is followed by optimization of pharmacokinetic properties like *in vivo* stability and oral availability leading to a therapeutic candidate. *De novo* methods do not require a natural substrate to start with.

A natural substrate peptide is broken down into smaller pieces which are tested for binding to the target protein in order to identify the minimum active sequence. At smaller length cell-permeability of peptides is often improved [184], but renal clearance has to be kept in mind [185]. Site-directed mutagenesis in the form of alanine scans is then performed, where every amino acid is substituted with alanine one after the other and binding against the target is again tested, which enables the identification of amino acids essential for binding as well as redundant sites for possible substitutions [186]. By substituting non-essential amino acids with a different amino acid in the next step, the potency and selectivity of the peptide can be improved, a task that might be accomplished by display methods. In display methods like phage display and mRNA display, a large library of peptides is generated and selected by target binding [187]. Peptide libraries can also be used as a *de novo* tool [188], which can be assisted by computational methods [189]. Synthetic peptides are commonly not cell-permeable per se. The insertion of cell-penetrating peptide (CPP) elements like penetratin or transcription trans-activating (TAT) sequences can overcome this limitation [190]. Other drawbacks of peptides mentioned earlier are a short half-life

and poor oral bioavailability. Peptides are inherently prone to degradation by proteases, but identification and substitution of proteolytic sites [191], incorporation of N-methylated, D- and non-natural amino acids [184] as well as cyclization [192] can extend peptic half-life from minutes to hours or even days [157], as the native targets for proteases are peptides formed from L-amino acids. Susceptibility to renal clearance can be improved by lipidation, conjugation to albumins or globulins and PEGylation [157]. Stability enhancing methods like cyclization and addition of non-natural amino acids have a similar effect on oral availability, circumventing or slowing the enzymatic metabolization of peptides in the gastrointestinal tract [157, 193].
As mentioned above, peptide-based inhibitors have been developed for NATs [180] and used in *in vitro* studies to selectively enrich NATs and interacting proteins [194]. For the clinically significant BET proteins, a peptide-based inhibitor has been reported recently [195] but its usage of acetyllysine restricts *in vivo* applications due to endogenous deacetylation. The recent discovery of a stable KAc mimic, a 1,2,4-triazole amino acid inspired by JQ1 [196], gives hope to use the advantages of peptide inhibitors to address BRDs.

1.4.3 Mixture based positional scanning synthetic peptide combinatorial libraries (PS-SPCL)

Peptide drug discovery is often accomplished by screening chemical or biological combinatorial libraries. Two different screening strategies can be distinguished: the phenotypic screening monitoring the effects compounds induce in live cells, tissues or organisms, leading to the identification of previously unknown targets. The target-based approach aims at new binders of known molecular targets [197, 198]. The libraries used for screening can be of biological or synthetic origin, each having different advantages and disadvantages including practical manners such as available equipment [199]. Biological libraries can be based on cellular and acellular display systems [200], with yeast display [201] as an example for a widely used cellular approach and mRNA display as an example for acellular methods [202]. The cell-free mRNA display possesses some major advantages over cellular approaches like yeast display, including the potential use of non-natural amino acids, larger library sizes up to 10^{15} individual peptides as well as the fact that no transformation of host cells is required [203, 204].
Chemical libraries are either synthesized and screened on a solid support or the peptides are cleaved off the solid support after synthesis and screened afterward. For the creation of large chemical libraries, the ”split-and-mix” method can be used, in which the first

amino acid of the peptide is coupled to resin beads in individual reaction vials. The beads are then mixed and equal portions are distributed amongst the reactors in which the next amino acid is coupled [205]. In each reactor, a different amino acid is coupled resulting in a library of peptides equaling X^n with X being the number of different amino acids, e.g. 20 for all canonical amino acids, and n being the length of the peptide. Due to the underlying concept of one-bead-one-compound, this approach is limited by the number of beads that have to exceed or be equal to the number of theoretical products [206]. The second strategy is the "pre-mix" or mixture-based method in which all amino acids are coupled at each step (X) in equimolar ratio or in so-called isokinetic mixtures that theoretically ensure an equimolar distribution of amino acids at a defined position in the peptides [207]. Isokinetic mixtures allow for the usage of matrices like modified planar cellulose which unlike bead-based polymer resins cannot be split, a prerequisite for the "split-and-mix" method.

To deconvolute chemical peptide libraries, iterative deconvolution or positional scanning (PS)/additive deconvolution can be used (the two strategies are shown in fig 2.8). In the iterative deconvolution, the library is divided into subsets, each including peptides with a defined amino acid at a specific position, while the other positions are randomized. After identifying the subset with the best performance, the defined position of this subset is retained and the subset is divided into further subsets. This process continues until the most active peptide is determined [208]. Using the additive deconvolution, again a defined amino acid with randomized residual positions is used as one subset with the difference that no result is retained for the next set and all subsets can be screened in parallel. The best performing amino acids of each subset make up the final peptide displaying superior binding. The advantage of this method is enabling parallel screening, which greatly reduces the overall screening time [209]. Other hit identification methods include mass spectrometry (MS)-based methods and fluorescence-based methods [210].

With the mixture-based approach applied to chemical peptide libraries, different kinds of matrices can be employed to screen for interacting peptides. One method is the SPOT synthesis, in which peptides are synthesized in spots on a functionalized cellulose membrane employing the Fmoc-strategy of SPPS [211]. Liquid handling robots facilitate the synthesis, which includes addition of the activated amino acid to the matrix, washing steps and Fmoc-removal until the desired peptide length is achieved. The combination with the mixture-based approach results in spots with a millimeter diameter containing several thousand peptides [212]. Binding can be tested with fluorescence-tagged proteins that enrich on spots with peptides exhibiting tight binding.

1.4.4 Proteomics

Proteomics is the quantitative determination of protein levels and its alterations under perturbations including drug treatments [213]. While different strategies like fluorescence- or radioactivity-based detection exist, mass spectrometry coupled with liquid chromatography is now the *de facto* standard method of quantifying proteins as they provide high resolution separation coupled with means of protein identification and quantification [214]. However, mass spectrometry is not intrinsically quantitative as the ionization yield of analytes is affected by the size, charge and hydrophobicity of the peptides. To address this problem, proteins can be labeled with stable isotopes metabolically, chemically or enzymatically and used as external or internal standards [215]. Thus, relative and even absolute quantification can be conducted by comparing the signal intensities of the labeled and the unlabeled peptides by mass spectrometry. Label-free quantification (LFQ) methods exist as well and can be divided into methods that use the signal intensity of any given peptide for relative quantification or methods which take the total number of peptide fragments identified for each protein into account. This method is based on the empirical observation that higher amounts of a protein correlate with more MS spectra of peptides originating from that protein [215]. LFQ methods do not require costly labeling agents, are generally less time-consuming and, caused by the fewer steps involved in the workflow, are less prone to technical errors than label-based techniques [216]. In recent years, the development of intensity determination and normalization procedures like MaxQuant have improved the relative quantification in MS-based LFQ [217, 218].

Chemoproteomics is a sub-class of proteomics which aims to unravel interactions between small-molecules and the whole proteome of any given cell or tissue [219]. One approach in this field is the activity-based protein profiling (ABPP) [220]. The probe molecules used in this approach consist of three distinct structural elements: the so-called warhead that binds to the active site of an enzyme, a linker group for fine-tuning the selectivity of the probe which simultaneously acts as a spacer between the warhead and the reporter tag which can be a fluorophore or an affinity tag for enriching targeted enzymes from celllysates. A general workflow of ABPP coupled with LC-MS readout and LFQ quantification is illustrated in 1.8.

An immobilized ABPP-probe is incubated with a cellular lysate followed by washing to remove non-binders. The bound proteins and complexes are eluted and digested using proteases like trypsin or lysozymes, which leaves predictable peptide fragments, which are easily detected by high resolution LC-MS/MS [221]. The tryptic peptides are then

desalted, separated by nano-HPLC, identified by tandem mass spectrometry (MS/MS) and ultimately quantified using data processing methods like MaxQuant coupled with protein databases and LFQ [222].

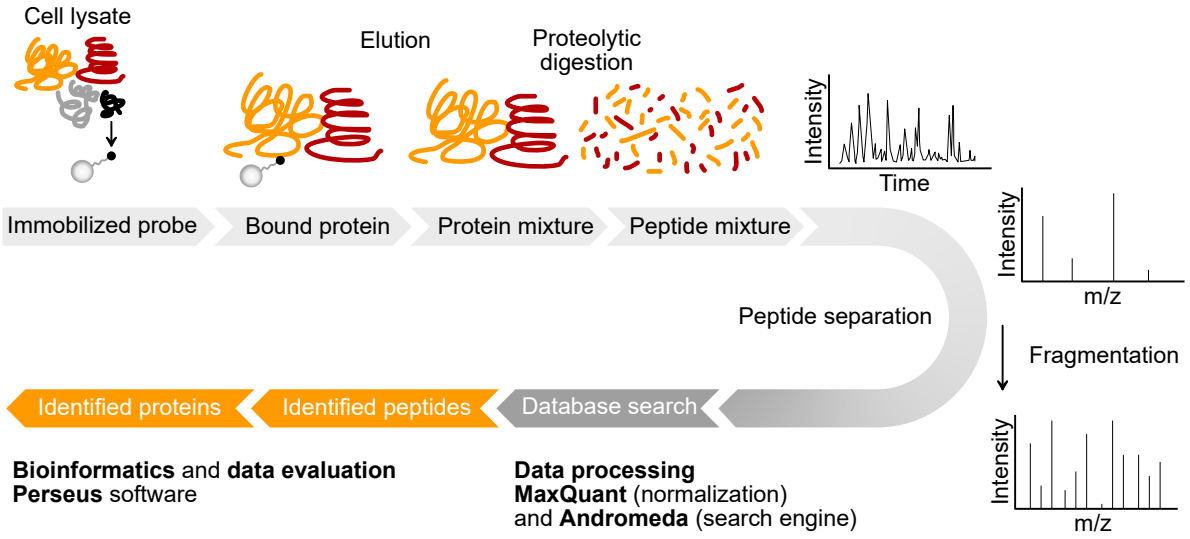


Figure 1.8: General principle of an ABPP-based proteomics workflow. Proteins bound to the probe are eluated and digested using proteases. Resulting peptide fragments are separated as well as analyzed and identified using LC-MS/MS. Corresponding spectra together with database searches lead to the identification of proteins which can be quantified and visualized using MaqQuant and Perseus. Adapted from [222].

1.5 Aims of study

This work can be divided into two parts: the main part focuses on the development of high-affinity peptide inhibitors of BRD3 and BRD4 for *in vivo* applications. The second part focuses on the interactome profiling of NAT complexes.
The aim of the first part of this thesis was to develop peptide inhibitors of the BET proteins BRD3 and BRD4 containing a stable acetyllysine mimic and optimize them in terms of length, binding affinity and protease stability. To do so, a library method for high-throughput screening had to be established and Brd constructs for pull-downs needed to be designed and expressed. As the starting point for the development of the peptide sequence, known native histone substrates of the different BETs [68] should be used. A recently developed, stable acetyllysine mimic [196] should be used as anchoring moiety enabling *in vivo* applications.
Peptide sequences undergoing this optimization procedure should be combined to a bivalent ligand addressing both Brds of BRD3 and BRD4 at the same time.
For the second part of this work, CoA-peptide conjugates with differing N-terminal amino acids should be synthesized and immobilized. These peptides should then be used in pull-down experiments with native cellular lysates. Consequently, label-free quantitative mass spectrometry was to be applied and the results should be compared to elucidate binding specificities and interactome data of NAT complexes. Both parts can be outlined as follows:

1. Development of BET peptide inhibitors
    a) Optimization of histone-derived peptide sequences in terms of length, affinity and stability for different BETs through the establishment of high-throughput library methods.
    b) Design and synthesis of a peptide construct containing the optimized binding sequences as well as means for *in vivo* experiments.
2. Establishing CoA-peptide conjugates for N-acetyltransferases
    a) Synthesis and immobilization of CoA-peptide conjugates.
    b) Pull-down experiments with cell extracts and proteomic analysis.

2 Results

2.1 Part 1: Development of ApmTri containing peptide substrates for BET proteins

JQ1 and other small molecule inhibitors bind with high affinity to the acetyllysine binding pocket of bromodomain 1 (BD1) and bromodomain 2 (BD2) of BRDs belonging to the BET-family [223]. The JQ1 derived non-natural amino acid 2-aminopimelic acid(3-methyl-1,2,4-triazole) (ApmTri) was shown to bind to the same binding pockets with its triazole moiety, exhibiting similar K_D-values as the native KAc substrate [196]. Starting from native substrate sequences, peptide-based inhibitors for BRD3 and BRD4 containing ApmTri were developed and optimized using SPOT synthesis, resulting in molecules suited for *in vivo* applications. The workflow is shown in 2.1.

A starting point for the optimization were known binding sequences of the two Brds of BRD3 and BRD4, respectively. Each of the four Brds bind to different acetylated sequences of the tails of the four core histone proteins with distinct affinities [66]. The optimization process started with the step-wise truncation of the native sequence in order to identify the minimal binding motif. Resulting peptides, termed deletion mutants, as well as the peptide with the original length were then synthesized in triplicates on a cellulose membrane (see a)). After deprotection, the membrane was incubated with the respective recombinant Brd fused to a fluorescent protein. Binding events were detected by the fluorescence intensity retained at each peptide spot. In a second step, all positions in the deletion mutants were replaced with every canonical amino acids creating a library and the resulting peptides were synthesized in triplicates on a membrane, as shown in b). The method employed for the library generation of these so-called affinity mutants is explained in chapter 2.1.5. After incubating the deprotected membranes again with the fluorescent protein, the peptide exhibiting the highest binding was chosen for the next step. In c), each L-amino acid in the optimized sequence, except for the central acetyllysine, was substituted with its D-enantiomer as well as with its N-methylated derivative. Replacing in total three L-amino acids per peptide resulted in so-called stability mutants. Binding experiments with the respective BRD fusion protein led to sequences that showed sufficient binding, but would be more stable against proteases. This sequence was included into the bivalent stable ligand design shown in d), which was designed to penetrate the cell membrane, reach the cell nuclei and inhibit BRD3 and BRD4.

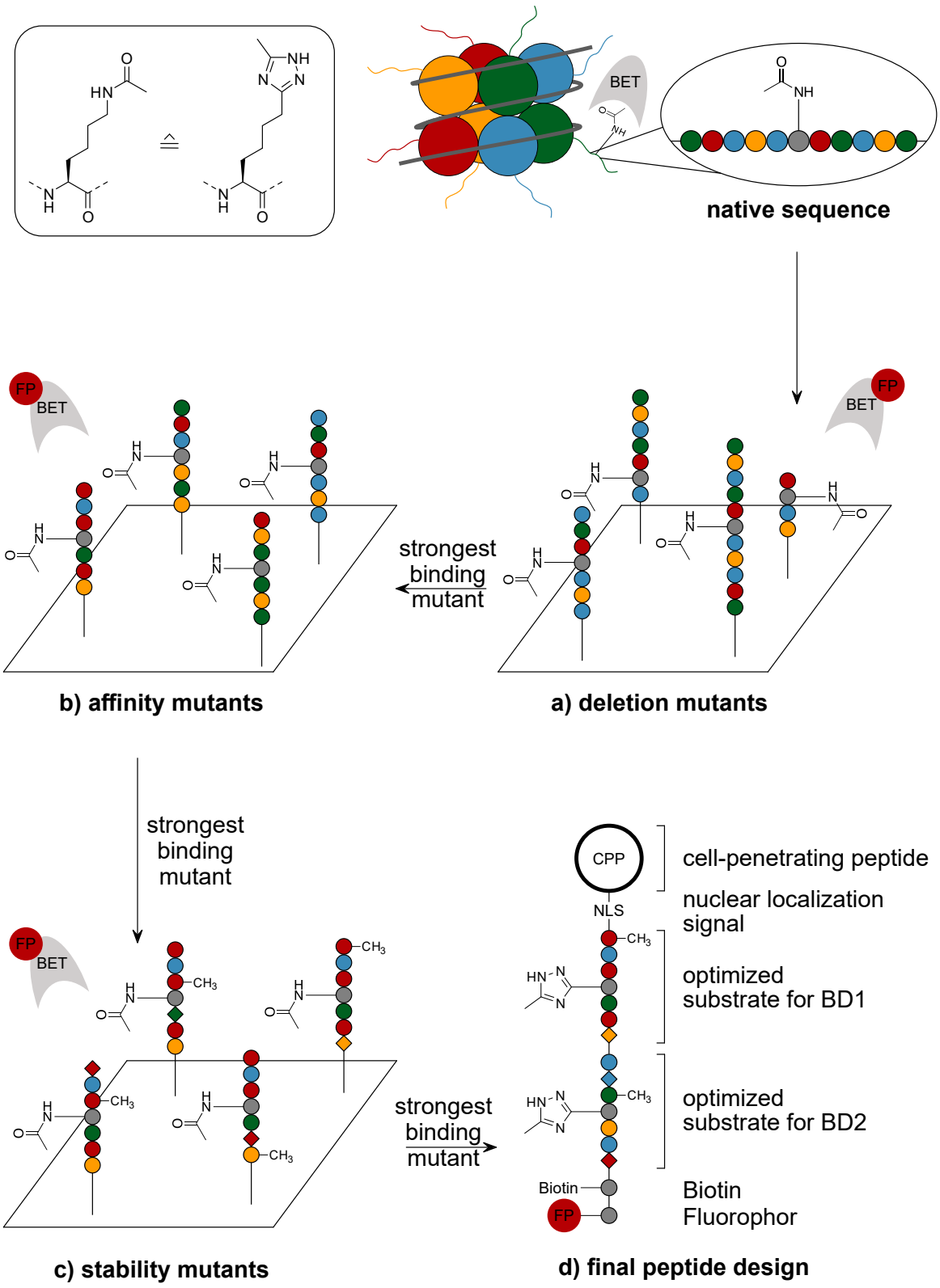

BET
native sequence
FP
BET
FP
BET
strongest
binding
mutant
b) affinity mutants
a) deletion mutants
strongest
binding
mutant
FP
BET
strongest
binding
mutant
CPP
NLS
cell-penetrating peptide
nuclear localization
signal
optimized
substrate for BD1
optimized
substrate for BD2
Biotin
FP
Biotin
Fluorophor
c) stability mutants
d) final peptide design

Figure 2.1: Strategy for the development of peptide-based inhibitors containing ApmTri as acetyllysine mimic. L-amino acids are represented as small colored spheres while the acetyllysine is indicated with an acetamide moiety. D-amino acids are shown as small diamond shapes and N-methylated amino acids are indicated with a methyl group. All four histone proteins that make up the nucleosome core particle are displayed as big colored spheres, while their tails stick out as thin lines. The bromodomain-containing protein is indicated as a gray semi-oval area including a binding pocket, the fluorophore is abbreviated as FP.

While incorporating the optimized sequence for BD1 as well as BD2 with ApmTri instead of KAc, the final design should also contain an N-terminal cell-penetrating peptide (CPP) for cellular uptake, a nuclear localization signal (NLS) for the transport into the nucleus, a fluorophore for fluorescence microscopy and a biotin to enable degradation studies. The optimal linker length between the substrates of BD1 and BD2 was determined in separate experiments.
In contrast to Fmoc-ApmTri-OH, Fmoc-Lys(Ac)-OH is a commercially available building block for automated SPPS. Because of that, most mutants in steps a), b) and c) synthesized on a membrane contained acetylated lysine. The results gained in these experiments were each verified re-synthesizing the best performing sequences with ApmTri and with KAc and conducting pull-down experiments with the respective BRD fusion protein. Additionally, crystal structure of peptides containing KAc compared with the same peptide containing ApmTri exhibited strikingly similar binding modes when bound by a bromodomain [196]. For these reasons, it was assumed that optimization results gained for peptides containing acetyllysine were applicable to peptides containing ApmTri.

2.1.1 Design of cell permeable bivalent peptide ligand for BRD3 and BRD4

The design of the cell permeable ligand is shown in figure 2.2. As reader proteins of post-translational histone modifications, BRD3 and BRD4 are located in the nucleus. To reach this sub-cellular location, the bivalent probes are required to be able to cross the cell- as well as the nuclear membrane. CPPs can be used to facilitate the transport of small molecules [224] and peptides across the cell membrane [225]. NLS on the other hand interact with nuclear transport factors, thereby enabling active import into the nucleus through the nuclear core complex [226]. By attaching an N-terminal CPP to an NLS via a disulfide bond, the peptide can pass through the cell membrane into the cytoplasm, where

the disulfide bond would be reduced, releasing the CPP and enabling the transport into the nucleus through the NLS.

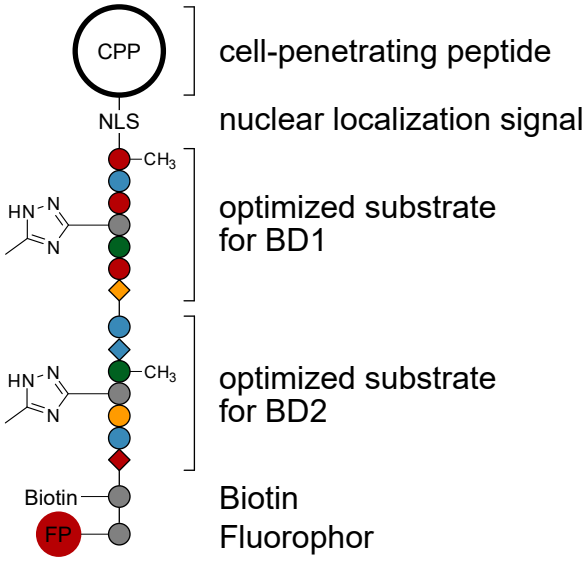


Figure 2.2: Design of the bivalent stable ligand for *in vivo* studies. L-amino acids are displayed as colored circles, D-amino acids are indicated with colored diamond shapes and N-methylated amino acids are indicated with a methyl group. The acetyllysine mimic ApmTri is represented as a gray circle with an attached triazole ring.

The NLS is covalently attached to the optimized and stabilized ligand of BRD3 and BRD4 (see chapter 2.1.5). The construct includes the acetyllysine mimic ApmTri instead of acetylated lysine because it cannot be altered by deacetylases [196]. The following biotin moiety enables immobilization on streptavidin and subsequent MS experiments. A C-terminal fluorophore allows *in vivo* fluorescence imaging of uptake and sub-cellular localization of the probe.

2.1.2 Synthesis of Fmoc-ApmTri-OH

The non-commercial building block Fmoc-ApmTri-OH needed to be synthesized in a first step. The route employed for the synthesis of (**7**) was developed by Dr. Sören Kirchgässner and is illustrated in 2.3.

In a first step, the α-amino group of the racemic 2-aminopimelic acid (**1**) was acetylated under basic conditions and afterward purified by RP-HPLC. The chiral resolution of the racemic mixture was performed with the enzyme acylase I, which only deacetylates

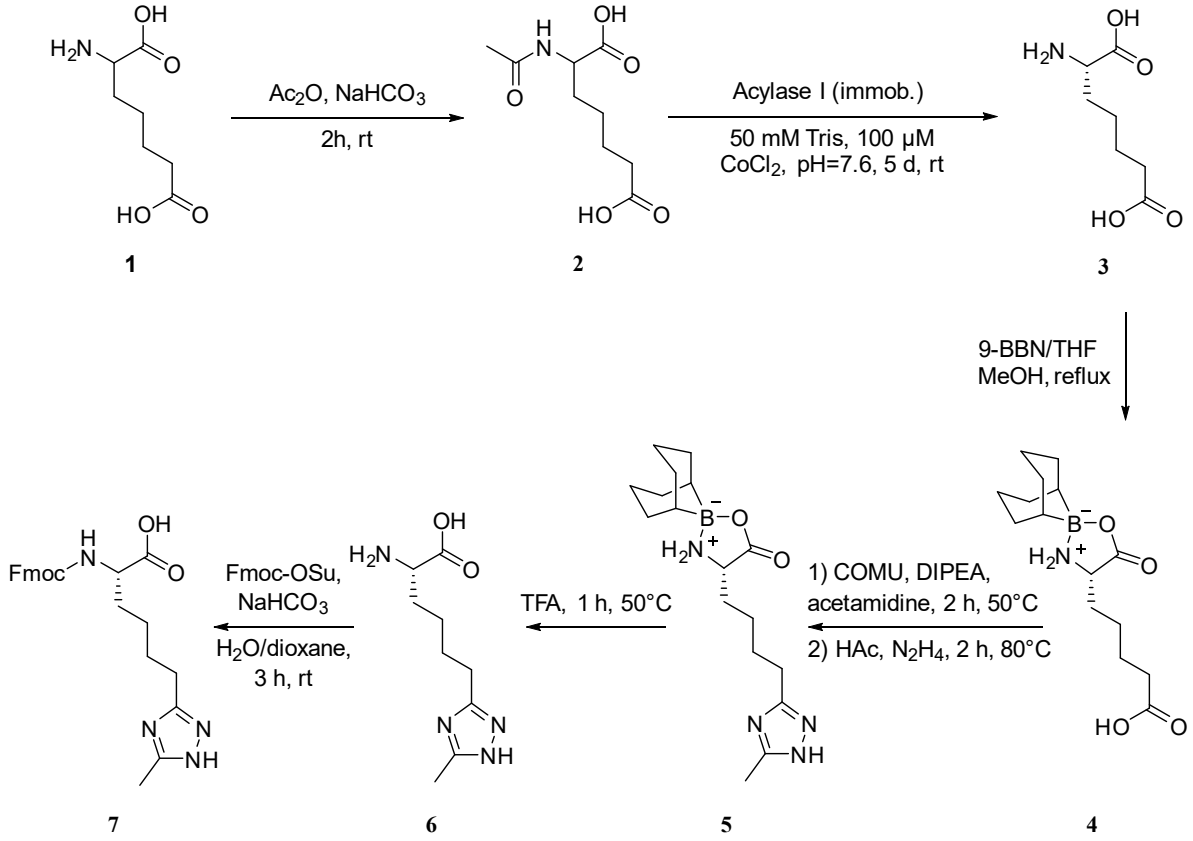


Figure 2.3: Synthesis route for Fmoc-ApmTri-OH (7) starting from racemic 2-aminopimelic acid (1).

the L-enantiomer, leading to a mixture of H-L-Apm-OH (**3**) and Ac-D-Apm-OH (**2a**) (not shown). 9-Borabicyclo(3.3.1)nonane (9-BBN) was used for the protection of the enantiopure H-L-Apm-OH (**3**), leading to 9-BBN-L-Apm-OH (**4**). After purification of (**4**) via preparative HPLC, 9-BBN-L-Apm(Triazole)-OH (**5**) was assembled in a one-pot reaction. In this synthesis, an acyl-amidine intermediate was first formed by reacting the unprotected carboxy group in **4** with acetamidine using the coupling agent COMU. Intramolecular cyclization then followed in a reaction similar to the Einhorn-Brunner reaction [227] by adding acetic acid and hydrazine, resulting in a 1,2,4-triazole ring with a methyl substituent at position 3. The final product was formed by first deprotecting **5** with TFA and then protecting the resulting H-L-Apm(Triazole)-OH (**6**) with Fmoc-OSu under basic conditions. A final RP-HPLC purification step resulted in the pure SPPS compatible product Fmoc-L-Apm(Triazole)-OH (**7**). The overall yield of the synthesis was 4.6 % including the 50 % loss in the chiral resolution. Figure 2.4 shows the LC-MS spectra of the purified final product.

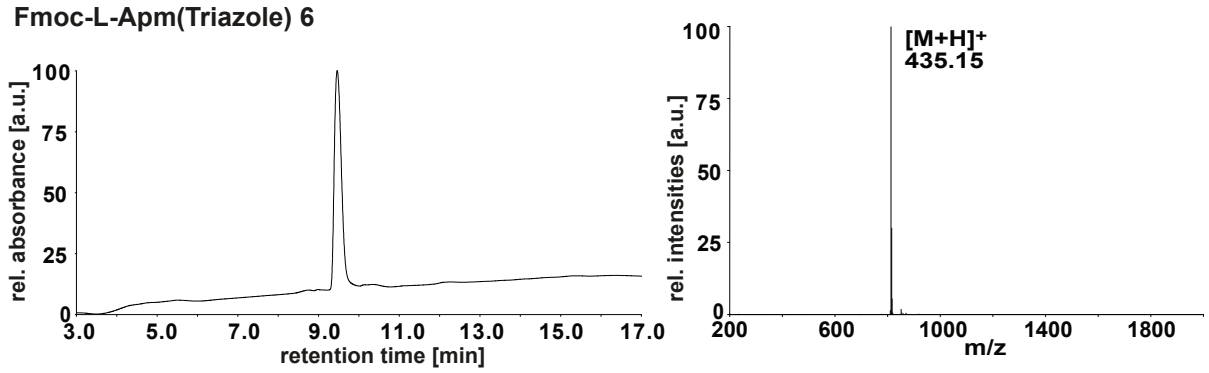


Figure 2.4: LC-MS spectrum of Fmoc-Apm(Triazole)-OH (7). The UV-spectrum on the left was recorded at λ= 218 nm, the MS-spectrum is displayed on the right side.

2.1.3 Design and recombinant expression of the bromodomain proteins BRD3(1), BRD3 and BRD4

Both the single Brds of BRD3 and BRD4 as well as the two single domains together were studied with respect to their binding properties towards different peptides. For this, constructs suitable for pull-down experiments had to be cloned for BRD3(1), BRD3(2), BRD4(1), BRD4(2), BRD3 and BRD4, the latter two bearing both bromodomains as well as the linker in between. The amino acid sequence of the Brds was adapted from [66] and the design of the constructs was developed and BRD3(2), BRD4(1) as well as BRD4(2) were already cloned by Dr. Sören Kirchgässner.

In the constructs, the sequence of the respective bromodomain or bromodomains was fused to a fluorescent protein on the DNA level. Including a fluorophore into each construct allowed relative quantification of the proteins to resin-bound as well as membrane-bound peptide probes. BRD3 and BRD4 were fused to mTagBFP2 [228], BRD3(2) and BRD4(2) were fused to TurboYFP [229], BRD3(1) was fused to mKate2 [230] and BRD4(1) was fused to TagGFP2 [231]. Each construct contained an N-terminal His_6-Tag and a C-terminal Strep-Tag for purification as well as a thrombin cleavage site and a tobacco etch virus (TEV) protease cleavage site between His_6-Tag-BRD and BRD-fluorophor, respectively. The two constructs BRD3-mTagBFP2 and BRD4-mTagBFP2 did not contain a C-terminal Strep-Tag. An alternative purification method is offered by the inclusion of the Strep-Tag and the protease cleavage sites allow for removal of the fluorescent protein and affinity tags. All six constructs are displayed in Figure 2.5, their exact sequence can be found in the appendix 4.1.

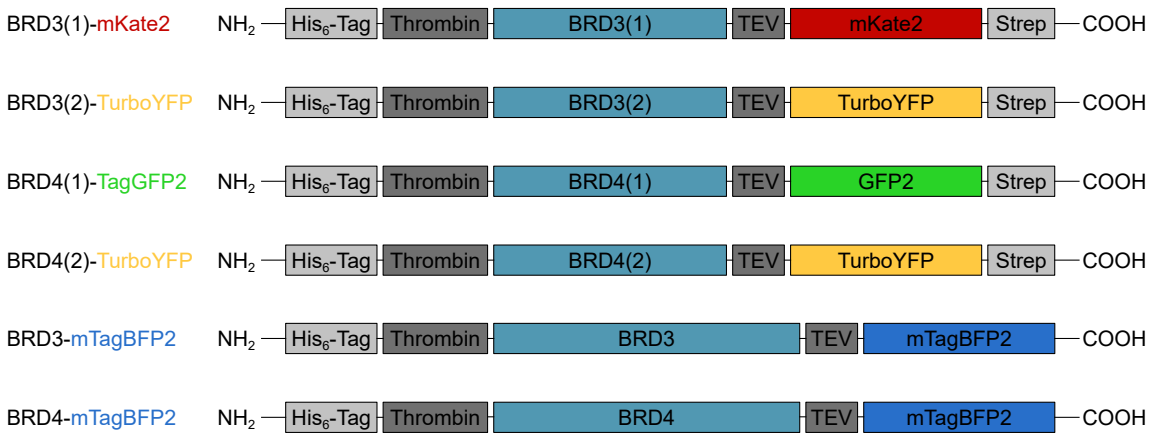


Figure 2.5: Overview of the six protein constructs used in this work. The affinity tags are marked in light gray, the protease cleavage sites in dark gray, the bromodomains are colored in azure and the fluorescence tags are highlighted in the color relating to their emission wavelength.

BRD3(2)-TurboYFP, BRD4(1)-TagGFP2 and BRD4(2)-TurboYFP were expressed and purified by Dr. Sören Kirchgässner, BRD3-mTagBFP2 and BRD4-mTagBFP2 were expressed and purified in cooperation with Yandan Zhao.

Table 2.1: Concentration of purified protein and used storage buffers.

Protein	**Storage buffer**	**Concentration**
BRD3(1)-mKate2	100 mM KCl, 20 mM HEPES, 20 % glycerine, pH 7.9	33.05 mg/mL
BRD3(2)-TurboYFP	100 mM KCl, 20 mM HEPES, 20 % glycerine, pH 7.9	23.92 mg/mL
BRD4(1)-TagGFP2	100 mM KCl, 20 mM HEPES, 20 % glycerine, pH 7.9	33.00 mg/mL
BRD4(2)-TurboYFP	100 mM KCl, 20 mM HEPES, 20 % glycerine, pH 7.9	22.80 mg/mL
BRD3-mTagBFP2	300 mM NaCl, 50 mM Tris-HCl, 10 % glycerine, 2 mM DTT, pH 7.5	2.18 mg/mL
BRD4-mTagBFP2	300 mM NaCl, 50 mM Tris-HCl, 10 % glycerine, 2 mM DTT, pH 7.5	2.19 mg/mL

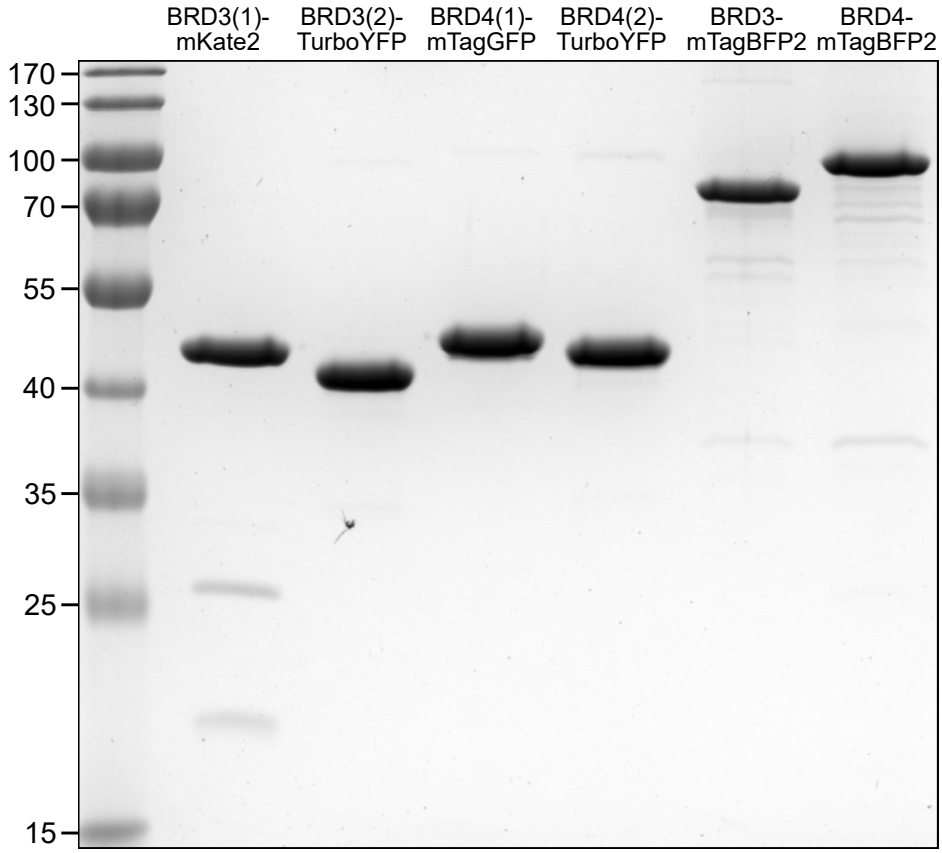


Figure 2.6: SDS-PAGE gel of all 6 bromodomain proteins used in this work.

An SDS-PAGE gel of the purified proteins is displayed in 2.6. Protein concentration and storage buffers are listed in Table 2.1.

2.1.4 Deletion mutants of histone-derived sequences

As a first step in the optimization of the peptide substrates, truncated peptide sequences on the basis of known binding substrates were synthesized on a SPOT membrane. The sequence AKRHR-KAc-VLRDN (termed H4K20ac) was chosen as a starting sequence for BD2 domains and the di-acetylated sequence SGRG-Kac-GG-Kac-GLGY (termed H4K5/8ac) was selected for BD1 domains as both exhibited the highest experimentally determined affinities in [68] for BRD3(2) and BRD4(1), respectively. BRD3(2) and BRD4(2) on the one hand and BRD3(1) and BRD4(1) on the other hand were grouped because they were expected to possess similar binding properties, resulting from sequence homologies of 80 % and 81 %, respectively [232]. All peptides were synthesized including a C-terminal Ahx which acted as a spacer between peptide and matrix. Each sequence

was truncated step by step, starting from the N- as well as the C-terminus, resulting in all possible truncation products. A study published in [66] mutated the linker between the two acetyllysines in the H4K5/8ac sequence and found that the native sequence bearing two glycines was optimal for BRD3(1) and BRD4(1). Based on this, the two glycines connecting both acetyllysines in the BD1 sequence were not removed in the truncation process or replaced in the following step. After the synthesis of the deletion mutants, the array-bound peptides were deprotected, washed and incubated with the corresponding BRD fusion protein. Bovine serum albumin was used to prevent unspecific binding and the concentration of the input proteins were adjusted to published K_D-values (6.8 µg/µL for BD1 binding to H4K5/8ac and 10.5 µg/µL for BD2 binding to H4K20ac). Membranes were washed and the fluorescence intensity of bound proteins was determined. An analysis of pull-down experiments with BRD3(2)-TurboYFP and BRD4(2)-TurboYFP is shown in figure 2.7.

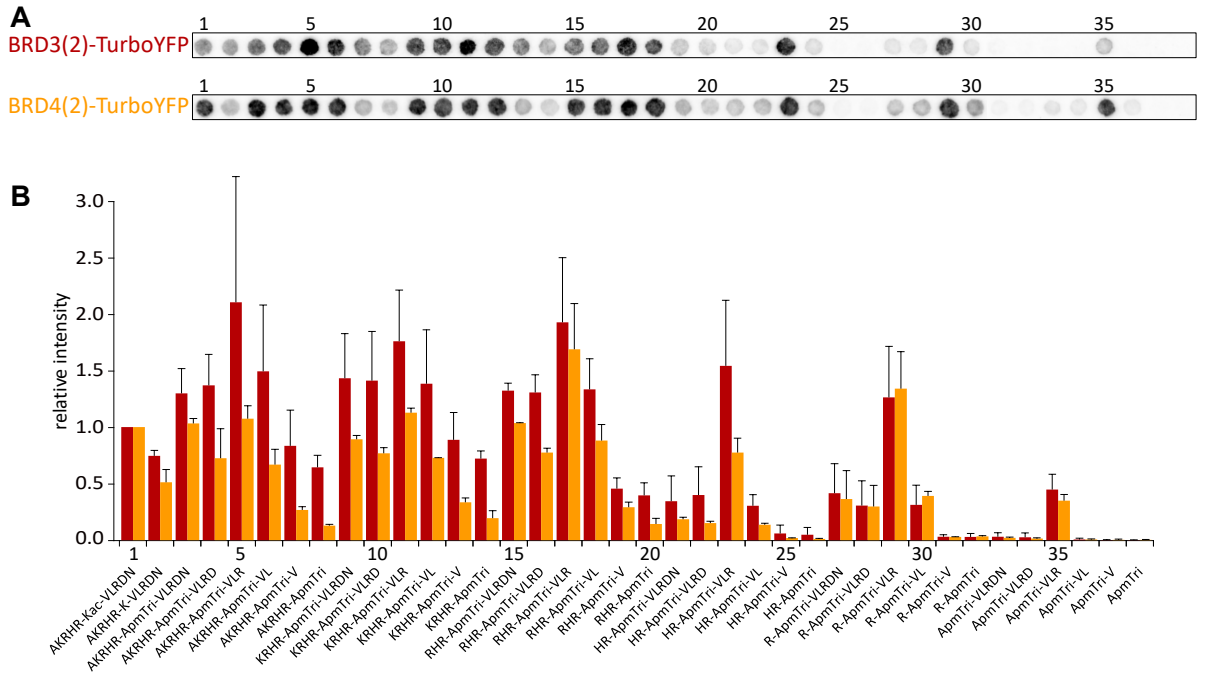


Figure 2.7: Results from pull-down experiments with truncated H4K20ApmTri and BRD3(2) as well as BRD4(2) fusion proteins. A) The fluorescent intensity of the fluorescent protein on each spot bearing a different truncation mutant, numbers indicate sequence found in B). The quantified fluorescence intensity is displayed in B) as a bar graph, red bars representing the intensity for BRD3(2)-TurboYFP, yellow bars showing the intensity for BRD4(2)-TurboYFP. Values are normalized to the intensity of the non-truncated H4K20ac sequence serving as control and standard deviation is indicated (n=3 for BRD3(2)-TurboYFP and n=2 for BRD4(2)-TurboYFP).

Fluorescence intensity of BRD3(2)- and BRD4(2) fusion proteins diminishes as expected with decreasing peptide length. For the N-terminal truncation, the slope of reduction is low until the point where the first arginine is truncated, at which the intensity drops significantly for both proteins. Truncation of the C-terminus shows a clear pattern of an increase in fluorescence intensity for valine-leucine-arginine over not only the more truncated peptides but also the native sequence. The increase for these three amino acids is even observable in the truncation products without the five N-terminal amino acids, which show little fluorescence for all other C-terminal mutants. A different method to visualise the data was used for the identification of the truncation product that displays the best binding towards BRD3(2) and BRD4(2) in table 2.3.

Table 2.3: Results from pull-down experiments with truncated H4K20ApmTri and BRD3(2) as well as BRD4(2) fusion proteins by total relative intensity of N- and C-terminal sequences. Highest values are underlined.

Total of N-terminal fluorescence intensities						
Protein	**AKRHR-**	**KRHR-**	**RHR-**	**HR-**	**R-**	**-**
BRD3(2)	<u>7.741</u>	7.591	6.742	2.700	2.358	0.515
BRD4(2)	3.898	4.047	<u>4.815</u>	1.283	2.458	0.403
Total of C-terminal fluorescence intensities						
Protein	**-VLRDN**	**-VLRD**	**-VLR**	**-VL**	**-V**	**-**
BRD3(2)	4.844	4.818	<u>9.038</u>	4.834	2.270	1.843
BRD4(2)	3.531	2.735	<u>6.354</u>	2.815	0.949	0.519

Here, the relative fluorescence intensities of all peptides starting with a certain N-terminal sequence or ending with a specific C-terminal sequence were summed up. In case of the N-terminal truncation, the sequence RHR- shows the highest sum of relative intensities for BRD4(2) and a still high relative intensity for BRD3(2). The C-terminal truncation experiments resulted in the -VLR minimal sequence for both proteins. As a result of the deletion mutant experiments for BD2, the complete minimal sequence RHR-ApmTri-VLR (termed sH4K20ApmTri) was chosen for BRD3(2) and BRD4(2).
Deletion mutant experiments for BD1 were performed with the aforementioned sequence SGRG-Kac-GG-Kac-GLGY. After expression of the BRD3(1)-mKate2 construct and

pull-down experiments, the fluorescent protein was found incompatible with fluorescent properties of the membrane. Based on this, only BRD4(1) was used to screen peptides in all three steps of the optimization experiments. The results of the pull-downs with truncated peptides and BRD4(1)-TagGFP2 are displayed in figure 4.1 of the appendix. Peptides truncated beyond the N-terminal Arg residue showed no BRD4(1) recruitment above background. For the C-terminus, the removal of leucine resulted in a steep reduction of fluorescence intensity. The overall relative intensity gradually declined with more truncation and there is no clear intensity peak as in the pull-down experiments for the BD2 domain, leading to the choice of a sequence with only one amino acid removed at the N- as well as the C-terminus. GRG-KAc-GG-KAc-GLG (termed sH4K5/8ac) was therefore chosen as the minimal sequence as a balance between relative intensity and peptide length.

2.1.5 Affinity mutants and verification

The previous chapter focused on reported acetylation sites as recognition sequences for Brds. This chapter describes the search for new binding sites based on screening of randomized peptide arrays referred to as affinity mutants arrays. To achieve this, every amino acid in the deletion mutant peptide was replaced with every canonical amino acid except cysteine. Cysteine was excluded to prevent the formation of inter- or intramolecular disulfide bonds. The theoretical library covers 19^6 different peptides for all of the 6 variable positions. This number cannot be synthesized by the SPOT method. Instead, a mixture based synthetic peptide combinatorial library (SPCL) for SPOT SPPS was developed. In this method, two amino acids per spot were defined, while the four other positions contained an isokinetic mixture of the 19 amino acids mentioned above. This resulted in 19^4 contained in each spot which had two residues in common. In each experiment, 19^2 peptide mixtures with two unique fixed amino acids per spot were synthesized in triplicates on spot-membranes.

Deletion mutants for BD1 and BD2 both had six variable positions, resulting in three sets of experiments. Two deconvolution strategies were tested for the SPCL, shown in figure 2.8. In the iterative strategy, 19^2 different subsets are screened, each including two fixed amino acids while the other 4 positions are randomized. The amino acid pair of the best performing subset is retained for the second set, in which again 19^2 subsets are screened, with the remaining two positions being randomized. This process is repeated until the peptide with the best binding properties towards the target protein is deduced. The additive method screens each subset independently, which means that no amino acid

is retained on any position. Ultimately, the additive method was chosen because intensity stained spots of tight binding sequences were randomly distributed over the array and did not cluster, which proved advantageous for the readout. The additive strategy was ultimately favored, on the one hand because the difference between spots was stronger and best binding amino acids were easier to deduce. On the other hand, it also offered the possibility to select multiple combinations of N_x per set which could be combined and independently verified. Some of the best performing peptides from the iterative strategy were nonetheless tested in the verification experiments for the affinity mutants, but they were found to enrich the targeted bromodomain not as strong as the peptides resulting from the additive strategy, as shown in fig 2.10.

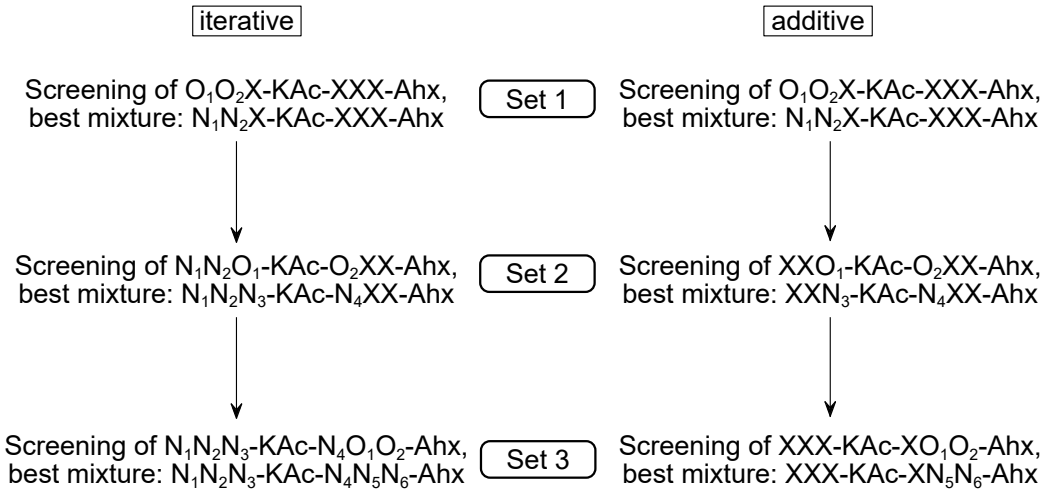


Figure 2.8: Two deconvolution strategies shown for the deletion mutant for BRD3(2) and BRD4(2). O_1 and O_2 are the amino acids each varied with 19 canonical acids per set, N_x are the amino acids chosen after each set showing the highest binding and X is an equimolar mixture of 19 amino acids.

As mentioned before, each set of affinity mutant arrays was synthesized in triplicates. After the synthesis, the peptides were deprotected on the membrane, washed and incubated with the respective fluorescent BRD fusion protein at a concentration of 10.5 µM for BD2 proteins and 6.8 µM for BRD4(1), based on reported K_Ds. Applying a recycling protocol to the membrane allowed for multiple pull-downs with different bromodomains. Fluorescence signals were recorded, normalized and the average of the triplicates displayed as a heat map. As a representative example, the first set of affinity mutant screening of BRD3(2)-TurboYFP is shown in figure 2.9.

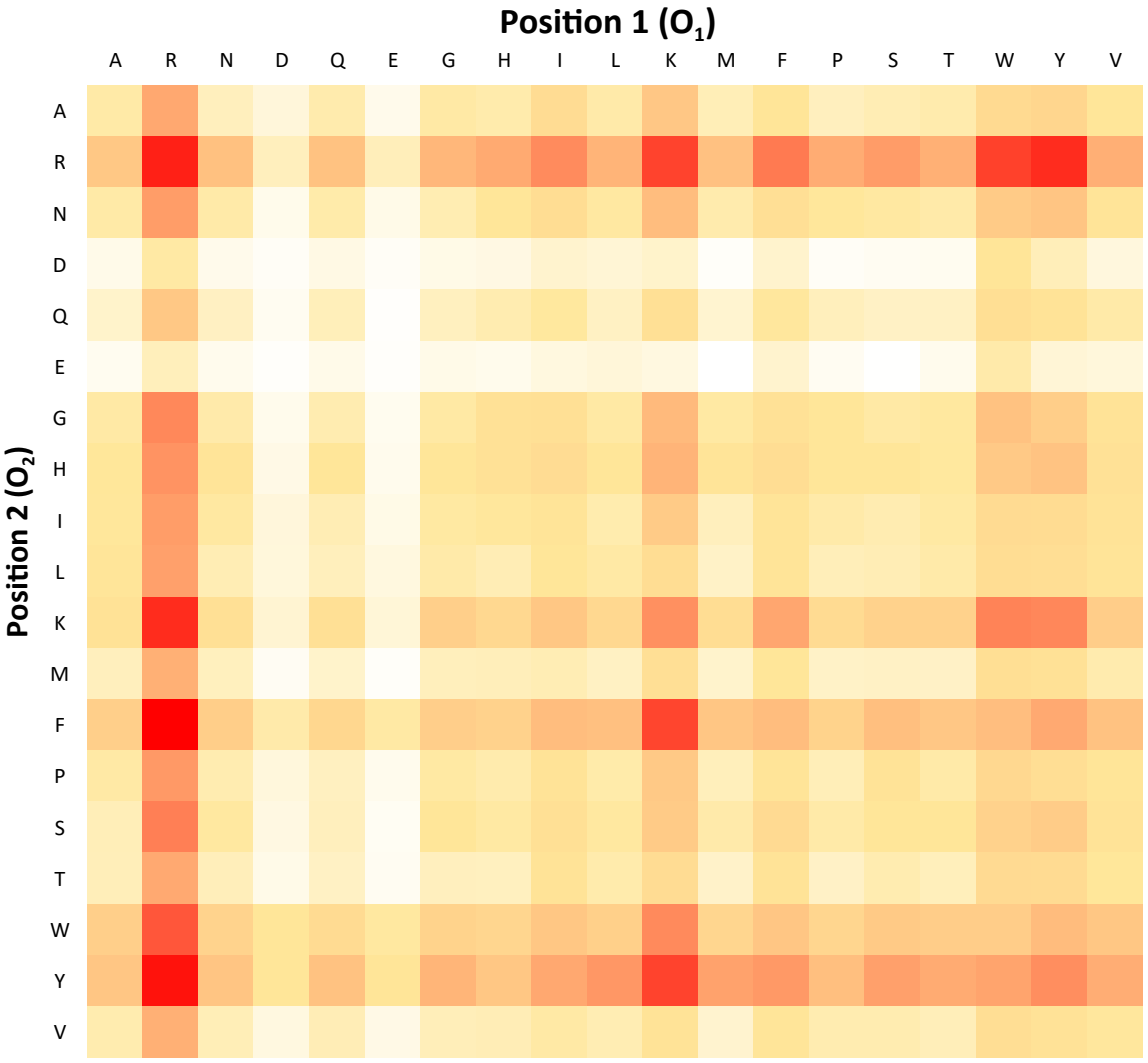


Figure 2.9: Average of fluorescence intensity of three pull-down experiment with BRD3(2)-TurboYFP and the mixture based SPCL for the peptide sequence O_1O_2X-KAc-XXX-Ahx normalized and displayed as a heat map. The amino acid for position O_1 is set for the indicated column, while the amino acid for position O_2 is set for the indicated row. A color gradient indicates the results, in which red displays the highest fluorescence and white displays the lowest fluorescence in this set.

While acidic amino acids weakened binding at both positions, basic amino acids as well as aromatic amino acids led to tighter binding of the BRD-fluorophore fusion protein. Arginine-histidine at position 1 and 2, as found in the sH4K20ac sequence, only led to average fluorescence in this pull-down experiment. The strongest binding was observed for the spots containing the amino acid pairs arginine-phenylalanine and arginine-tyrosine at position 1 and 2. These two pairs were selected for the first two positions of the

affinity mutant for BRD3(2). The heat map for the second set addressing the 3rd and 4th position in the peptide can be found in the appendix (figure 4.2). Spots with peptides containing arginine in position 3 and 4 showed again strong binding, while lysine did not show binding as prominent as in the first set of affinity mutants. Arginine-valine as the combination found in the sH4K20ac sequence at position 3 and 4 led to intermediate fluorescence, comparable to the results of the pair arginine-histidine at position 1 and 2 of the sH4K20ac sequence in figure 2.9. The strongest fluorescence intensity was found if the peptides included arginine-proline and arginine-tryptophan at position 3 and 4. In the heat map showing the results for the fixed positions 5 and 6 (figure 4.3), arginine is the most prominent amino acid in the 6th position, resulting in high fluorescence intensity in pair with many other amino acids in the 5th position. The sH4K20ac sequence contains valine-arginine in these two positions and this pair shows a relatively high fluorescence intensity compared to the results of the sH4K20ac sequence in the two previous sets 2.8. Peptide spots containing phenylalanine-arginine, tryptophan-arginine, tyrosin-arginine and leucine-arginine at position 5 and 6 showed high normalized fluorescence after pull-down with BRD3(2)-TurboYFP and were selected to be validated.

BRD3(2) and BRD4(2) share 80 % sequence identity and were expected to show similar binding patterns (the sequences of the domains is shown in 4.1). The affinity mutant experiments for BRD4(2) confirmed this assumption, as shown in figure 4.4 of the appendix. BRD4(2)-TurboYFP shows high binding towards peptides with arginine at the first and second position. Since too many charged residues enhanced the chance of unspecific binding, pairs containing more than one basic amino acid were omitted from further investigations. The highest fluorescence intensity for position 1 and 2 was observed for peptides containing arginine-phenylalanine and tyrosine-arginine at these positions. In the second set of the screening, the interaction profile of BRD4(2) (figure 4.5) was again similar to that of BRD3(2) (figure 4.2). The selected sequence for position 3 and 4 was arginine-tryptophan, which is the same result as for BRD3(2). In the third set, BRD3(2)-TurboYFP showed high affinity to multiple sequences (figure 4.3), while in the experiments with BRD4(2)-TurboYFP only two pairs of amino acids exhibited strong fluorescence signals (figure 4.6). These two pairs, tryptophan-arginine and phenylalanine-arginine, were also found amongst the hits for BRD3(2). While the leucine-arginine pair found in the sH4K20ac sequence showed high normalized fluorescence intensity in the screening for BRD3(2), only an intermediate fluorescence intensity was observed in the experiments with BRD4(2).

The basis for the affinity mutant experiments with BRD3(1)- and BRD4(1) fluorophore

fusion proteins was the deletion mutant sH4K5/8ac (GRG-KAc-GG-KAc-GLG-Ahx). As mentioned before, BRD3(1)-mKate2 could not be used for screening because the intrinsic fluorescence of the membrane overlapped with the signal of the fluorophore. The screenings of set one, two and three of the affinity mutants for BRD4(1)-TagGFP2 are displayed in the appendix in figures 4.7, 4.8 and 4.9, respectively. While the amino acid pair glycine-arginine found in the sH4K5/8ac sequence did not show strong fluorescence intensity, the inverted pair showed the strongest binding and was therefore selected in the first and second position for further experiments (figure 4.7). Positively charged amino acids in general lead to strong fluorescence intensities, which was also found in the library screenings with BD2 bromodomain-fluorophore proteins. Spots with peptides containing glycine in combination with leucine or isoleucine in position 3 and 4 exhibited high fluorescence intensity after pull-down (figure 4.8). Notably, serine, threonine and alanine in combination with leucine and isoleucine exhibited intermediate strong fluorescence intensities, while arginine again showed high affinity towards the BRD4(1)-TagGFP2 in combination with several amino acids. Glycine-leucine was ultimately selected as the set amino acid pair in these positions. In the last set of affinity mutants screening of BRD4(1), a clear preference for peptides with arginine in position 6 was detectable (figure 4.9), exhibiting strong fluorescence intensity signals in combination with many amino acids. Tyrosine-arginine was chosen as the set pair for position 5 and 6, again omitting arginine-arginine to reduce unspecific binding.

The results of all three affinity mutant experiments using a mixture based SPCL are summarized and displayed in table 2.5.

Table 2.5: Best performing amino acid pairs in all three sets of affinity mutant experiments for BRD3(2), BRD4(2) and BRD4(1) fusion proteins deconvoluted with the additive strategy.

	Position 1 + 2	**Position 3 + 4**	**Position 5 + 6**
Sequence: $N_1N_2N_3$*-KAc-*$N_4N_5N_6$*-Ahx*			
BRD3(2)	RF; RY	RW	FR; WR; YR; LR
BRD4(2)	RF; RY	RW	WR; FR
Sequence: $N_1N_2N_3$*-KAc-GG-KAc-*$N_4N_5N_6$*-Ahx*			
BRD4(1)	RG	GL	YR; WR; PR

In total 40 potential binding sequences of BRD3(2)/BRD4(2) were unraveled, including results of the additive and iterative deconvolution. A further SPOT membrane was synthesized containing only these hits, including not only acetyllysine but also ApmTri as well as lysine at the central position. Pull-downs were performed in triplicates. Since BRD3(2) and BRD4(2) performed similarly in all three sets of affinity mutant screenings, BRD4(2) was not tested in this experiment. As the pool of best performing binding sequences for BRD4(1) was much smaller than for the BD2 bromodomain proteins, they were directly verified in pull-down assays on peptide probes immobilized on Sulfolink™ resin. Fluorescence of bound protein was quantified, normalized and the means displayed as a bar graph in figure 2.10.

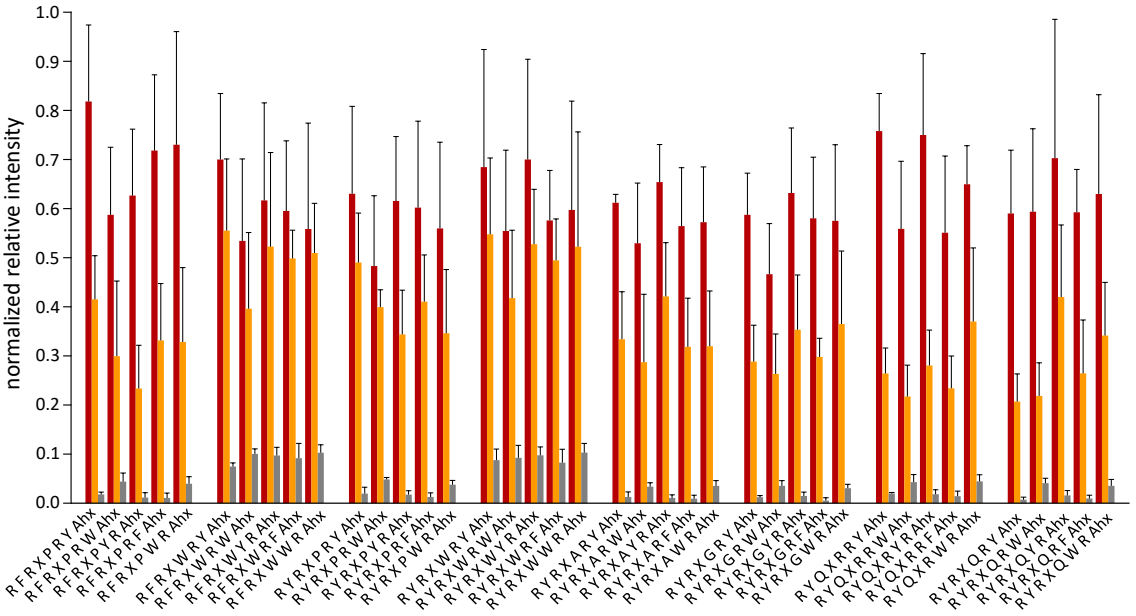


Figure 2.10: Min-max normalized fluorescence intensity of the pull-downs with BRD3(2) and the best performing affinity mutants. Red bars represent the results of the peptides incorporating acetyllysine, orange graphs represent ApmTri-containing peptides and gray graphs represent peptides containing lysine. Standard deviation is indicated (n=4 for KAc, ApmTri; n=5 for Lys).

In general, ApmTri was able to recruit BRD3(2) in all the tested probe sequences, but the normalized fluorescence signal was reduced when compared to the KAc probes indicating weaker binding. However, some of the tested sequences recruited BRD3(2) almost as effective as the KAc peptides. Arginine at position 3 and proline at position 4 resulted in the best performance for acetyllysine containing peptides. Amongst all ApmTri containing

peptides, the sequences incorporating arginine at position 3 and tryptophan at position 4 show the highest normalized fluorescence intensity. On the other hand, the lysine controls of these sequences show a relatively high fluorescence intensity, indicating unspecific binding contributing to the enrichment of the BRD-fluorophore fusion protein. Yet, the average signal intensity of BRD3(2) bound to ApmTri-containing probe was still 5-fold higher than the lysine probe.

It should further be noted that screening of the deletion mutant arrays showed enhanced Brd binding of the lysine control (fig 2.7). However, pull-downs with Sulfolink™ peptides showed that binding to the lysine control of BD2 domains on SPOT membranes does not result in strong binding to the control on Sulfolink™ resin.

To explore if the matrix influenced the binding between the control and the affinity probes as well as to verify the findings of the SPOT array, peptides were resynthesized, purified and immobilized on Sulfolink™ resin. Binding of BRD-fluorophore fusion proteins to immobilized peptides was tested in pull-down assays. Five sequences of the strongest binders in the affinity mutant array experiments shown in table 2.5 were synthesized, once containing ApmTri and once containing lysine as control, resulting in a total of 10 peptides. After purification, all peptides were immobilized on Sulfolink™ resin and pull-down experiments were performed with BRD3(2)-TurboYFP and BRD4(2)-TurboYFP at 10.5 μM in triplicates. Washing was followed by elution of bound proteins, and eluates were analyzed by SDS-PAGE. Representative Coomassie-stained SDS-PAGE gels of pull-downs with BRD3(2) including statistical analysis of quantified band intensity are shown in figure 2.11. All immobilized ApmTri-containing peptide probes enrich BRD3(2) significantly stronger than the lysine control. BRD3(2) was most efficiently retained on probe 3 (RFR-ApmTri-WYR) but the difference in enrichment on other ApmTri-containing probes was only significant for probe 1. The finding of the pull-down experiments with BRD4(2)-TurboYFP with this set of probes are shown in figure 4.11 of the appendix.

Collectively, pull-downs with BRD4(2) showed similar interaction patterns with the ApmTri-containing probes as BRD3(2), confirming the utility of these sequences for targeting the BD2s of BRD3 and BRD4. The difference between the probe and the control hints to the observation of smaller differences between positive and negative control on the membranes in previous experiments (figure 2.10) being caused by matrix effects. Summarizing the findings so far, the step-wise truncation of the H4K20ac sequence delivered efficient 7-mer custom probes for BRD3(2) and BRD4(2), but when KAc was replaced by ApmTri, the binding efficiency was reduced. In order to uncover tight binding sequences with ApmTri as an anchoring residue, a combinatorial approach was used delivering ApmTri-containing

probe P3 as its most efficient binder. To illustrate this development, pull-downs with probes that paved the way to P3 were performed and analyzed side by side.

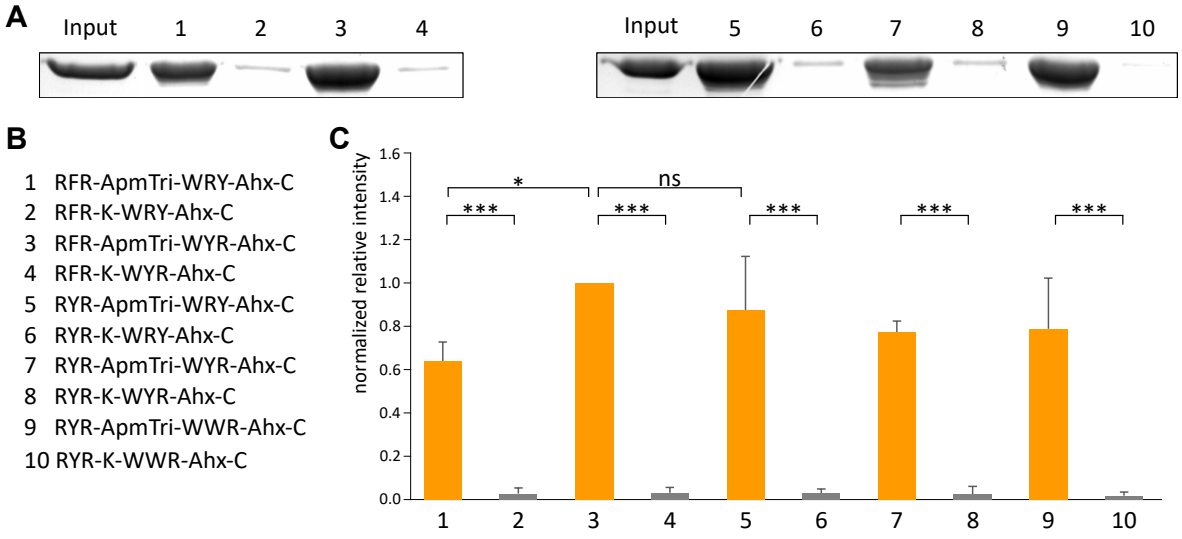


Figure 2.11: Pull-down experiments for BRD3(2)-TurboYFP with affinity mutants. A) Coomassie-stained SDS-PAGE bands with BRD3(2)-TurboYFP and the probes. B) Display of the sequences of each peptide. C) The quantified intensity after the pull-down was normalized to the intensity of the probe exhibiting the strongest band and is displayed as the mean (n=3). Orange graphs represent ApmTri-containing peptides and gray graphs represent peptides containing lysine. P-values were calculated using ANOVA and Tukey's HSD test for multiple comparisons. *=$p<0.05$; **=$p<0.01$; ***=$p<0.001$; ns=not significant.

In order to verify the gain in affinity, the native sequence, the deletion mutant sH4K20 and the optimized affinity mutant P3 were immobilized and pull-downs were conducted using BRD3(2)-TurboYFP. P3 was chosen as it displayed the strongest enrichment of BRD3(2) in the previous experiment (figure 2.11). Fluorescence intensity was normalized to functionalization of the probes. The pull-downs were performed in triplicates, allowing for statistical analysis as displayed in figure 2.12. This comparison showed that the interaction between BRD3(2) and P3 was significantly improved when compared to sH4K20ApmTri and was now on a similar level as the acetylated wild-type sequence. The deletion mutant sH4K20ac also displayed significantly higher enrichment of the Brd-fluorophore protein in comparison to the non-truncated H4K20ac sequence, underlining the benefit of the truncation experiments. No enrichment of the Brd-fluorophore fusion protein was seen on the lysine control probe.

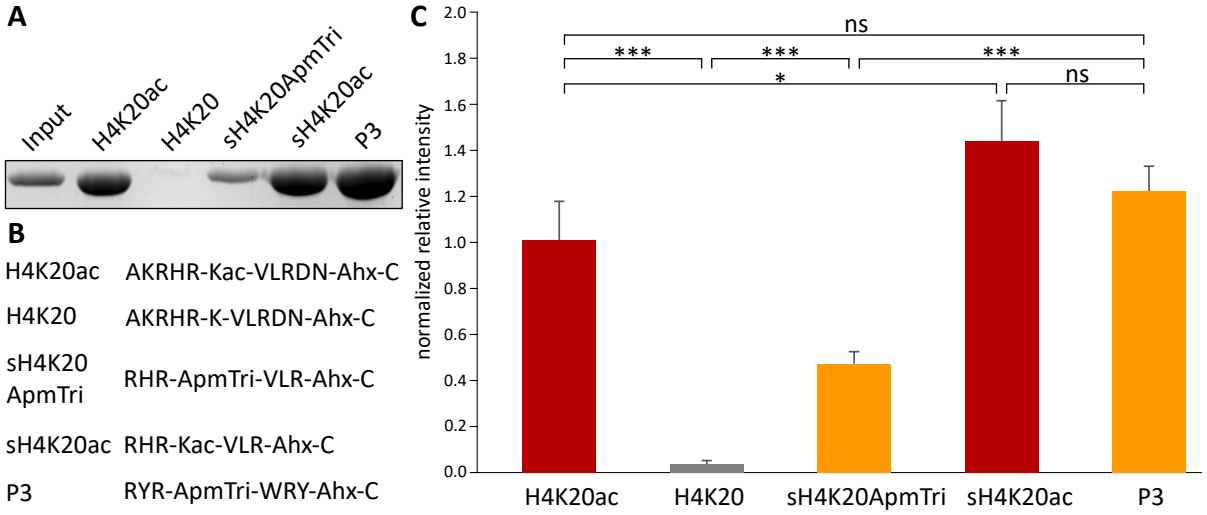


Figure 2.12: Pull-down results for BRD3(2)-TurboYFP with native sequence, deletion mutant and affinity mutant. A) Coomassie-stained SDS-PAGE bands with BRD3(2)-TurboYFP. B) Sequences of the peptides used in the pull-down. C) Red bars represent the results of the peptides incorporating acetyllysine, orange graphs represent ApmTri-containing peptides and gray graphs represent peptides containing lysine. The quantified intensity was normalized to the input and is displayed as the mean (n=3). P-values were calculated using ANOVA and Tukey's HSD test for multiple comparisons. *=p<0.05; **=p<0.01; ***=p<0.001; ns=not significant. The P3 sequence with lysine instead of ApmTri does not mediate binding of BRD3(2)-TurboYFP as illustrated in 2.11.

Eight peptides were immobilized via Sulfolink™ and pull-downs with BRD3(1) and BRD4(1) fusion proteins were conducted. These eight peptides were composed of three of the optimized sequences containing two ApmTris (as shown in table 2.5) as well as the same sequences with two lysines as a control. The starting sequence was also included, once containing acetyllysine and once containing lysine. Results as well as the sequences of the peptides are displayed in the appendix 4.10. The three optimized sequences containing ApmTri display enrichment of both proteins in comparison to the lysine probe. BRD3(1)-mKate2 and BRD4(1)-TagGFP2 display very strong interaction towards the KAc-containing probe H4K5/8ac. In contrast to the peptide sequences optimized for BD2 proteins, the peptides bearing ApmTri as a binding moiety were not able to outperform the native sequence.

2.1.6 Stability mutants and verification

After enhancing the strength of the interactions between the BRDs and the ApmTri-containing peptide, the next objective was to enhance the stability of the peptide against

proteases. D-amino acids readily enhance proteolytic stability, as the native targets for proteases are peptides composed of L-amino acids. N-methylated amino acids can also increase the half-life of peptides *in vivo*. The methyl group blocks the cleavage site of proteases, which leads to a decrease in proteolytic efficiency [233]. The impact of both of these stabilizing residues to enhance the proteolytic resistance were investigated in this work. The so-called stability mutant peptides were synthesized in three rounds on SPOT membranes, in the first round only substituting one amino acid at a time with its N-methylated or D-derivative, in the second round substituting two and in the third round three residues.

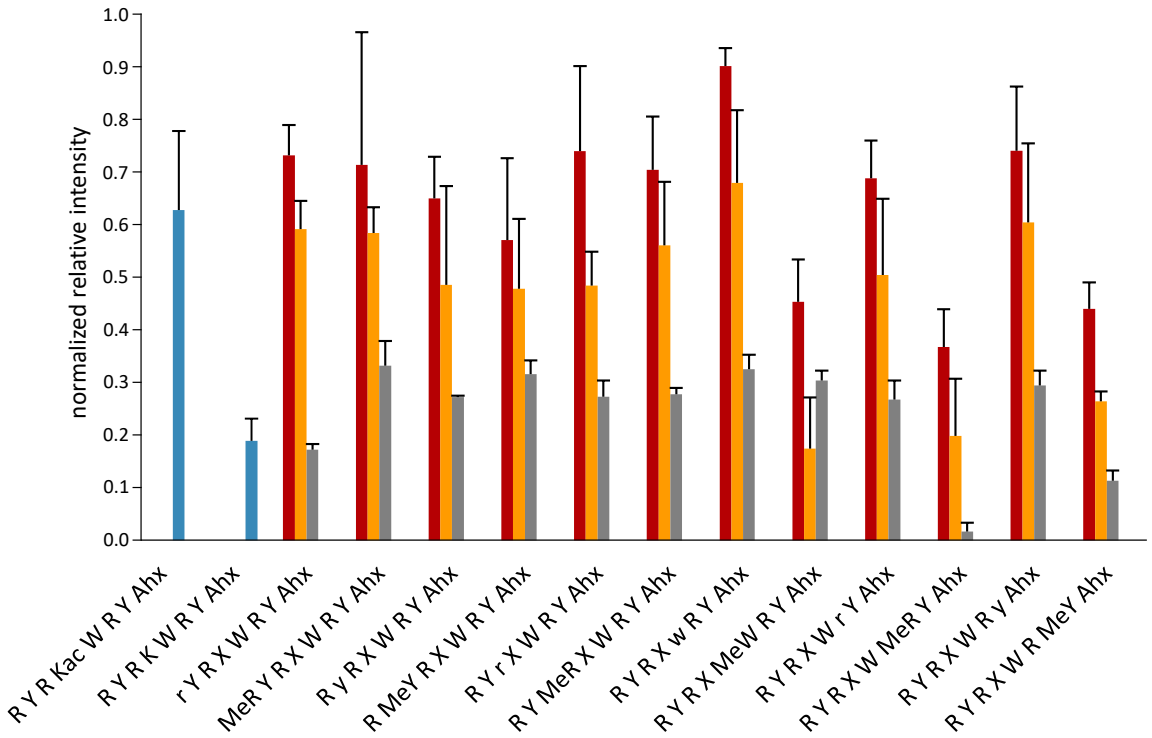


Figure 2.13: Normalized fluorescence intensity of pull-down results for BRD3(2)-TurboYFP with the first round of stability mutants of P5 (RYR-ApmTri-WRY-Ahx). The positive and negative control are displayed as blue bar graphs. Red bars represent the results of the peptides incorporating acetyllysine, orange graphs represent ApmTri peptides and gray graphs represent peptides containing Lys. Standard deviation is indicated.

Peptides were synthesized in triplicates on a membrane, deprotected and suspected to pull-downs with BRD3(2)-TurboYFP. After washing, fluorescence of bound fusion protein was recorded and visualized. Amongst the outcome of the sequences tested, results for

peptide 5 (for sequence see 2.11) are analyzed and depicted as a representative example. In the first round, substituting only one amino acid per peptide with its N-methyl or D-derivative did not impact the enrichment of the bromodomain significant in most of the peptides, as shown in figure 2.13.

While the peptides containing acetyllysine overall enriched BRD3(2) stronger than the peptides containing ApmTri, the difference was not significant. Notably, the control peptides containing lysine showed only very limited enrichment of the bromodomain protein. This effect was observed before (fig 2.11) and is most likely again attributed to differences between SPOT membrane and the Sulfolink™ resin. The ApmTri-containing sequence showing the strongest enrichment of the BRD-fluorophore construct in figure 2.13 (RYR-ApmTri-wRY-Ahx) was chosen for the next round of stability mutant experiments. In the second round of stability mutant arrays, the substituted amino acid of the first round was fixed and all other amino acids were substituted with their N-methylated and D-derivatives, leading to peptides that each contained two N-methylated and D-amino acids. N-methylated amino acids that led to poor enrichment of the bromodomain protein in the first round were omitted for the second round.

Results of the second round for BRD3(2) are depicted in fig 2.14. With these findings, a third round of stability mutant arrays was designed, now fixing two substituted amino acids and substituting the remaining positions with their N-methylated or D-derivates. Three combinations of N-methylated and D-amino acids from the results of round two were tested in round three, since the enrichment of the protein on these sequences was equally strong. The outcome is displayed in figure 4.12 of the appendix. In this round, the overall enrichment of the bromodomain starts to diminish in comparison to the sequence of the previous round. As an example, this effect is observed in the mutants for the sequence rYR-X-wRY-Ahx. Here, all peptides containing an additional third N-methylated or D-amino acid show lower enrichment than the sequence containing only the D-arginine and D-tryptophan at the first and fifth position, respectively. Overall, it was observed that substituting two amino acids in the sequence of the final affinity mutants did not diminish the enrichment of the bromodomain protein strongly. However, D-amino acids performed much better than N-methylated amino acids, with the latter leading to a strong decrease of enrichment of the bromodomain protein at most positions. With more than two positions substituted with N-methylated or D-amino acids in the peptide sequence, binding of the bromodomains was continuously weakened, resulting in the selection of peptides containing two N-methylated or D-amino acids as the final stability mutants. These finalized probe sequences were used for bivalent ligands of BRD3 and BRD4 designed in the next step.

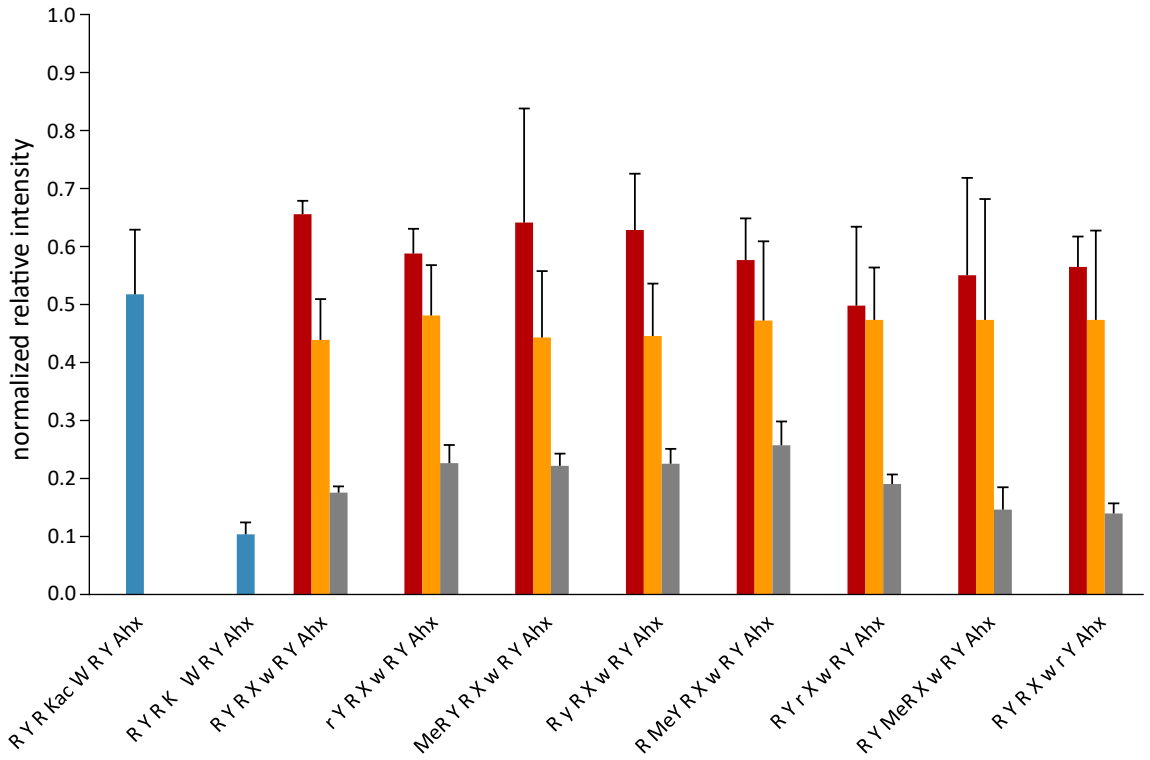


Figure 2.14: Min-max normalized fluorescence intensity of pull-down results for BRD3(2)-TurboYFP with the second round of stability mutants of P5 (RYR-ApmTri-WRY-Ahx). The positive and negative control are displayed as blue bar graphs. Red bars represent the results of the peptides incorporating acetyllysine, orange graphs represent ApmTri peptides and gray graphs represent peptides containing Lys. Standard deviation is indicated.

2.1.7 Investigation of the optimal linker length

In the next step of the design process, the optimized and stabilized ApmTri-containing sequences were combined to a bivalent ligand addressing both Brds of BRD3 and BRD4 simultaneously. Since the distance between the two bromodomains of the native bromodomain proteins BRD3 and BRD4 is defined, the linker length connecting the two inhibitor sequences appears to be crucial for efficient binding. To explore the influence of the length between the optimized inhibitor sequences for BD1 and BD2 on the recruitment of the bromodomains, three peptides with differing linker lengths were synthesized and immobilized, pull-down experiments were performed with full-length BRD3 and BRD4. As the ligand for BD1, the H4K5/8ac (SGRG-KAc-GG-KAc-GLGY) sequence was chosen and for BD2 and the affinity mutant 5 (figure 2.11) was used substituting ApmTri with

KAc. The choice for BD1 to include the acetyllysine-containing H4K5/8ac sequence was made because at the time of this experiment, no affinity mutants for BD1 were available. The peptides were synthesized and cleavage off the resin was followed by purification via RP-HPLC. To allow linkage of the N-terminal and the C-terminal peptide, a clickable system was developed.

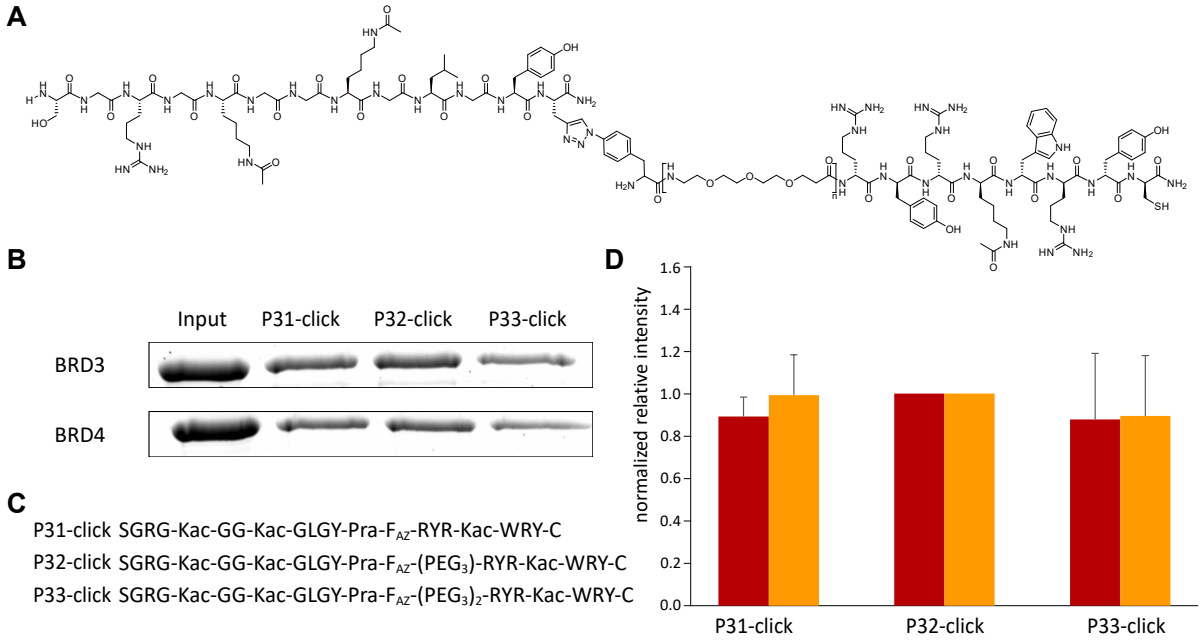


Figure 2.15: Linker investigation with pull-down experiments with BRD3-mTagBFP2 and BRD4-mTagBFP2. A) Structure of the probes, with n for PEG_n ranging from 0-2. B) Coomassie-stained SDS-PAGE bands of pull-downs. C) Sequence of the probes. D) Quantified intensity for BRD3 is displayed as red bars, intensities for BRD4 is displayed as yellow bars. The quantified intensity was normalized to the intensity of the probe exhibiting the strongest band and is displayed as the mean (n=3 for BRD3 and n=2 for BRD4).

The substrate sequence of BD1 (N-terminal peptide) and the substrate sequence of BD2 (C-terminal peptide) were coupled using Cu-catalyzed azide–alkyne cycloaddition (CuAAC). The variable linker was inserted into the C-terminal peptide. Sequences of the peptides can be found in figure 2.15. After linkage, the peptides were repurified via HPLC and the pure peptides were immobilized on a Sulfolink™ resin. Pull-downs with BRD3-mTagBFP2 and BRD4-mTagBFP2 are displayed in figure 2.15. While the binding to the probe with a single PEG_3-linker (P32-click) appeared to be stronger, the proteins were enriched nearly independently, hinting towards a minor impact of the linker lengths tested here. One polyethylene glycol (PEG_3) linker was chosen as the linker length for the bivalent ligand

construct as a compromise. After determining that the linker length between the substrate peptides for BD1 and BD2 domains had little influence on the enrichment of BRD3 and BRD4, enrichment depending on the acetylation state of the peptide was determined.

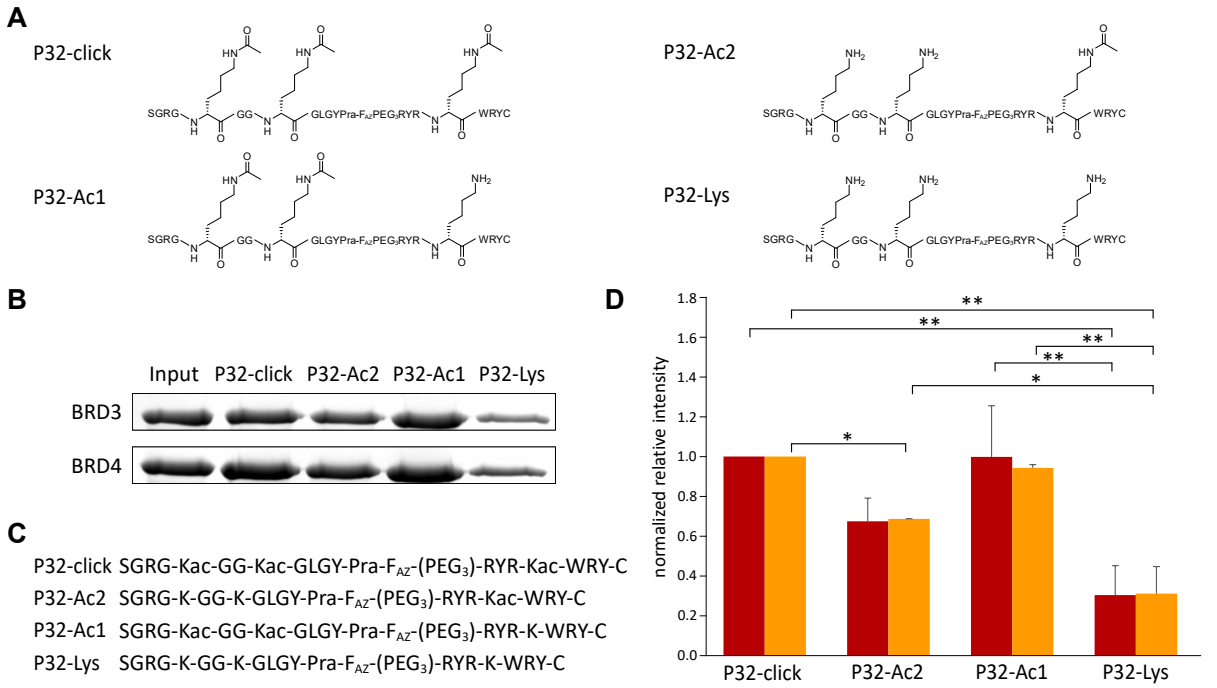


Figure 2.16: Linker test pull-down results for BRD3-mTagBFP2 and BRD4-mTagBFP2 with peptide P32 and controls. A) Structure of the probes. B) Coomassie-stained SDS-PAGE bands of pull-downs. C) Sequence of the probes. D) Quantified intensity for BRD3 is displayed as red bars, intensities for BRD4 is displayed as yellow bars. The quantified intensity was normalized to the intensity of the probe exhibiting the strongest band and is displayed as the mean (n=3). P-values were calculated using ANOVA and Tukey's HSD test for multiple comparisons. *=p<0.05; **=p<0.01; ***=p<0.001; ns=not significant. Significant p-values are displayed.

Data in a review by Filippakopoulos et al. [68] showed that BD1 bromodomains only showed weak binding towards mono-acetylated histone sequences, but displayed a higher affinity to di- or poly-acetylates sequences. BD2 bromodomains bound to mono-acetylated histone sequences with K_D-values similarly to BD1 binding di-acetylated peptides [68]. To explore the individual contribution of the ligand sequences to the overall binding of BRD3/BRD4, the peptide P32-click was synthesized in different acetylation states. The peptides were synthesized, purified and linked together by CuAAC. After purification of the peptide constructs, they were immobilized on Sulfolink™ beads and pull-downs

were conducted in triplicates with BRD3 and BRD4. Peptide sequences and quantified interaction profiles observed in pull-downs are displayed in figure 2.16. BRD3 and BRD4 showed similar enrichment on the probes. While the proteins were enriched the strongest on P32-click containing KAc in both sequences, the enrichment on P32-Ac1 was similar. In the peptide where only the peptide sequence for BD2 was acetylated, the enrichment was weaker, in the case of BRD4 this effect was significant. The lysine control P32-Lys exhibited the weakest binding. The findings suggest that the BD1 substrate peptide contributes the most to the overall binding. This could be due to the fact that this peptide is bound by the BD1 as well as the BD2 domain of the two bromodomain proteins tested in the pull-down, while the substrate peptide for BD2 would show much weaker binding by the BD1 domain.

2.1.8 Synthesis of cell-penetrating peptides

While small molecules showing a moderate polarity are often able to passively diffuse through the cell-membrane, most peptides are too polar to enter the cell by passive diffusion [234]. To overcome this challenge, biomolecules can be attached to cell-penetrating peptides (CPPs) which facilitate their cellular uptake. CPPs can be categorized according to their physicochemical properties into a cationic, an amphipathic and a non-polar group. They can enter the cell through endocytosis, passive diffusion and translocation through the formation of inverted micelles [235]. In this work, five different cationic CPPs were synthesized for N-terminal attachment to the bivalent stable ligand (shown in 2.1.1). The sequences and the yields of the CPPs are shown in table 2.7.

Table 2.7: Sequences and yields of CPPs synthesized in this work. Cyclic sequences are underlined and references for the sequences are provided.

Name	Sequence	Yield	Source
sC18*	GLRCRLRKFRNK	16.33 mg	[236]
Poly-D/L-Arg	C-PEG3-KRrRrRrRrRE	2.3 mg	[237]
Poly-Arg	C-PEG3-RRRRRRRRR	5.2 mg	[238]
Penetratin	C-PEG3-RQIKIWFQNRRKWKK	8.33 mg	[239]
Nucleus targeting 3RK	C-PEG3-RrRK	7.12 mg	[240]

The linear CPPs poly-arg, penetratin and the nucleus targeting 3RK were synthesized via automated peptide synthesis on a 50 µmol scale each. N-terminal cysteine and a PEG-linker were introduced manually, the peptides were cleaved from the resin, purified by RP-HPLC and stored. Analytical data of purified peptides is displayed in figure 4.15 of the appendix.

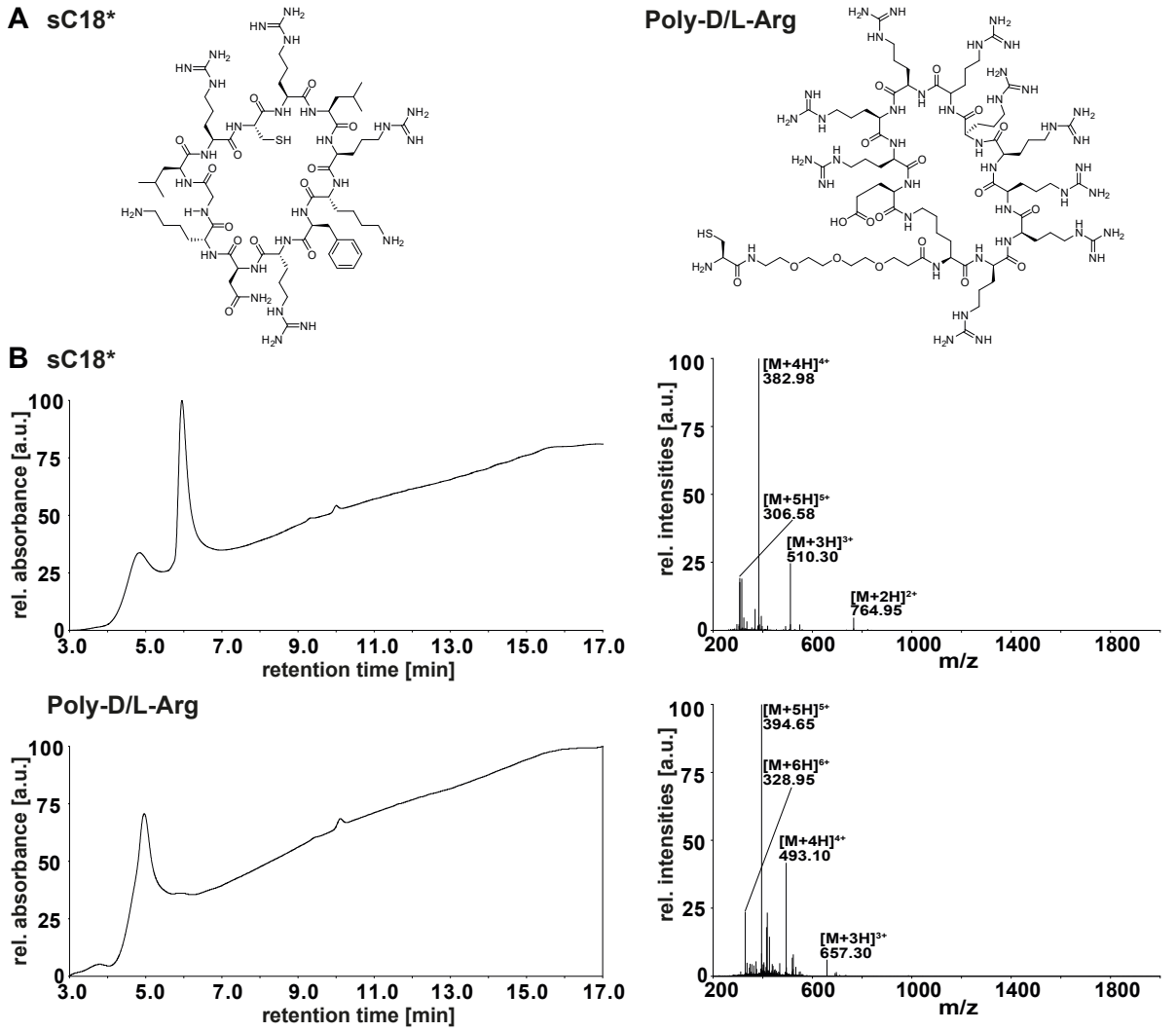


Figure 2.17: Structures and LC-MS spectra of cCPPs Poly-d/l-Arg and sC18*. A) Structures of the cyclic peptides. B) LC-MS spectra, left side: UV-chromatograms, right side: MS spectra.

Head-to-tail and side-chain-to-tail cyclization of cyclic CPPs (cCPPs) was conducted in solution, which required peptides having protected side-chains, an unprotected C-terminus as well as an unprotected N-terminus or side-chain moiety, respectively. The peptides

were synthesized on a TCP resin and the Lys(Dde) in poly-D/L-Arg was deprotected with hydrazine. Cleavage from the resin was conducted using mild acidic conditions, leaving remaining side-chain protection groups in place. Crude peptides were washed and lyophilized. The peptides were then dissolved in a high volume of DMF to achieve high dilution, thereby lowering the chance of multimer formation, and cyclization was carried out using HBTU, Oxyma and DIPEA. After the reaction, the now cyclic peptides were purified via RP-HPLC, lyophilized and stored. Structures and analytical data of Poly-D/L-Arg and sC18* are shown in figure 2.17.
In total, three linear and two cyclic CPPs were successfully synthesized and purified. With the completion of the CPP synthesis, all building blocks for the envisioned cell permeable bivalent BRD3/BRD4 ligand are provided allowing assembly and test in live cells in the near future.

2.1.9 Pull-down with cyclic peptides as BET inhibitors

While this project aiming at optimizing linear ligand sequences of BD1 and BD2 of BRD3 and BRD4 was ongoing, a screening campaign for cyclic BRD3 and BRD4 peptide ligands containing KAc was reported, resulting in binders in the picomolar range [195]. Several of the reported ligands exhibited strong selectivity towards a specific Brd while one candidate (3.2B) in particular bound to BD1 and BD2 of BRD3 and BRD4 with a mean K_D-value of 14 to 32 nM. In this work, the acetylated (3.2B-KAc) and the ApmTri substituted peptide (3.2B-ApmTri) as well as a lysine control (3.2B-K) were investigated for their ability to bind BRD3(2) and BRD4(1).

Table 2.9: Name and sequence of cyclic BET inhibitors as well as the Sulfolink™ connector peptide. The cyclic part of the peptides is underlined. KAc represents acetylated lysine and 4-azido-phenylalanine (F_{AZ}) as well as propargylglycine (Pra) were the functional groups required for joining the peptides via CuAAC.

Name	Sequence
3.2B-KAc	WSWLCK-KAc-YNLIH-$(PEG_3)_2$-Pra
3.2B-ApmTri	WSWLCK-ApmTri-YNLIH-$(PEG_3)_2$-Pra
3.2B-K	WSWLCK-K-YNLIH-$(PEG_3)_2$-Pra
Connector peptide	F_{AZ}-GSGAhx-C

For this, a linker peptide was ligated to the C-terminus of each of the three cyclic peptides via CuAAC, the conjugates were immobilized on Sulfolink™ resin and pull-downs were carried out. The sequences of the cyclic peptides and the linker peptide can be found in table 2.9.

The C-terminal propargylglycine and the two PEG linkers were coupled manually, while the rest of the peptide was synthesized automatically. After deprotection of the N-terminal Fmoc protection group, bromoacetic acid was conjugated to the free N-α-group using bromoacetic anhydride. Cleavage off the resin was followed by HPLC purification. While cyclization was planned to be conducted by gradually increasing the pH of the dissolved peptide solution, it was found that the bromoacetic linker readily reacted with the thiol moiety of cysteine directly after the HPLC purification. To rule out that the Nε of the lysine adjacent to cysteine reacted with the N-terminal bromoacetic moiety (see sequences in table 2.9), a spectroscopic test using Ellman's reagent (DTNB) was conducted in order to analyze if free thiol is still present. 3.2B-KAc, 3.2B-ApmTri and 3.2B-K together with a peptide containing a free thiol and one containing no thiol were incubated with DTNB solution at a final peptide concentration of 10 µM and the absorbance at 412 nm was recorded.

The analysis of the Ellman's assay is displayed in figure 4.13 of the appendix. While the thiol control peptide bearing a free cysteine showed a strong absorbance at 412 nm, the three peptides 3.2B-KAc, 3.2B-ApmTri and 3.2B-K showed absorbance at the background level, indicating cyclization via the cysteine residue.

After confirmation of the intended cyclic products, the peptides were ligated to the connector peptide via CuAAC, followed by purification and ultimately immobilization on Sulfolink™ resin. Pull-downs were conducted in triplicates with BRD3(2)-TurboYFP and BRD4(1)-TagGFP2. The interaction profile of BRD3(2) is displayed in figure 2.18. The cyclic ApmTri-containing probe 3.2B-ApmTri was able to enrich BRD3(2) strongly, while the KAc control exhibited even stronger binding. The lysine control was not bound by BRD3(2).

Pull-down experiments with BRD4(1) on the other hand showed that the protein would not bind to 3.2B-ApmTri under selected conditions while still showing strong enrichment on the KAc probe, as shown in figure 4.14 of the appendix. In contrast to BRD3(2), the enrichment of BRD4(1) on 3.2B-ApmTri was not significantly stronger than on the lysine control. Overall, the acetylated cyclic peptide 3.2B-KAc reported in [195] did show a strong enrichment of BRD3(2) and BRD4(1), while only BRD3(2) was strongly enriched on the peptide incorporating ApmTri (3.2B-ApmTri).

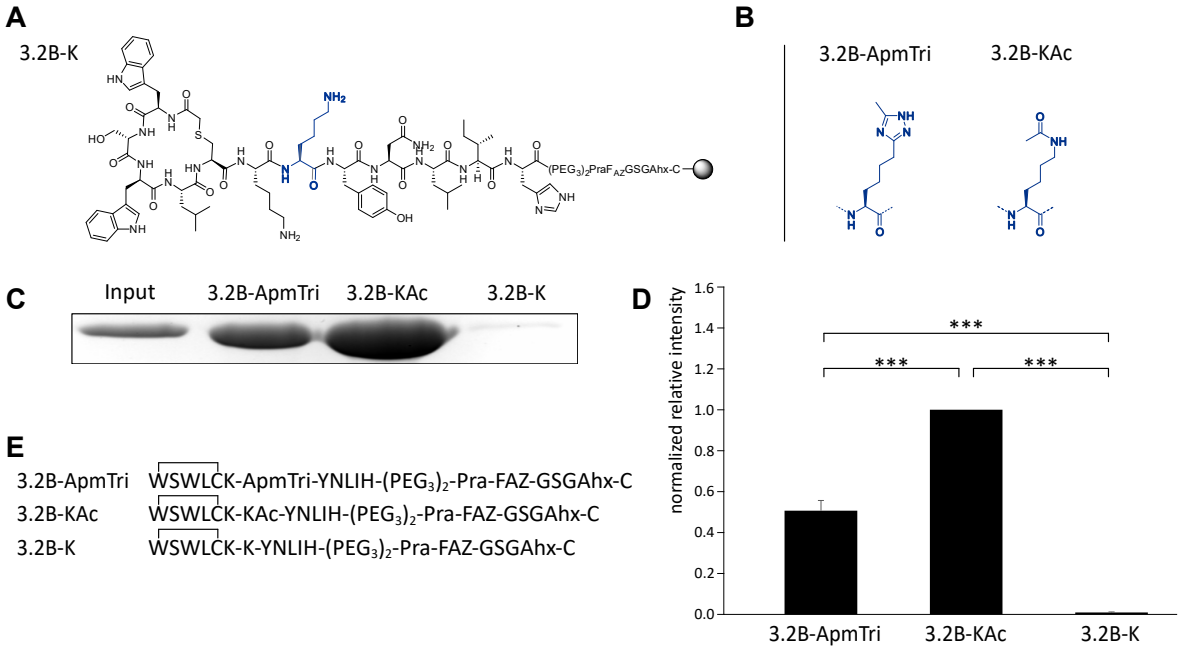


Figure 2.18: Pull-downs of BRD3(2)-TurboYFP with 3.2B-KAc, 3.2B-ApmTri and 3.2B-K. A) Structure of 3.2B-K. B) ApmTri and KAc replacing lysine in the sequence of 3.2B-ApmTri and 3.2B-KAc, respectively. C) Coomassie-stained SDS-PAGE bands of pull-downs with BRD3(2)-TurboYFP and the probes. D) The quantified intensities of the pull-downs were normalized to the intensity of the probe exhibiting the strongest band and are displayed as the mean (n=3). P-values were calculated using ANOVA and Tukey's HSD test for multiple comparisons. *=p<0.05; **=p<0.01; ***=p<0.001; ns=not significant. E) The sequences of the three probes, cyclization is indicated.

2.2 Part 2: Bisubstrate inhibitors for proteomic profiling of N-acetyltransferases

Lys-CoA is an established bisubstrate inhibitor that binds to KATs [179]. By implementing the concept of CoA bisubstrates in a sequence context, Lys-CoA and N-terminal-CoA (N-CoA) probes were designed that showed an enhanced specificity towards individual KATs [179] or NATs [180], respectively. Immobilized N-CoA-peptide conjugates targeted endogenous NATs from cell lysates and allowed for MS/MS assisted interactome studies [194]. The probes showed enrichment of the catalytic and auxiliary units of the NatA, NatB, NatC, NatD and NatE complexes. While the NatE related NAA80 also showed binding of the probes, this interaction was weak when compared to other NATs.

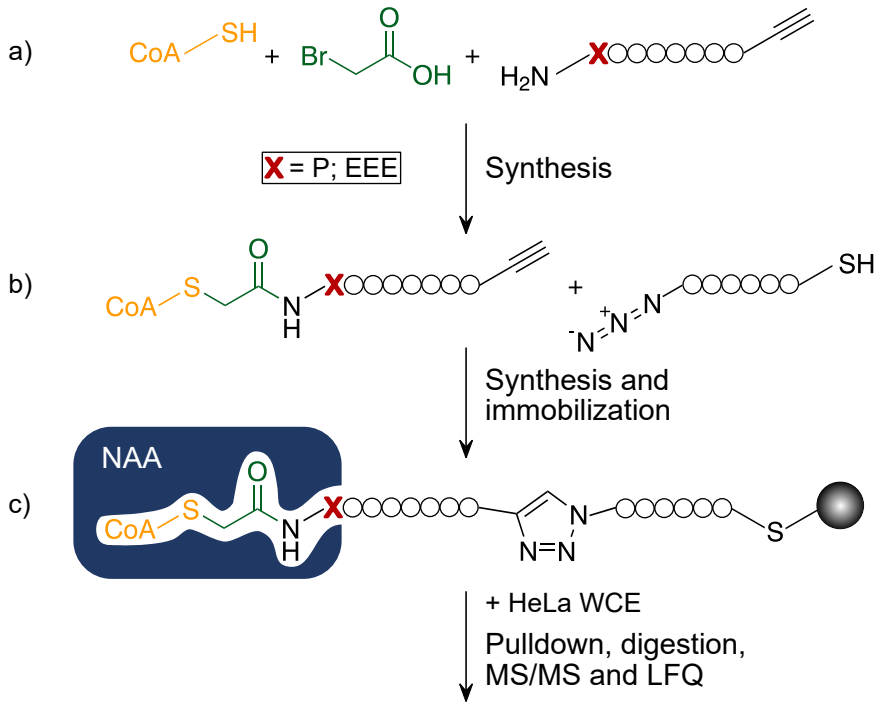


Figure 2.19: Strategy for synthesizing bisubstrate NAT inhibitors and acquiring LFQ data of pull-downs with whole cell extract (WCE). In step a), the peptides bearing either proline or three glutamic acids at the N-terminus (colored red) were linked to CoA (colored yellow) through a bromoacetic acid (colored green). In b), to immobilize the conjugate, it was coupled via CuAAC to a peptide scaffold containing a thiol group, which in turn was linked to functionalized agarose beads. Shown in c), NAAs (colored blue) can bind to the immobilized N-CoA-peptide conjugate, which thus serves as a way to pull-down endogenous NAT complexes.

Human N-α-acetyltransferases show specificity to a wide range of N-terminal amino acids, an exception is proline, which is not acetylated by known NATs [241]. To further elaborate the specificity of bisubstrate N-CoA-peptide probes, mainly focusing on NAA80 and potential, however so far unknown, N-proline acetyltransferases, two N-CoA-peptide conjugates were synthesized in this work and used in interactome profiling.
The strategy of N-CoA-peptide conjugate synthesis and subsequent experiments is displayed in 2.19. The workflow was developed by Dr. Julia Sindlinger and is published together with the results from this part of this thesis in [194]. The pull-down experiments were performed using HeLa WCE and involved the two peptides synthesized in this work, as well as two other peptides with similar structure including CoA-alanine and acetylated alanine at the N-terminus which were synthesized by Dr. Julia Sindlinger (2.11). The eluted proteins were digested and subsequently analyzed by high resolution LC-MS/MS. Label free quantification (LFQ) provided quantitative information about protein enrichment on probes. The data is displayed as volcano plots.

2.2.1 Synthesis of N-CoA-peptide conjugates

Peptides P11, P12 and P16 were synthesized via SPPS using standard Fmoc-strategy, their sequences are shown in 2.11.

Table 2.11: Overview of the peptides for NAT experiments.

Peptide	**Sequence**
P11	PGKAcGG-AhxAhx-Pra
P12	EEEIGKAcGG-AhxAhx-Pra
P16	F_{AZ}-GSG-AhxC
Synthesized by Dr. Julia Sindlinger	
CoA-P13-click	CoA-AGKAcGG-AhxAhx-Pra-F_{AZ}-GSG-AhxC
P14-click	AcAGKAcGG-AhxAhx-Pra-F_{AZ}-GSG-AhxC

The minimal sequence of GKAcGG in every peptide was chosen to focus on the effect of the N-terminal amino acids on the selectivity towards certain NATs. The acetylated lysine in the sequence stems from the probe design being applicable as NAT- as well as KAT-substrates, and originally served as a control that does not bind acetyltransferases.

Before cleavage of the peptides, the N-terminus of P11 and P12 was coupled with bromoacetic acid. Different conditions for the coupling were tested (fig 2.20).

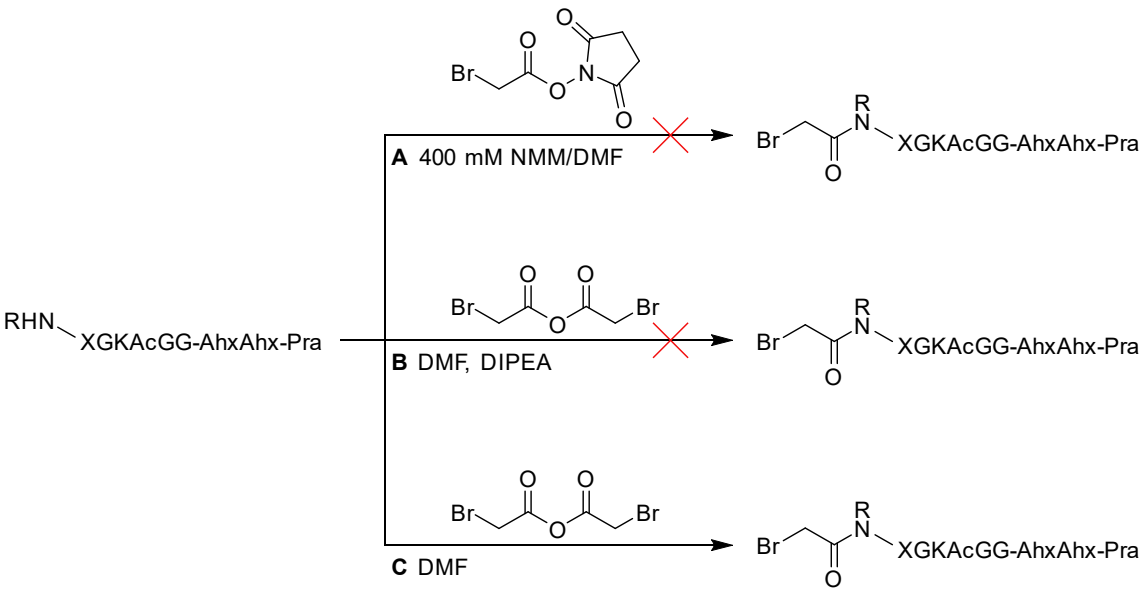


Figure 2.20: Strategies tested for the bromoacetylation of the N-terminus of P11 and P12. Method A) used the bromo acetic *N*-hydroxysuccinimide (NHS) ester which yielded only partial conversion and impurities, method B) used bromoacetic anhydride and DIPEA, which yielded more product but still a large amount of impurities, method C) omitted DIPEA which resulted in full conversion

The first approach A) in fig 2.20 used an NHS activated ester of bromoacetic acid which is attacked by the amino group of the N-terminal amino acid, leading to the bromoacetylated peptide. LC-MS studies revealed small amounts of product and significant impurities (4.16A of the appendix), thus method B) was tested using bromoacetic anhydride and DIPEA as a base. As the LC-MS analyses showed product but again significant amounts of impurities (fig 4.16B of the appendix), the coupling reaction was repeated once with 5× the amount of DIPEA and once without DIPEA, to investigate the influence the base had on the reaction. The amount of impurities significantly increased with the approach using more DIPEA and was relatively low without DIPEA. This lead to method C), where the coupling was conducted using bromoacetic anhydride without any base. As a result, the bromoacetylation of the N-terminus was significantly improved, yielding near complete conversion of the educt to the desired product. MS-data showing the differences in conversion and relative intensity of the impurities are displayed in the appendix (4.16).

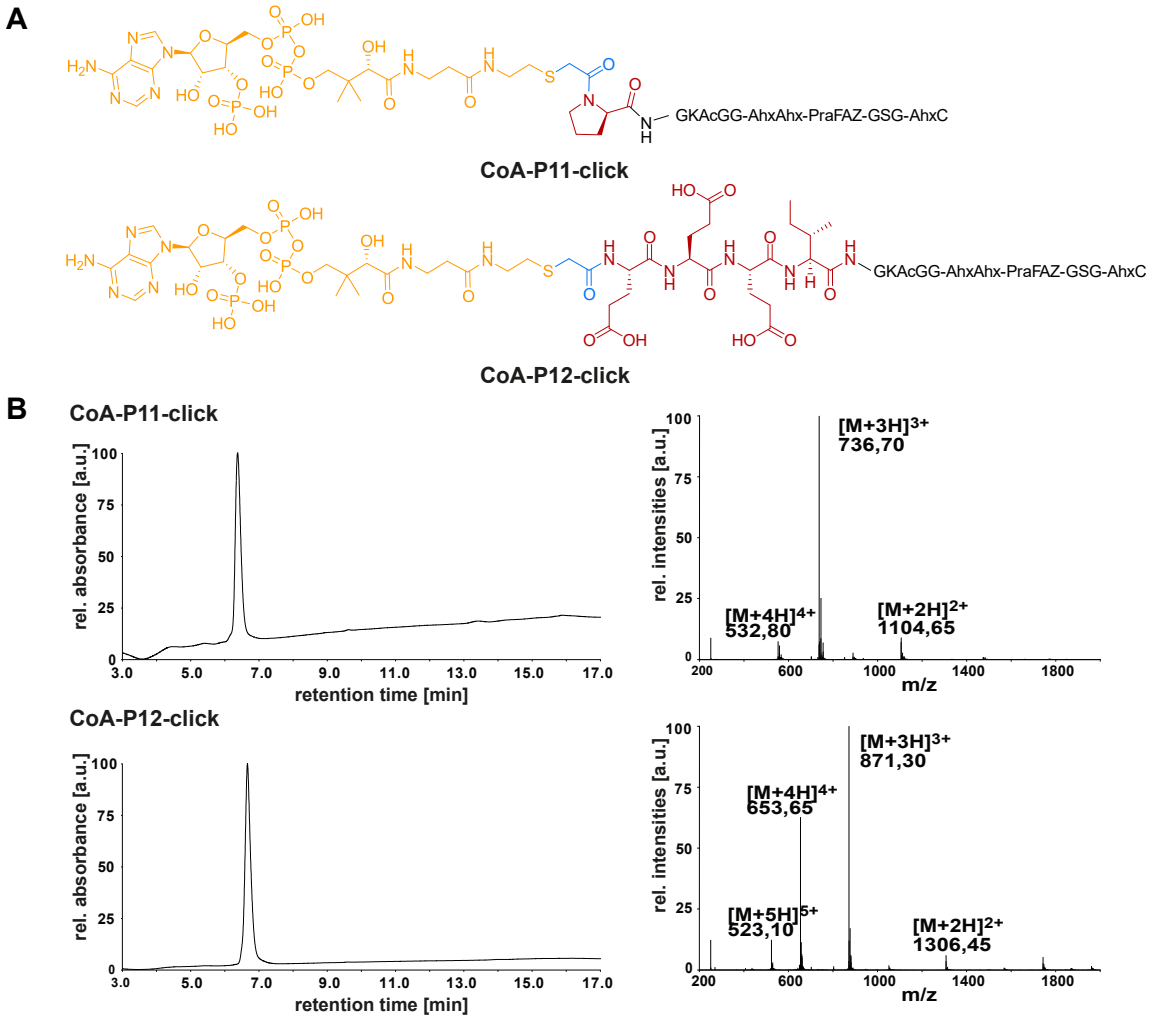


Figure 2.21: Structure and LC-MS analyses of peptides CoA-P11-click and CoA-P12-click. A Chemical structures of CoA-P11-click and CoA-P12-click, the color labeling is the same as in 2.19 with CoA in yellow, the acylic linker in blue and the selectivity inducing N-terminal amino acids in red. **B** UV (at 218 nm) and MS spectra from LC-MS analysis for CoA-P11-click (top) and CoA-P12-click (bottom). Assigned m/z values are displayed in the MS spectrum.

The bromoacetylated peptides P11-BrAc and P12-BrAc as well as P16 were cleaved off the resin using standard cleavage conditions. After purification by HPLC, peptides P11-BrAc and P12-BrAc were coupled with CoA. The CoA-conjugated peptides were repurified using HPLC and afterward separately linked to P16 using CuAAc, resulting in the two final constructs CoA-P11-click and CoA-P12-click.

The necessity for the click reaction was due to the possible interference of thiols with

the CoA coupling. A cysteine in the sequence of P11-BrAc or P12-BrAc would have inevitably led to intramolecular cyclization or multimer formation after the cleavage of the trityl protection group on the thiol, rendering the CoA coupling impossible. That is why the thiol was introduced afterward and therefore still allowed for immobilization on SulfoLink™ beads. The sequence and the analytical data of the purified final products CoA-P11-click and CoA-P12-click are displayed in 2.21. The peptides were eventually linked to iodoacetyl-functionalized agarose beads via their C-terminal cysteine and used for pull-down experiments.

2.2.2 Western blots of different HeLa cell extracts

Different batches of HeLa WCE were tested for homogeneous expression of NAAs and utilized for subsequent interactome profiling by LC-MS. Therefore, a select group of NAA antibodies (NAA10, NAA25, NAA50 and NAA80) was chosen to test for NAA presence and relative quantities in three different HeLa WCEs. NAA10 and NAA50 associate with other catalytic and auxiliary subunits forming the NatA complex that shows specificity towards alanine and serine, among others [104], however free NAA10 was shown to acetylate acidic N-termini [40]. NAA25 is part of the NAT complex NatB that can bind to proteins containing Glx or Asx after the N-terminal methionine [104]. NAA80 has a distinct specificity towards the N-termini of β- and γ-actin [104] which display three consecutive aspartic acids or glutamic acids, respectively. These four NAAs were expected to show specificity towards each of the three probes.

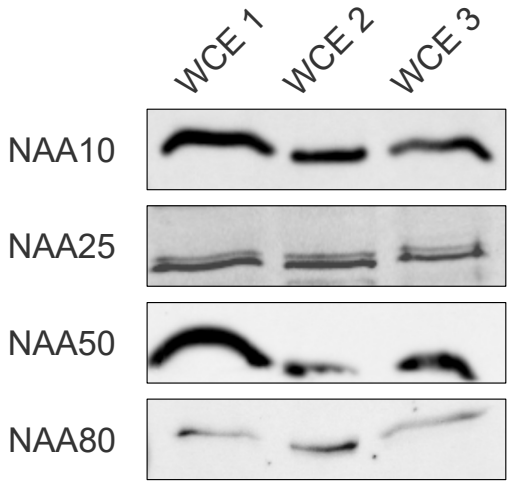


Figure 2.22: Western blot results of WCEs separated via SDS-PAGE and incubated with antibodies against NAA10, NAA25, NAA50 and NAA80.

The HeLa WCEs were separated via SDS-PAGE, blotted onto membranes using the western blotting method, and afterward separately incubated with the four antibody solutions. The developed blots in 2.22 showed presence of all four NAAs in each WCE. Lower abundance of NAA50 and NAA80 was observed in WCE2, on the other hand NAA10 and NAA25 were each present at relatively the same level in all three WCEs. While NAA10 and NAA50 showed very strong bands in WCE1, WCE3 showed intermediate strong bands for all four NAAs. WCE3 was chosen as the extract used in follow-up experiments.

2.2.3 Interactome studies with N-CoA-peptide conjugates

Label-free quantification used in proteomics allows for the straight-forward and fast acquisition of interactome data. Relative intensities between two proteome samples can be determined using the intensity-based MaxLFQ procedure as part of the MaxQuant software [218].

Eluates from pull-down experiments with WCE3 and immobilized probes CoA-P11-click, CoA-P12-click, CoA-P13-click and P14-click were reduced and alkylated with chloroacetamide to prevent disulfide formation. The proteins were then digested using the proteases LysC and trypsin. Resulting peptides were separated on an UHPLC system and subsequently analyzed by high resolution MS/MS. LC-MS/MS experiments as well as raw data processing with MaxQuant were performed by Dr. Jürgen Eirich (group of Prof. Dr. Iris Finkemeier, University of Münster). The -$\log_{10}$ transformed p-value was plotted against $\log_2$ fold enrichment changes of identified proteins from CoA-P11-click and CoA-P12-click against CoA-P13-click and P14-click. The resulting four volcano plots are displayed in 2.23.

When plotting the interactomes of the CoA-conjugates against the N-acetyl controls (B and D in fig 2.23), proteins enriched on CoA-conjugates over acetyl controls are located in the right section of the volcano plot. Similarly, proteins that were enriched stronger on the CoA-conjugates in comparison to the enrichment on the CoA-Ala control are located on the right side of the respective volcano plots (A and C in fig 2.23). Data points of proteins that are significantly enriched are located right and left of the two dashed vertical lines, indicating an enrichment of $\log_2 \geq 0.58$ (equals ≥ 1.5 enrichment). Statistical significant enriched proteins exhibit an alpha level of $p \leq 0.05$ and respective data points are located above the dashed horizontal line at $-\log_{10} p = 1.3$. A total of 3287 proteins were identified in the input, 4239 in P-CoA/A-CoA, 4327 in P-CoA/Ac-A, 4284 in E-CoA/A-CoA and 4307 in E-CoA/Ac-A. Relevant protein hits were grouped depending on their function.

The groups are color-coded and data points were colored accordingly. The structures of the four N-CoA-peptide conjugates used for the interactome studies are displayed in A). The Ac-A probe missing the N-terminal Coenzyme A served as a control, while the A-CoA probe was a bisubstrate inhibitor for several NATs, thereby acting as specificity deducing comparison for P-CoA and E-CoA.

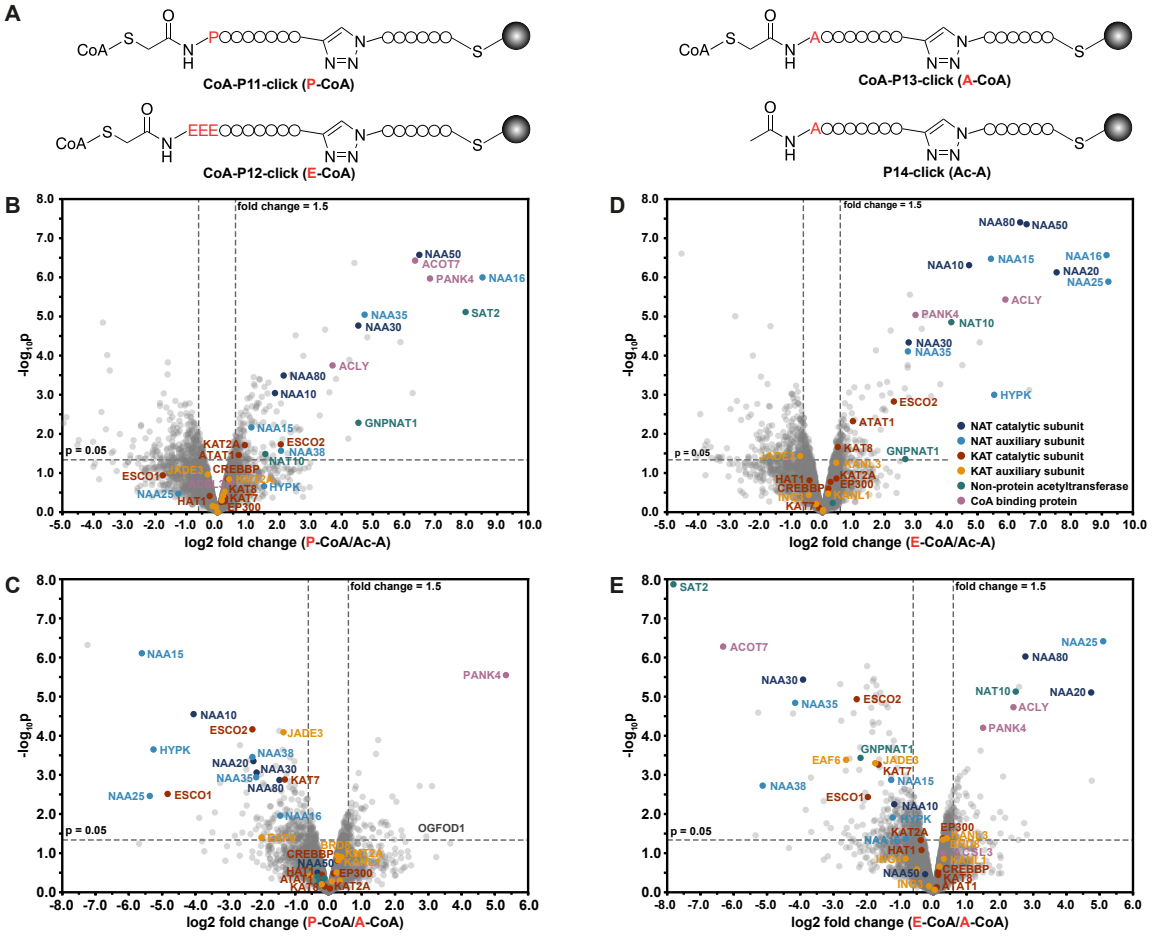


Figure 2.23: Probe structures and volcano plots of pull-downs with CoA-P11-click, CoA-P12-click, CoA-P13-click and P14-click. A) Structures of all four probes used in the pull-down and subsequent mass experiments. B) P-CoA versus Ac-A. C) P-CoA versus A-CoA control. D) E-CoA versus Ac-A. E) E-CoA versus A-CoA control.

The interactome B) of P-CoA versus the Ac-A control shows enrichment of the main catalytic and auxiliary NATs as well as some catalytic KATs. Several non-protein acetyltransferases and CoA binding proteins were also found enriched on P-CoA. However, in the volcano plot C) it becomes apparent that the quantitative enrichment, especially of NATs and KATs, is much lower on P-CoA than on A-CoA. While almost all NATs, KATs

and non-protein acetyltransferases were depleted in interactome C), the CoA binding protein pantothenate kinase like protein (PANK4) stood out as strongly enriched on P-CoA. Another protein still enriched on the P-CoA probe vs. A-CoA was 2-oxoglutarate and Fe(II)-dependent oxygenase domain-containing protein 1 (OGFOD1).

All catalytic and auxiliary NATs found are enriched on the E-CoA probe in comparison to the Ac-A control. Particularly noteworthy is NAA25, which was the only auxiliary NAT showing no enrichment on P-CoA in the volcano plot B), but is the strongest enriched protein found on E-CoA in D). Some of the catalytic KATs that have been enriched on P-CoA vs. Ac-A also show enrichment on E-CoA vs. Ac-A, though the enrichment is lower in comparison to NAT associated proteins in both interactomes. The enrichment of non-protein acetyltransferases and CoA-binding proteins on E-CoA vs. Ac-A shows a similar interaction pattern as in P-CoA vs. Ac-A, the major exception being the cytosolic acyl coenzyme A thioester hydrolase (ACOT7). ACOT7 was highly enriched on P-CoA vs. Ac-A, while there was no enrichment found on E-CoA vs. Ac-A. The comparison of E-CoA with A-CoA in the volcano plot E) even showed a high relative enrichment of ACOT7 on the alanine probe. Volcano plot E) shows depletion of all NAT proteins enriched in the interactome D) except for NAA20, NAA25 and NAA80 which still were strongly recruited to E-CoA. PANK4 also shows enrichment on E-CoA in E), although the enrichment is less pronounced as on P-CoA vs. A-CoA in C). The ATP citrate lyase (ACLY) and the N-acetyltransferase 10 (NAT10) are enriched on E-CoA vs. A-CoA in contrast to P-CoA, where the low enrichment observed vs. the Ac-A control was depleted after comparing it with A-CoA.

NAA10 as part of the NatA complex exhibits specificity towards N-termini starting with alanine and serine, however free NAA10 has been found to acetylate acidic N-termini [40]. To investigate whether the CoA-peptide conjugates can also detect this shift in substrate specificity, pull-downs with recombinant expressed NAA10-TurboYFP and peptides E-CoA, A-CoA and Ac-A were conducted. The assays are displayed in figure 2.24. Interaction between NAA10 and A-CoA as well as Ac-A is weak, while it was found to be enriched on the P12-probe. This is in line with expected specificity, implicating that the NAA10 in the cell lysate is mostly localized in the NatA complex.

The experiments showed that the overall broad initial enrichment on both of the CoA probes vs. the Ac-A control did not allow simple readout of specificity. Only the interactomes that compared the enrichment on both probes with the enrichment on A-CoA deduced specificity in a straightforward way. The numerical $\log_2$ fold enrichment of all relevant proteins is displayed in 2.12.

Table 2.12: Overview of selected proteins and their log_2 fold enrichment in all four interactomes. Proteins were grouped according to their function and their log_2 fold values for each experiment are reported. Bold log_2 fold changes indicate significant enrichment ($log_2 \geq 0.58$) as well as statistical significance ($p \leq 0.05$).

UniProt ID	Protein	log_2 fold change			
		P-CoA vs. Ac-A	P-CoA vs. A-CoA	E-CoA vs. Ac-A	E-CoA vs. A-CoA
NAT catalytic subunit					
P41227	N-α-acetyltransferase 10	**1.85**	**-4.06**	**4.72**	**-1.18**
P61599	N-α-acetyltransferase 20	-	**-2.26**	**7.54**	**4.74**
Q147X3	N-α-acetyltransferase 30	**4.53**	**-2.16**	**2.78**	**-3.92**
Q9GZZ1	N-α-acetyltransferase 50	**6.49**	-0.33	**6.58**	-0.25
Q93015	N-α-acetyltransferase 80	**2.13**	**-1.48**	**6.37**	**2.67**
NAT auxiliary subunit					
Q9BXJ9	N-α-acetyltransferase 15	**1.09**	**-5.61**	**5.43**	**-1.27**
Q6N069	N-α-acetyltransferase 16	**8.53**	**-1.45**	**9.15**	**-0.83**
Q14CX7	N-α-acetyltransferase 25	-1.27	**-5.37**	**9.21**	**5.10**
Q5VZE5	N-α-acetyltransferase 35	**4.73**	**-2.17**	**2.75**	**-4.15**
Q9NX55	Huntingtin-interacting protein K	1.50	**-5.26**	**5.53**	**-1.22**
KAT catalytic subunit					
Q5SQI0	α-Tubulin N-acetyltransferase 1	**0.68**	-0.20	**0.98**	0.10
Q5FWF5	N-acetyltransferase ESCO1	-1.77	**-4.83**	-	**-1.97**
Q56NI9	N-acetyltransferase ESCO2	**2.03**	**-2.29**	**2.30**	**-2.30**
Q92830	Histone acetyltransferase KAT2A	**0.88**	0.05	0.45	**-0.38**
O95251	Histone acetyltransferase KAT7	0.23	**-1.31**	-0.11	**-1.65**

Non-protein acetyltransferase					
Q96EK6	Glucosamine 6-phosphate N-acetyltransferase	**4.53**	-0.33	**2.67**	**-2.19**
Q9H0A0	RNA cytidine acetyltransferase	**1.53**	-0.29	**4.15**	**2.47**
Q96F10	Thialysine N-ε-acetyltransferase	**7.99**	-0.13	0.33	**-7.81**
CoA-binding protein					
P53396	ATP citrate lyase	**3.70**	0.21	**5.89**	**2.41**
O00154	Cytosolic acyl coenzyme A thioester hydrolase	**6.36**	0.20	-0.16	**-6.32**
Q9NVE7	4'-phosphopantetheine phosphatase	**6.84**	**5.33**	**3.00**	**1.49**
Prolyl hydroxylase					
Q8N543	Prolyl 3-hydroxylase OGFOD1	0.16	**2.94**	0.01	**2.56**

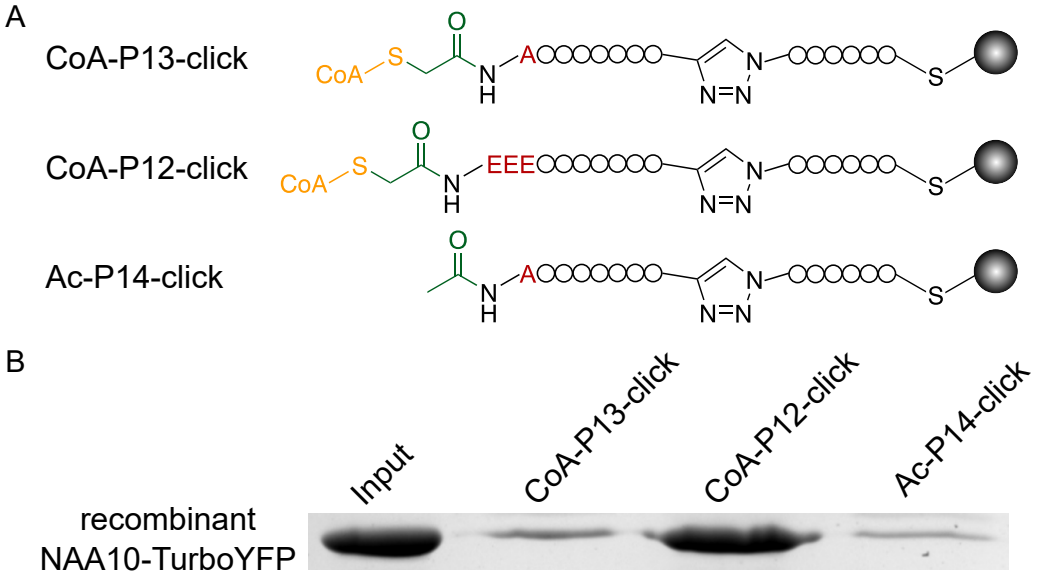


Figure 2.24: SDS-PAGE analysis of pull-down conducted with recombinant NAA10-TurboYFP and the peptide probes CoA-P13-click, CoA-P12-click and Ac-P14-click. A) Structure of the immobilized peptide probes. B) SDS-PAGE gel showing the bands of the inputs as well as the eluted pull-down samples.

3 Discussion

Protein acetylation plays an important role in cell homeostasis and disease development. Nt-acetylation is introduced by mono- or multisubunit NATs which possesses inherent substrate specificity with regard to N-terminal amino acids of proteins. Using the same co-substrate as NATs, KATs catalyze the acetylation of lysines. The common co-substrate acetyl-CoA was used to create the bisubstrate inhibitor Lys-CoA. Applying the same design concept to conjugate the N-terminus of peptides with CoA resulted in probes targeting NATs. Analysis of changes in abundances of individual NATs and KATs, which can be helpful in understanding health and disease models, would require specific probes coupled with a technique to quantify and simultaneously identify bound proteins.
Lysine acetylation is one of the PTMs employed by eukaryotes to govern gene regulation. The installation of an acetyl group on lysines located on histone tails directly affects chromatin accessibility and thereby regulates gene expression. Not only does the charge neutralization of the acetylated lysine lead to a more loosely packed chromatin, bromodomain-containing proteins can bind to acetyllysine and recruit different chromatin remodeling factors that in turn also influence transcription levels. Aberrant bromodomain recruitment is found in different types of cancer, with the NUT midline carcinoma being a prominent example which is caused by the oncogenic fusion-proteins BRD3-NUT and BRD4-NUT. Inhibiting Brd recruitment to their target site is of great interest, leading to the development of different BRD- and especially BET inhibitors. Most reported small molecule BET inhibitors however do not discriminate between individual members of the BET family, likely contributing to the high toxicity observed in pre-clinical trials. Peptides present a new potential class of BET inhibitors that might be developed to possess high affinity and high specificity combined with low toxicity.

3.1 Part 1: Development of ApmTri containing peptide substrates for BET proteins

While some acetylated histone-derived peptides are bound by multiple BET bromodomains, other acetylated sequences on the histone tails are only bound by single bromodomains [66] which might be an indicator that they have different functions in the cell [99]. Interactions between the amino acids flanking the acetyllysine and the loop regions of the bromodomain, which are not conserved between members of the BET subfamily [65], contribute to the

different substrate specificities. Small molecular inhibitors like JQ1 show high affinity to BET proteins but have limited capabilities to interact with the specificity-inducing areas of the bromodomains due to their small size and targeting of the binding pocket, resulting in low specificity within the BET subfamily. To address the need of high-affinity and selective BET inhibitors in research and medical applications, a peptide-based histone-derived inhibitor was optimized and stabilized using combinatorial libraries. Incorporating the stable acetyllysine mimic ApmTri [196], the intrinsic deacetylase sensitivity of acetylated lysines was circumvented. Additionally, the amino acids flanking the central ApmTri potentially allow for specificity towards single bromodomains.

3.1.1 Library experiments

Native substrate sequences were used as the starting peptides to determine the minimum length of the peptides at which binding was still deemed sufficient. This step reduced the size of the library in the subsequent step. Following library experiments aimed at improving the strength of interaction between the Brds and the ligands as well as stabilizing the peptides for *in vivo* applications. While the choice of the mono-acetylated sequence H4K20ac as a starting point for optimizing the binding motif of the BD2 domain was straightforward based on reported affinity [68] and medium sequence length (11 amino acids), the choice of the starting sequence for binding motifs of BRD3(1) and BRD4(1) H4K5/8ac was more difficult. In [196] it was shown that replacing acetyllysine with ApmTri in a mono-acetylated sequence led to a 2-fold increase of the K_D-value while the replacement of both acetylated lysines in a di-acetylated sequence with ApmTri increased the K_D-value approximately four times. Choosing a poly-acetylated sequence showing an initial high affinity in which more than two acetyllysines had to be replaced by ApmTri was thought to lower the binding even more, leading to the selection of not the sequence displaying the lowest K_D-value in [68] but the least number of acetyllysines paired with a still high affinity. A search for mono-acetylated sequences for BRD4(1) was conducted by Sebastian Kalkuhl under my supervision in his B.Sc. thesis. However, the sequence with the highest enrichment of BRD4(1) amongst the tested mono-acetylated sequences still interacted weaker with BRD4(1) than the H4K5/8ac sequence (mean band intensity of 4:1). These findings led to the choice of the H4K5/8ac sequence as the starting sequence for the development of peptide inhibitors of the BD1 domains.

The first and the last two amino acids of the sequence AKRHR-KAc-VLRDN (H4K20ac) for BD2 were truncated, leaving the sequence RHR-KAc-VLR (sH4K20ac) without major

reduction in the binding assay with the BD2 proteins as seen in fig 2.7. Surprisingly, the removal of aspartic acid four residues downstream of the KAc site even lead to a strong increase in binding in comparison to the native sequence. Aspartic acid is negatively charged in the cellular environment (pKa = 3.90 [242]). It is reasonable to assume that by truncating aspartic acid, electrostatic repulsion that were present before are lost, leading to stronger binding between peptide and Brd. However, this cannot be confirmed by the structure of BRD3(2) in complex with H4K20ApmTri (PDB: 8B5A), since the aspartic acid is not close to the protein surface. SGRG-Kac-GG-Kac-GLGY (H4K5/8ac) as the starting sequence for the optimization of BD1 binders was only shortened to the sequence GRG-Kac-GG-Kac-GLG (sH4K5/8ac) while also accepting a minor reduction in affinity. The gradual decline in fluorescence intensity with continuing truncation observed in the binding assays is coherent with the assumption that all amino acids of the native sequence (H4K5/8ac) contribute to the binding, thus removing them lead to weaker attraction between Brd and substrate peptide.

Many design methods for libraries such as mRNA- and phage display allow *de novo* discovery of binding sequences, omitting the steps of truncation employed for libraries starting from natural substrates (fig 1.7). However, the truncation step was necessary for the display method used in this project, as it is limited in the number of residues randomized in the design process. The coupling efficiency of the automated SPOT synthesis decreased after ~15 amino acids, which limits the length of the peptide. Furthermore, the synthesis prohibited length variation for peptides synthesized in one single experiment which required the determination of a suitable length beforehand. Thus, a fixed peptide length determined in the truncation experiments was used for probing BRD3 and BRD4 binding with the SPOT synthesis. Consequently, the next step of affinity mutant screening is not a true *de novo* screening but rather an optimization of the native sequence.

Several peptides were obtained from the affinity mutant screening experiments that show strong binding of the respective bromodomains, as seen in fig 2.10. Pull-downs with affinity mutants on membranes for BRD4(1) were first conducted over 1 h at room temperature, the same condition as for the pull-downs with the BD2 proteins. The results however showed strong preference for arginine-rich sequence, again hinting towards unspecific binding. After extending the incubation period to overnight at 4 °C, the results were more defined and good binding was observed for the native sequence, implying a better pull-down condition. These results can be found in the appendix in figures 4.7, 4.8 and 4.9. Conditions for the pull-down done on resin were conducted at room temperature for 1 h, which might explain the relatively weak binding of the optimized sequences in comparison

to the native sequence seen in fig 4.10. Additionally, the aforementioned weaker enrichment of BRD4(1) on H4K5/8ApmTri in comparison to H4K5/8ac most likely also contributes to the weaker interaction [196].

From a library of 19^6 = 47,045,881 different peptides for each of the BDs, 5 sequences for BD1 domains and 5 sequences for BD2 domains were selected for the following stabilizing experiments. With a sequence diversity of almost 5×10^7, the library method employed is rather small in comparison to cellular approaches like phage display (1×10^9), cell-surface display methods ($1 \times 10^{8\text{-}10}$) and especially acellular approaches like mRNA display that theoretically offer up to 1×10^{15} individual peptides [200]. However, cellular and acellular libraries are hampered by the limitation of incorporating only natural amino acids. This limitation does not apply to SPOT libraries. While non-natural amino acids have been introduced to cellular and cell-free approaches mainly by reassigning the codons like the amber stop codon UAG [243, 244], they are often restricted to the introduction of a single non-natural amino acid to the library, making them impractical for stability screenings including different N-methylated and D-amino acids.

Random nonstandard Peptide Integrated Discovery (RaPID) is a display method developed in recent years which offers huge library sizes paired with the ability to reprogram the entire 64 codons of the genetic code with the usage of so-called flexizymes [245]. This method could have been applied to this project, but it comes with high expenditure of materials and requires specific knowledge for its usage. The applicability is underlined by the publication of the work of Patel et al. [195] where RaPID was used to discover cyclic peptide-based inhibitors for the BET proteins BRD2, BRD3 and BRD4 that showed K_D-values as low as 100 pM and specificities to individual BDs of up to 10^6. The considerable library size of RaPID enabled the selection of high affinity and selective BRD3 and BRD4 binding sequences, but without focus on *in vivo* stability, the peptides do not contain N-methylated and D-amino acids. The cyclic nature of the peptides will provide enhanced resistance against proteases, but cell permeability and oral availability would require further adaptations. Moreover, sensitivity against deacetylases remains an issue, as the sequences incorporate acetyllysine in contrast to a stable KAc mimic like ApmTri employed in this work, thereby prohibiting the application of the inhibitors *in vivo*. It was shown that incorporation of ApmTri into one of the sequences lead to binding of BRD3(2) 2.18, binding to BRD4(1) was however greatly diminished when acetyllysine was substituted with ApmTri 4.14. This leads to the conclusion that the peptides developed in [195] could be explored for strong binding and selectivity after exchanging KAc with ApmTri. However, it is likely that not all BET bromodomains can be addressed with this approach.

The optimized affinity mutants for BRD3(2) and BRD4(2) were subjected to substitution with N-methylated and D-amino acids, resulting in the sequences bearing two N-methylated and D-amino acids without major changes in the interaction profile 2.14. There are two main approaches of stabilizing peptides against proteases: cyclization and substitution of amino acids with their N-methylated and D-derivatives. N-methylation as well as D-amino acids have been shown to convey resistance against metabolic enzymes and can be found in several pharmacological active natural peptides like cyclosporin and bouvardin [246]. Additionally, D-amino acids are rarely recognized by endogenous proteases [247]. Cyclisation on the other hand offers an intrinsic resistance against exopeptidases, provided it involves the N- and the C-terminus, while the rigidity that comes with cyclisation also renders endopeptidase cleavage more challenging [248]. At the start of this project, two peptide designs for *in vivo* studies were discussed, one involving a linear construct that would require an additional step of stabilizing by implementing N-methylated and D-amino acids, and one involving a cyclic construct. Both designs are displayed in fig 3.1 Eventually, the linear construct was favored as initial experiments showed that the membrane on which the SPOT synthesis was conducted does not offer great potential for straightforward creation of cyclic libraries.

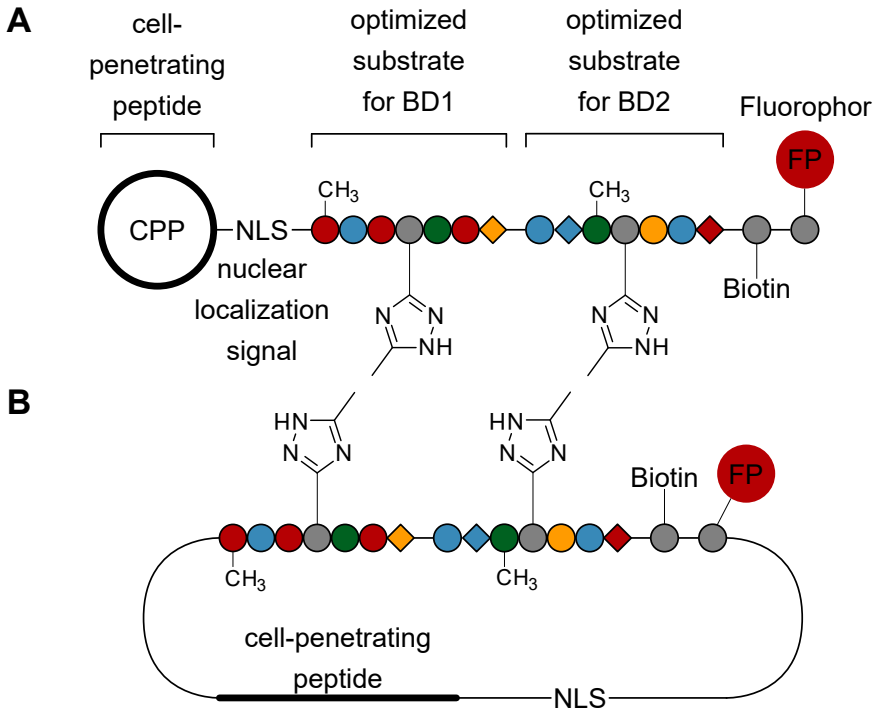


Figure 3.1: Linear and cyclic design of final peptide. A) The final peptide design as discussed in 2.1.1. B) A macrocyclic peptide design.

A drawback of linear peptides is that conformational flexibility commonly lowers the affinity and flexibility towards the binding partner due to entropic effects [246], but the bivalent nature of the final inhibitor is thought to counteract this loss in affinity. Comparing linear and macrocyclic compound libraries directly has not been a major focus of research yet, but the work of Gao et al. [249] suggests that cyclic compound libraries achieve stronger interactions than their linear counterpart.

3.1.2 Investigation of linker length

In order to investigate the effects of the linker length between the substrate sequences of BD1 and BD2 in the bivalent peptide ligand, three different lengths were tested in pull-down experiments. The two Brds in BRD3 and BRD4 each bind to sequences containing acetylated lysines or ApmTri. Consequently, avidity effect were expected in the pull-downs with both substrate sequences combined to one polypeptide chain. Avidity is the combined strength of multiple affinities of singular non-covalent interactions, however it is more than just the sum of these affinities. When two reactants are held in close proximity by an enzyme or due to a previous binding event of one site of a bivalent binder, their so-called effective molarity (EM) is increased [250]. The effective molarity can be expressed as the ratio between the rate constants of the intramolecular (first order kinetics) and the corresponding intermolecular reaction (second order kinetics):

$$EM = \frac{k_{intra}(s^{-1})}{k_{inter}(l \cdot mol^{-1} \cdot s^{-1})}$$

Binding of one bromodomain to the respective substrate sequence would therefore increase the EM of the second binding sequence, thus facilitating the second binding event. This cooperative effect also depends on the molar ratio of protein and ligand, the nature of the linker and if the linker length is able to assist or prohibit the binding [251]. The results of the pull-downs with different linker lengths give no definitive conclusion on avidity effects since all enable Brd binding with only subtle differences. As seen in 2.15, the enrichment of BRD3 and BRD4 on peptide P32-click with one PEG linker acting as a spacer was the strongest. However, quantifying the signal intensities of the pull-downs indicates that there is no significant difference in the interaction profile.

Pull-downs are a fast method to analyze the binding between a protein and a substrate. One of their disadvantages is that the result depends on the k_{off} and is not influenced by the k_{on}. However, avidity is expected to have an effect on the k_{on} because if one Brd already bound to the substrate sequence, the other binding event would be supported and

the k_{on} would be greater. For testing and especially for quantifying avidity effects, other methods would need to be employed.
The linker can influence the EM of the second binding site on the substrate peptide, and thereby the avidity between BRD protein and inhibitor, in many different ways. Interactions between the linker and the protein surface, rigidity and above all linker length must be taken into account. A study investigating the effect of PEG linkers with differing length between a substrate and an enzyme on the EM revealed that if the linker is too short, the EM is the lowest while it is unsurprisingly the highest when the linker length is optimal [252]. If the linker is too long, however, it was observed that the EM did not decrease as abruptly as expected as displayed in fig 3.2.

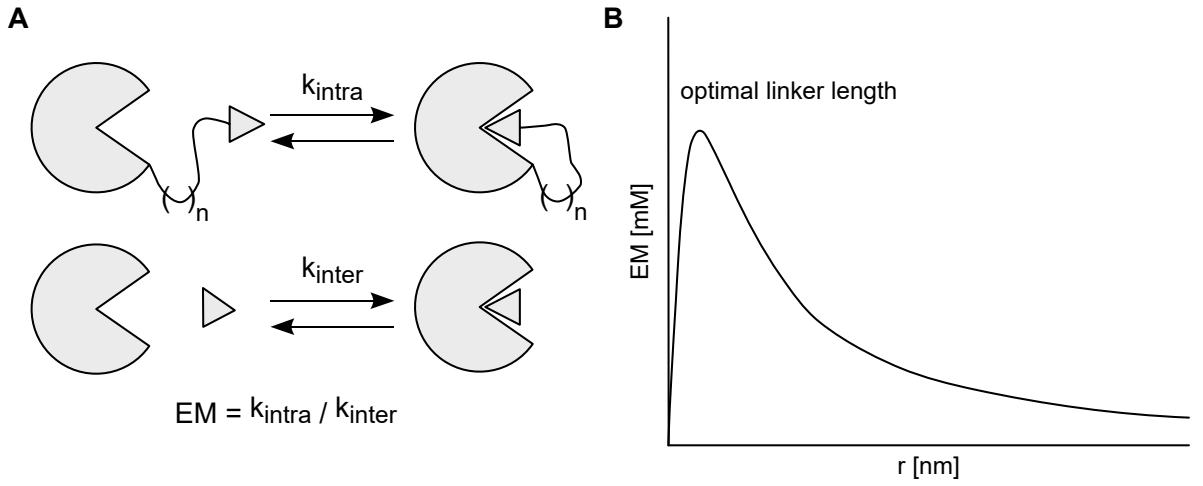


Figure 3.2: Effective molarity as a fraction of the reaction rate constants. A) Depending on the circumstances, the same reactants can interact in an intra- or intermolecular fashion, affecting the effective molarity. B) EM is highest if the linker length is optimal, but it does not decrease as sharp when the length of a flexible linker is too long. Adapted from [252].

This would lead to the conclusion that while scanning for an optimal linker length, a flexible linker that is longer than the distance between the binding sites should be chosen as a starting point. In this work, the distance between both bindings sites, BD1 and BD2 in each of the two bromodomain proteins, was not known for binding in its native state. However, AlphaFold predicts a distance of 34 Å (AF-O60885-F1) between the two conserved asparagines Asn140 and Asn433 in BRD4, while the longest linker length tested is around 31 Å in its extended form [253]. Although the confidence of the unstructured regions determining the distance between BD1 and BD2 is very low, the comparison

indicates that the spacer between the two substrate sequences tested could have been too short to observe avidity effects.

3.1.3 Cellular permeability and localization

Since the two target proteins in this work, BRD3 and BRD4 as well as their oncogenic counterparts BRD3-NUT and BRD4-NUT, are located in the cell nucleus, the peptide-based inhibitors had to be optimized for the cellular uptake and the nuclear localization. For the former task, five cell-penetrating peptides were successfully synthesized and purified, which can be readily linked to the final bivalent stable inhibitor via a disulfide bond. Based on their concentration, the uptake mechanism of these five cationic CPPs can vary [235]. In high concentrations, cationic peptides penetrate the cell-membrane directly in an energy-independent fashion by first attaching to negative components of the membrane surface, which is followed by membrane destabilization due to pore formation or membrane thinning [235]. Severe limitations accompany the second endocytosis-dependent uptake mode: the CPP and its cargo could be trapped in the endosomes and do not reach the cytoplasm or other organelles [238]. Direct penetration of the cell membrane is thus favored. Hackenberger and co-workers proved the delivery of CPP-bound cargo to the cell using the cyclic Poly-D/L-Arg and the linear Poly-Arg also synthesized in this work [237, 238]. The uptake of the CPP-conjugates was efficient but decreased with increasing size of the cargo. However, addition of the unbound linear Poly-Arg improved cell-penetration of CPP-conjugates in comparison to a control group with no additives [237]. It is suspected that the unbound Poly-Arg weakens the cellular membrane supported by covalent binding via its free thiol moiety (sequence shown in table 2.7), creating nucleation areas on the cell membrane which allow for effective delivery of CPP-cargo conjugates into the cell [237]. Another benefit of this method is the lower concentration of the CPP-conjugate needed for efficient cellular uptake, alleviating the toxicity of the molecules. The same type of CPP was considered in this project [237].

It has been shown that basic peptide sequences incorporating a high number of arginines and/or lysines tend to accumulate at the nucleolus [254], a compartment of the eukaryotic nucleus which is the site for ribosome biogenesis. This was also found in earlier studies [237, 238], in which not only cell-penetration but also crossing of the nuclear membrane and localization in the nucleolus was reported for two different CPPs. Generally, nuclear localization signals (NLS) and nucleolar localization signals (NoLS) seem to some degree overlap [254, 255]. However, targeting and enrichment especially of therapeutic CPP-

peptide conjugates at the nucleolus could be unfavourable as their molecular target might be localized in a different region of the nucleus. To circumvent the immediate nucleolar localization of the BRD-targeting inhibitor developed in this work, the peptide was designed in a way that the N-terminal CPP is conjugated via a disulfide bond that can be cleaved in the reducing environment of the cytoplasm [256]. Additionally, some CPPs do not serve as NLS at the same time, so a further NLS sequence is required. Thus, after the permeation of the cell membrane the CPP is released, leaving a nuclear localization signal at the N-terminus that mediates entry into the nucleus. As mentioned before, the favored membrane penetration mechanism of CPPs at low concentrations is the endocytosis-mediated translocation, which often leads to the peptides being trapped in endosomes [257]. Finding the optimal way for the BRD-inhibitors to reach their nuclear target therefore remains a challenge that most likely requires different approaches to be tested. It should also be noted that when administered to an organism, the disulfide bond must be able to withstand reduction in the blood plasma while at the same time it must be reducible in the cytoplasm. Because the concentrations of free cysteine as the most abundant free thiol in blood plasma is a thousand times lower than the primary reducing agent glutathione in the cytoplasm [258], this challenge can be overcome. However, the steric hindrance of the disulfide bond has to be in a certain range to make it resistant enough against reduction in the blood plasma but labile enough for reduction in the cytoplasm.

3.1.4 Suitability of final peptide construct for *in vivo* applications

Providing BRD peptide-based inhibitor displaying sufficient affinity which are suited for *in vivo* studies and, ultimately, clinical application, was the aim of this work. As discussed in the introduction (chapter 1.4.2), most peptides bear significant drawbacks in terms of stability and bioavailability that make them initially less favorable for clinical application. These disadvantages however can be addressed, which lead to the final peptide design displayed in figure 2.2. To enhance the resistance of the construct against endoproteases, amino acids were substituted with N-methylated and D-derivatives, which is discussed in chapter 3.1.1. Vulnerability of the N-terminus of the construct towards exoproteases should be reduced when cyclic or D-amino acids-containing CPPs are conjugated to the peptide in an N-terminal fashion and the C-terminus should provide enhanced resistance if a D-prolin is present [259]. These adaptations render the construct better suited for *in vivo* applications in research, but it is only an initial step towards medical application.

Most often, the preferable route of drug administration is the oral route because it is convenient for the patient. Peptides however generally display low oral availability due to metabolic enzymes and biological barriers present in the gastrointestinal system [260]. For developing biological active and orally available peptidic drugs, two general strategies have been proposed [261]: As a first method, the structure of the peptide is determined based on its affinity towards its target, after which the structure is optimized for oral availability, retaining the affinity. In the second strategy, a proteolytically stable scaffold fitted for oral administration is developed which can be functionalized to achieve biological activity. The first strategy was considered in this work, improving overall proteolytic stability and oral availability by incorporation of N-methylated and D-amino acids to a sequence that shows high bioactivity. N-methylated amino acids can increase oral bioavailability, but no universal rule for this effect has been described as of yet [261]. Length and cyclization status of a peptide also contributes to its oral availability, as shorter and cyclic peptides are more enzymatically stable whereas more degradation is observed for longer and linear peptides [193]. A good example of cyclic and N-methylated peptides found in nature is cyclosporin A, a hepta-N-methylated cyclic peptide composed of 11 amino acids that displays oral availability between 15 and 50 %, depending on the pharmaceutical formulation. The delivery system of a peptide drug generally can have a major influence on its oral availability and the application of suitable systems like lipid-based, polysaccharide-based and inorganic particles could improve the oral application of the peptide-based inhibitor developed in this work [260]. If the final bivalent stable ligand presented in this work is stable enough to be administered orally is questionable since only a few N-methylated amino acids could be incorporated as the only means that typically enhance oral availability of peptides. Parenteral administration would be favorable as the peptide would not have to be optimized in terms of withstanding the major proteases encountered in the gastrointestinal tract as well as traversing cellular and mucus barriers as optimization of stability can diminish affinity as seen in 2.1.6.

If peptides are used in *in vivo* experiments, it is crucial to determine toxicity with wet lab and/or *in silico* methods. Computer-based methods predict toxicity of peptides by sequence comparison or by employing machine-learning models. However, most in silico methods like ToxIBTL [262] and ToxinPred [263] are limited to canonical amino acids as inputs and would not be able to predict toxicity of the peptides in this work. Wet lab assays like the lactate dehydrogenase leakage assay [264] or the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay [265] could be applied to analyze the toxicity of the bivalent stable ligand. These assays measure the integrity of the cellular

membrane as a marker for cell viability and cytotoxicity, as many toxic peptides disrupt cell membranes [266]. Although the toxicity of the bivalent ligand is hard to predict, it is likely that the CPP conjugated to its N-terminus could lead to toxic effects by compromising the integrity of the cellular membrane. Depending on their chemical nature and cellular uptake process, CPPs can weaken cellular membranes and lead to leakage of protons, proteins and metal ions [267]. This effect however, can be addressed by adjusting the concentration of the peptide [268]. Lastly, most CPPs are not cell type specific [268]. The CPP-conjugated bivalent stable ligand could therefore lead to reduction of BRD3 and BRD4 recruitment in healthy tissue, which in turn could lead to toxic effects as well.
The final bivalent stable ligand design provides means of cell permeation and nuclear uptake and displays some modifications that enhance resistance towards proteolytic enzymes. A biotin moiety and a fluorophore located at the C-terminus should enable *in vivo* studies with live cell Förster resonance energy transfer or degradation assays. In summary, the activity of the bivalent ligand in the current state in whole organisms is unlikely and will most likely require further optimization steps due to unknown metabolic stability and toxicity.

3.2 Part 2: Bivalent inhibitors for proteomic profiling of N-acetyltransferases

Determination of NAT/KAT levels and activities in cell cultures and patient-derived disease tissues could improve the understanding of the involvements and functions of these protein families in diseases. A method for addressing this task was used in this work focused on substrate specificity of NATs in cell lysates towards Nt-CoA-peptide conjugates. Label-free quantification employing MS/MS analysis was used to quantify the interactome of each probe. Two probes bearing proline or three consecutive glutamic acids at the N-terminus as well as a control probe with N-terminal alanine and probe with an acetylated N-terminus were synthesized and immobilized. Eluates from pull-downs with cell lysates were digested, purified and ultimately analyzed via MS/MS spectrometry. Resulting data were analyzed and protein enrichment on one probe vs. enrichment on another probe was plotted against the $\log_p$-value indicating statistical significance as volcano plots.
The two catalytic subunits NAA20 and NAA80 as well as the auxiliary subunit NAA25 show strong enrichment on the E-CoA probe as shown in figure 2.23. This interaction indicates their specificity towards acidic N-terminal amino acids: NAA20 and NAA25

constitute the NatB complex that preferentially binds and acetylates N-terminal iMet residues followed by Asx/Glx residues [107]. Although the iMet is missing, the three consecutive glutamic acids seem to imitate the natural substrate sufficiently to enrich both N-acetyltransferases over the A-CoA control. NAA80 is the sole component of acetyltransferase NatH which has a very narrow substrate specificity towards β- and γ-actin [104]. Both proteins possess three aspartic acids or glutamic acids, respectively, at the N-terminus after iMet removal, a motif mimicked by this CoA-probe.

The main catalytic and auxiliary NATs show enrichment in the volcano plot of P-CoA vs. the control Ac-A. However, when compared to other CoA-peptide conjugates like E-CoA, it becomes evident that P-CoA enriches NATs only weakly (fig 2.23). This is in agreement with the finding that none of the NATs discovered so far acetylate N-termini starting with proline: An evaluation of 1200 N-terminal peptide fragments from proteins of *drosophila* cells by Goetze and co-workers found that an N-terminal proline after iMet removal is generally not acetylated [269]. Similar findings were made when analyzing 742 human and 379 yeast protein N-termini [270]. No Nt-acetyltransferase able to acetylate N-terminal prolines has been reported so far.

While all NATs are depleted in the interactome data of P-CoA compared to A-CoA, the CoA-binding protein pantothenate kinase 4 (PANK4) stands out as highly enriched (fig 2.23). PANK4 acts as part of the biosynthesis of CoA and, like other pantothenate kinases, is feedback inhibited by free CoA and Acetyl-CoA [271]. It appears like the PANK4 binds to the probe via its CoA/Acetyl-CoA binding site which results, together with less competition of NATs and KATs, in a strong enrichment on the immobilized peptides. The strength of enrichment however is still surprising and might hint towards a yet undiscovered function of PANK4.

Another enzyme that was enriched on the P-CoA probe is 2-oxoglutarate and Fe(II)-dependent oxygenase domain-containing protein 1 (OGFOD1) which was found to catalyze the 3-hydroxylation of Pro-62 in the small ribosomal protein S23 [272]. The P-CoA might resemble this site and enrich OGFOD1 over probes without N-terminal Pro.

The CoA-peptide probes displayed selectivity towards the respective NAT-complexes which leads to the question if the probe design could be used for the development of NAT-inhibitors. For this, an optimization of affinity and selectivity would be desirable which could be done using library-based display methods. Biological display methods could not be used, as genetic encoding of the N-terminal bromoacetylated amino acid would not be possible due to its reactivity, prohibiting subsequent conjugation of CoA. The RaPID method might be suitable for this task, as it was proven that even smaller peptides

could be aminoacylated to tRNA, which was used for successful *in vitro* translation [273]. Chemical methods like the method developed in this work are very well suited for this task as a bromoacetylated amino acids can be easily integrated, as all nucleophiles are protected, which is followed by CoA conjugation resulting in the final probe. As only the first 2-3 amino acids of a protein determine the specificity of NATs, the library would be small in size which makes the display method using SPOT synthesis even more favorable.

Functionality of the linker in the first part of the work is solely the connection between the substrate sequences for BD1 and BD2 and, as shown in chapter 3.1.2, linker length, especially when too long, has only a small negative impact on the binding. This is in stark contrast to the second part of the work, where the linker functions as a mean to simulate the approximation of both substrates' en route to the transition state. The optimal length of the linker connecting CoA and the amino acid was first analyzed using serotonin N-acetyltransferase (SNAT) and then successfully adopted for KAT inhibitors: Khalil et al. were the first to determine the linker length of a Lys-CoA-like inhibitor by testing different lengths connecting a CoA moiety and a serotonin and observed their influence on the binding of SNAT [274]. The acetyl spacer exhibited the best binding, while even two extra methylene groups reduced binding 10-fold. Addition of a third methylene group decreased binding 100 times in comparison to the acetyl moiety. The acetyl linker was later adapted for lysine acetyltransferases [179] and was also applied to the NAT inhibitors developed in the second part of this thesis. The linker length for NAT inhibitors however is not as well studied, but one report explored the effect of one additional methylene group in comparison to the linker determined in [274] and found that it increased the affinity of NatD towards a CoA-SGRGK inhibitor by 10-fold [275].

In contrast to ligands for bromodomains, which can displace oncogenic proteins from chromatin, the need for NAT inhibitors is not as pressing. Diseases correlated with NATs are often caused by mutations in the genes coding for NATs [104, 134] and lead to loss of function of the respective NATs which results in developmental and neurological disorders [104, 136]. Loss-of-function mutations cannot be treated with NAT specific inhibitors, however the role of NATs in cancer might be taken into consideration. Overexpression of some NATs has been shown to accelerate cancer proliferation and improve cancer cell survival, but the role of NATs like NAA10 in cancer remains rather ambiguous [140–143] (discussed in chapter 1.2.4). A better understanding of the functions of NATs in cancer is a prerequisite for application of potent NAT inhibitors in the future.

3.3 Outlook

In this work, a bivalent stable ligand was developed to enable *in vivo* binding to BRD3 and BRD4. Further analyses and experiments are needed to implement the ligand as a biotechnological probe, and further concepts can be envisioned:

Chemical mixture-based peptide libraries were used for seeking high affinity ligands of the four bromodomains in pull-down experiments. However, the interactions were only investigated with pull-down assays, so determining the dissociation constant of the ligand-Brd complexes by microscale thermophoresis (MST) or isothermal titration calorimetry (ITC) would provide further insights into the binding mode. Moreover, structural elucidation of the bromodomain proteins in complex with the optimized ligand would help understand these interactions on molecular level and would lay a base for theoretical analysis of attraction and repulsion of each individual amino acid, potentially explaining the enhanced binding of the affinity mutants.

Library screening experiments with the aim of lowering protease vulnerability of the peptide ligands led to the incorporation of N-methylated and D-amino acids impairing this interaction as judged by pull-downs. The next step would be testing the stability of peptides in cell lysates or in live cells, using the biotin moiety at the C-terminus of the bivalent stable ligand to enrich the peptides from the lysate followed by MS analysis. If these experiments show that degradation stability is low, the same library method could be applied to substitute more L-amino acids in the peptide with their N-methylated or D-derivatives.

New peptide-based ligands have been developed that target either BD1 or BD2 of the bromodomain-containing proteins BRD3 and BRD4. Selectivity towards a single Brd was not a selection criterion in the library screenings and was neither pursued nor tested. The toxicity of small molecule BET inhibitors like JQ1 is probably linked to its low selectivity amongst the members of the BET family, leading to the conclusion that off-target effects are a factor that prohibit its clinical application so far. Therefore, the library experiments could be repeated with the goal of identifying peptide sequences that discriminate between individual bromodomains resulting in selective peptidic bromodomain inhibitors that only target a single Brd in BRD3 and BRD4. Selective BET inhibitors would not only bear the advantage of potentially lower toxicity in medical applications, they would also offer the opportunity to observe abundance of single bromodomains in different tissues and disease cell lines by capturing with the probe and subsequent MS analysis.

Ultimately, the final bivalent stable ligand makes the peptide-based inhibitors suitable

for *in vivo* applications for research and ideally therapeutic purposes. Some therapeutic targets like disease-related proteins have shown up-regulation of expression levels when subjected to chemical perturbations like inhibitors, proving a resistance mechanism against this drug [276]. Adjustment of drug doses could easily approach toxicity limits, prohibiting their effective therapeutic use. Overexpression of BRD4 is one of the resistance mechanism towards BET inhibitors and was found to be addressable by coupling the BET inhibitor with a proteolysis targeting chimera (PROTAC). PROTACs are a class of heterobifunctional molecules with one moiety binding to the target protein and another moiety binding to a E3 ubiquitin ligase which, as part of the proteasomal degradation pathway, polyubiquitinilates the target protein which leads to its degradation [277]. Including a peptide-based BET inhibitor into a PROTAC, in contrast to a small molecule BET inhibitor, would potentially result in a highly selective ligand.

In the second project, a CoA-peptide conjugate was used to observe relative enrichment of NATs based on the N-terminal amino acid sequence of the probe. As discussed before in chapter 3.2, the sequence of the CoA-conjugates could be further improved in terms of affinity, selectivity and *in vivo* stability when using a library screen as in the first project. The resulting optimized probes could be applied for extended studies of substrate selectivity or even serve as NAT inhibitors.

The linker between the CoA moiety and the N-terminus has to be short for the CoA-conjugate to fit into the active site of NATs resulting in maximum affinity towards the acetyltransferase. The linker length was adapted from experiments with KATs and Lys-CoA-based peptides, but a recent paper showed that an additional methylene group improves binding in the case of NatD and Lys-CoA-based probes [275]. In future experiments, this linker could be combined with NatD-specific probes of the same design employed in this work. Furthermore, it could be tested if the addition of one methylene group also enhances the enrichment of other NAT complexes than NatD.

Finally, probes incorporating a third N-terminal amino acid can be developed and specificity for NAT complexes with overlapping substrates like NatC and NatE [104] could be tested.

4 Materials and methods

4.1 Materials

4.1.1 Suppliers

Suppliers for chemicals, consumables and equipment, the location of their headquarters and their abbreviated forms are listed in table 4.2 of the appendix. In this work, the short form of the suppliers is used.

4.1.2 Chemicals

Fmoc-protected canonical amino acids and 3-[bis(dimethylamino)methyliumyl]-3*H*-benzotriazol-1-oxide hexafluorophosphate (HBTU) were purchased from GL Biochem. Further Fmoc-protected building blocks as well as N-methylated and D-derivatives were acquired from Iris, Carbolution, Merck, Bachem and Sigma Aldrich. Resins and membranes for peptide synthesis were purchased from Rapp Polymere, Intavis and Merck. Solvents for synthesis and chromatographic systems were purchased from Sigma Aldrich, Th. Geyer, VWR and Biosolve. Other chemicals were acquired from AppliChem, Bachem, Biozym, Carbolution, Iris, Lonza, Merck, Carl Roth, Sigma Aldrich, TCI, Thermo Fisher, VWR and Wako, unless stated otherwise. Consumables were purchased from Carl Roth, Sarstedt, MultiSynTech and Eppendorf. The water used in this work was supplied by a Milli-Q® system (Merck) as ultrapure water with a resistivity at 25 °C of 18.2 MΩ · cm and a total organic carbon (TOC) content of less than 10 ppb.

4.1.3 Biological materials

LB-medium was purchased from Carl Roth. The bacterial strains *E. coli* BL21 (DE3) and *E. coli* XL1-Blue were both acquired from Agilent. Kits for DNA purification and extraction including buffer, chemicals and solutions were purchased from VWR and Thermo Fisher.

4.1.3.1 Enzymes and buffers

Table 4.1: Enzymes and buffers for plasmid processing (Thermo Fisher).

Enzyme	**Catalog #**
FastDigest NdeI	FD0583
FastDigest HindIII	FD0504
FastDigest PstI	FD0614
T4 DNA Ligase	EL0016
Buffer	**Catalog #**
FastDigest buffer (10×)	B64
T4 DNA Ligase buffer (10×)	B69

4.1.3.2 Plasmids and primers

Synthetic genes were codon-optimized for *E. coli* and purchased from Thermo Fisher.

Table 4.2: Plasmids containing gene of interest and their respective cloning sites.

Backbone	**Gene of interest**	**Cloning site**
pMA-RQ	BRD3.1-mKate2-Strep	5′-NdeI, 3′-HindIII
pMA-T	BRD3-mTagBFP2	5′-NdeI, 3′-HindIII
pMA-T	BRD4-mTagBFP2	5′-NdeI, 3′-HindIII
	Empty expression vectors	
Vector	**Tags**	**Supplier**
pET28a(+)	$^{N}His_6$, $^{C}His_6$	Merck

Table 4.3: Primers for sequencing.

Primer	**Sequence (5′ to 3′)**
T7 Promoter forward	TAATACGACTCACTATAGGG
T7 Terminator reverse	GCTAGTTATTGCTCAGCGG

4.1.4 Instruments

All instruments used in this work are listed in table 4.4. The instrument, their specification and the manufacturer are given.

Table 4.4: Instruments and manufacturers.

Instrument	Specification	Manufacturer
Centrifuge	Megafuge™ 16R	Thermo Fisher
	Sorvall™ RC 6 Plus	Thermo Fisher
	Centrifuge 5420	Eppendorf
Chromatograph	Äkta™ Prime Plus	Cytiva
Clean Bench	MSC-Advantage™	Thermo Fisher
Freeze dryer	Alpha™ 2-4 LDplus	Christ
Electroporator	Micropulser™	Bio-Rad
Gel system	PerfectBlue Mini M	VWR
	Mini-PROTEAN® Tetra cell	Bio-Rad
	Mini Trans-Blot® Cell	Bio-Rad
Homogenizer	Emulsiflex C5	Cytiva
HPLC	ProStar 210	Varian
	LC-10AT VP	Shimadzu
	LC-20AR	Shimadzu
Imager	ChemiDoc™ MP	Bio-Rad
Incubator	Multitron Standard	Infors HT
Mass spectrometer	LC-MS2020	Shimadzu
	Exploris 480	Thermo Fisher
	Reflex IV	Bruker
Peptide synthesizer	Syro I	MultiSynTech
	ResPep SL	Intavis
Rotary evaporator	RV 10	VWR
Shaker	Thermomixer® compact	Eppendorf
UHPLC	EASY-nlc 1200	Dr. Maisch
UV/Vis spectrometer	GENESYS™ 10S	Thermo Fisher
	NanoDrop™ One	VWR

4.1.5 Software

The software used to record and analyze data in this work is shown in table 4.5.

Table 4.5: Software and application.

Software	Application
Affinity Designer 1.10.4.1198	Editing graphics
ApE - A plasmide Editor 3.0.4	DNA sequence analysis
ChemDraw Professional 20.0.0.41	Drawing of chemical structure
Galaxie Chromatography Data System	Control software HPLC
Image Lab 6.0.1	Gel and membrane imaging
LabSolutions 5.91	LC/MS and HPLC data analysis
MaxQuant 2.0.3.0	MS/MS raw data processing
Microsoft Excel 2019	General data handling
Mmass 5.5.0	MS data analysis
MultiPep 4.2.20	Control software ResPep SL
Perseus 1.16.15.0	MS/MS data analysis
PyMol 2.0.7	3D structures of molecules
R 3.5.0	Statistical processing
Vision lite 4.0	Control software UV/Vis spectrometer

4.2 Methods

4.2.1 Analytical methods

4.2.1.1 Analytical HPLC

Analytical reversed-phase high-performance liquid chromatography (RP-HPLC) was conducted on a Shimadzu LC-10AT VP system equipped with a Reprospher 100 C18 column (250 × 4.6 mm, 5 µm, 100 Å, Dr. Maisch) using the solvents HPLC-A (0.1 % trifluoroacetic acid (TFA) in water) and HPLC-B (80 % acetonitrile (ACN), 0.1 % TFA in water). For each run, 100 µL of the sample was injected. Analytes were eluted with a gradient of 5-95 % HPLC-B for 20 min at a flow rate of 1.5 mL/min and the UV chromatograms were recorded at 218 nm and 280 nm.

4.2.1.2 Preparative HPLC

Preparative RP-HPLC purifications were performed either on a Varian ProStar 210 HPLC system or on an LC-20AR HPLC system from Shimadzu both equipped with a Reprosil C18 column (250 × 20 mm, 5 µm, 100 Å, Dr. Maisch) at a flow rate of 13 mL/min. For semi-preparative HPLC, a ReproSil 100 C18 column (250 × 10 mm, 5 µm, 100 Å, Dr. Maisch) and a flow rate of 4 mL/min was used. The crude sample was dissolved in a mixture of HPLC-A and HPLC-B, centrifuged (10 min, 5,000 × g, room temperature) and the supernatant was subsequently injected into the column. The compounds were separated using a gradient of 5-95 % HPLC-B for 40 min, unless stated otherwise, and the UV absorption was detected at 218 nm. Collected fractions were analyzed by LC-MS and product-containing fractions were combined and lyophilized.

4.2.1.3 Liquid chromatography mass spectrometry

Liquid chromatography mass spectrometry (LC-MS) measurements were carried out on a Shimadzu LCMS-2020 system with a Kinetex C18 column (100 × 2.1 mm, 2.6 µm, 100 Å, Phenomenex) serving as the stationary phase. LC-MS solvent A (0.1 % formic acid (FA) in water) and LC-MS solvent B (80 % ACN, 0.1 % FA in water) were used as eluents. After injection of the sample, a flow rate of 0.2 mL/min and a gradient of 5-95 % LC-MS solvent B was applied. The UV absorption was measured at 218 nm and the connected electron spray ionization (ESI) mass spectrometer was operated in positive mode.

4.2.1.4 MALDI-TOF mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry was performed on a Bruker Reflex IV MALDI-TOF mass spectrometer in reflection mode using α-cyano-4-hydroxycinnamic acid (CHCA) as a matrix substance (5 mg/mL CHCA, 0.1 % TFA, 50 % ACN in water). Equal volumes of sample and matrix were added on a polished steel target, pipetted up and down to achieve a homogeneous mixture and left to dry. Between 50 and 100 shots as well a laser intensity of 30 to 50 % were used to record the spectra. The spectra were analyzed with the MMass 5.5.0 software.

4.2.1.5 UV-Vis spectrophotometry

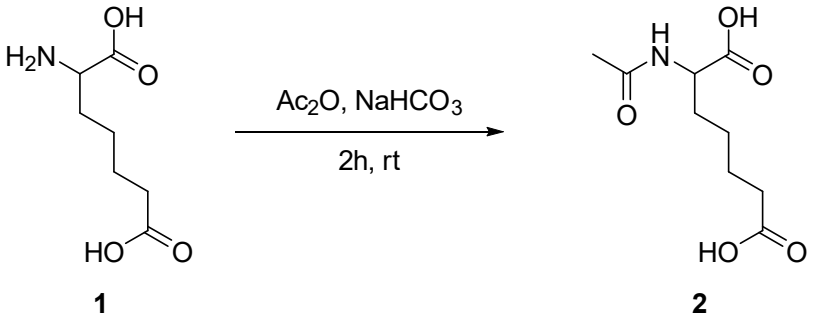
Determination of free thiols was conducted using the Thermo Scientific™ GENESYS™ 10S Ultraviolet-visible (UV-Vis) spectrophotometer (Thermo Fisher) by measuring the

absorbance of the sample at a wavelength of 412 nm. The method for quantifying the concentration of free thiols is reported in chapter 4.2.3.7. The same device was used to determine the concentration of bacterial cell cultures by recording the optical density at 600 nm (OD_{600}).

4.2.2 Building block synthesis

4.2.2.1 Synthesis of Fmoc-ApmTri-OH

Synthesis of Ac-Apm-OH (2)



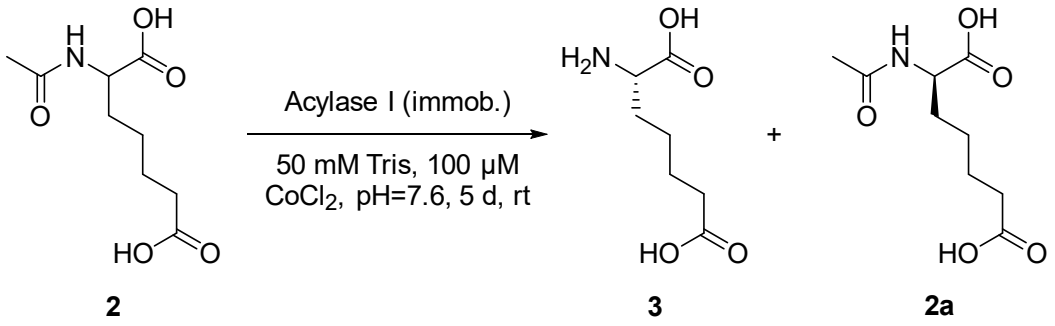
1 **2**

DL-Aminoheptanedioic acid (H-DL-Apm-OH) (**1**) (1.75 g, 10 mmol, 1 eq) was dissolved in water (20 mL) and the pH was adjusted to 8-9 using $NaHCO_3$ (1 M in water). Acetic anhydride (1.89 mL, 20 mmol, 2 eq) was added dropwise and the mixture was stirred for 30 min at room temperature. Lyophilization of the aqueous reaction solution overnight yielded crude white solid product which was dissolved in water, acidified by adding TFA and subsequently purified by RP-HPLC. Ac-DL-Apm-OH (**2**) was obtained as a white solid (2.146 g, 9.89 mmol, 99 %).

HPLC (5-50 % B in 40 min, 13 mL/min): $t_{ret.}$ = 7.9 min.

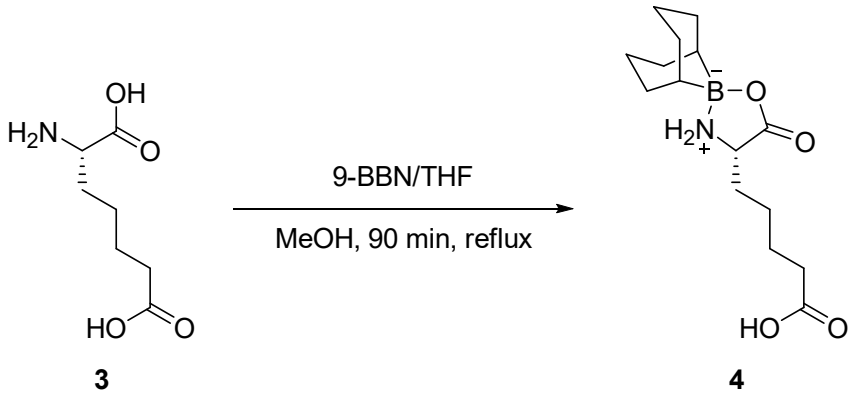
ESI-MS: $[M+H]^+$, m/z = 218.10 (calculated), 217.80 (found).

Synthesis of H-Apm-OH (3)



Ac-DL-Apm-OH (**2**) (2.146 g, 9.89 mmol) was dissolved in acylase I buffer (50 mM tris(hydroxymethyl)aminomethane (Tris), 100 µL $CoCl_2$, pH 7.6) and the pH value was adjusted to 7.6 with the help of NaOH solution (6 M). The reaction mixture was added to immobilized acylase I (from *Aspergillus sp.*, immobilized on Eupergit® C, Sigma Aldrich) (30 U, 450 mg) which was priorly washed with water (3 × 5 mL) and then equilibrated with Tris-buffer (50 mM Tris, pH 7.6) (3 × 5 mL). The resulting suspension was incubated for 5 days on a shaker at room temperature and the solution containing the product H-L-Apm-OH (**3**) as well as the by-product Ac-D-Apm-OH (**2a**) was subsequently lyophilized yielding a white solid.

Synthesis of 9-BBN-Apm-OH (4)



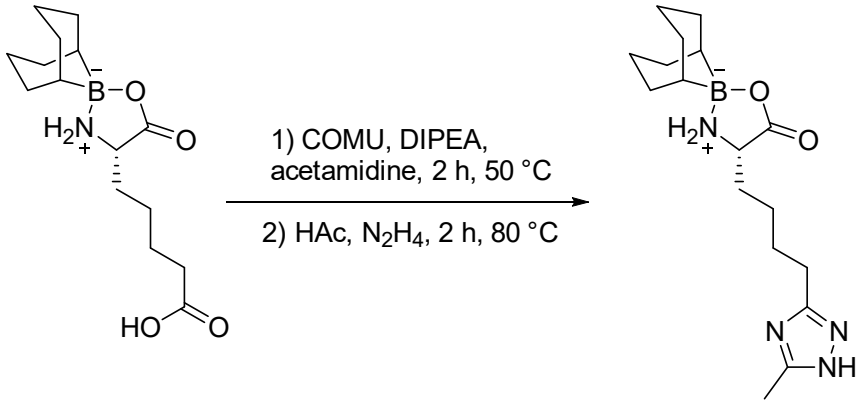
The lyophilized product mixture from the previous step containing **3** and **2a** was suspended in anhydrous MeOH (25 mL) under argon atmosphere and heated at reflux. 9-Borabicyclo[3.3.1]nonane (9-BBN) (0.5 M in THF, 10 mL, 5 mmol, 1.0 eq in relation to 50 % of **2**) was added drop-wise to the solution and the product mixture was furthermore

heated at reflux for 90 min until the solution became clear. The solvent was evaporated under vacuum, the solid product was dissolved and subsequently purified via preparative RP-HPLC. After combination and lyophilization of the product fractions, 9-BBN-L-Apm-OH (**4**) was obtained as a white powder (1.17 g, 3.96 mmol, 40 %).

HPLC (50-80 % B in 40 min, 13 mL/min): $t_{ret.}$ = 16.4 min.

ESI-MS: $[M+H]^+$, m/z = 296.20 (calculated), 296.00 (found); $[2M+H]^+$, m/z = 591.40 (calculated), 591.35 (found).

Synthesis of 9-BBN-ApmTri-OH (5)



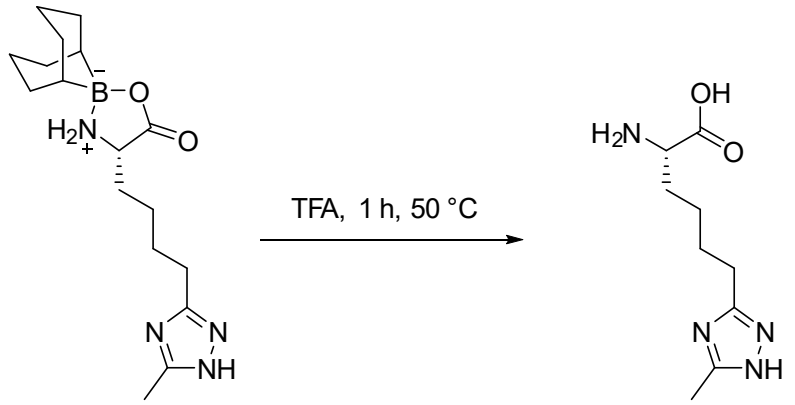
4 **5**

9-BBN-L-Apm-OH (**4**) (1.17 g, 3.96 mmol) was dissolved in DMF (15 mL) and heated at 50 °C. (1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate (COMU) (1.865 g, 4.36 mmol, 1.1 eq) and acetamidine hydrochloride (2.246 g, 23.7 mmol, 6 eq) were each dissolved in DMF (20 mL) separately. The COMU solution and *N*,*N*-Diisopropylethylamine (DIPEA) (4.04 mL, 23.7 mmol, 6 eq) were added to the educt and stirred at 50 °C for 5 min. The acetamidine solution was added and the mixture was stirred at 50 °C for 2 h. Then, acetic acid (2.26 mL, 39.6 mmol, 10 eq) and hydrazine monohydrate (0.288 mL, 5.94 mmol, 1.5 eq) were added and stirring was continued at 80 °C for an additional 2 h. DMF was removed under vacuum and the residue was lyophilized. The crude product was purified using RP-HPLC yielding 9-BBN-L-Apm(3-methyl-1,2,4-triazole)-OH (**5**) as a white solid (319.3 mg, 0.96 mmol, 24 %).

HPLC (5-95 % B in 50 min, 13 mL/min): $t_{ret.}$ = 23.0 min.

ESI-MS: $[M+H]^+$, m/z = 333.24 (calculated), 333.25 (found).

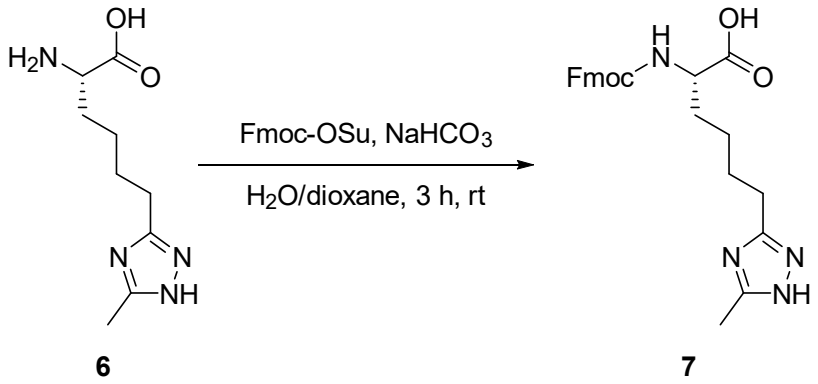
Synthesis of H-ApmTri-OH (6)



5 **6**

9-BBN-L-Apm(3-methyl-1,2,4-triazole)-OH (**5**) was directly dissolved in TFA (10 mL) and heated for 1 h at 50 °C under stirring for deprotection. The TFA was removed under reduced pressure and the residue was lyophilized. The resulting product H-L-Apm(3-methyl-1,2,4-triazole)-OH (**6**) was obtained as a yellow-brown solid.

Synthesis of Fmoc-ApmTri-OH (7)



6 **7**

6 was dissolved in water (7 mL) and the pH value was adjusted to 8-9 by addition of $NaHCO_3$ (1 M in water). *N*-(9-Fluorenylmethoxycarbonyloxy)succinimide (Fmoc-OSu) (364.14 mg, 1.06 mmol, 1.1 eq in relation to **5**) was dissolved in 1,4-dioxane (10 mL) and added to the aqueous educt solution. The reaction mixture was stirred for 2 h at room temperature after which the solvent was evaporated under vacuum and the residue was lyophilized. The crude product was purified using RP-HPLC and the applicable fractions

were combined and lyophilized. The final product Fmoc-L-Apm(3-methyl-1,2,4-triazole)-OH (**7**) was obtained as a white solid (200 mg, 0.46 mmol, 48 %).

HPLC (40-70 % B in 50 min, 13 mL/min): $t_{ret.}$ = 15.0 min.

ESI-MS: $[M+H]^+$, m/z = 435.20 (calculated), 435.20 (found).

4.2.3 Solid-phase peptide synthesis

All peptides in this work were synthesized via solid-phase peptide synthesis (SPPS) while applying the Fmoc-strategy. The functionalized cellulose membrane Amino-PEG_{500}-UC540 (loading capacity: 400 nmol/cm^2) and three different resins were employed as a solid support: the TentaGel® HL RAM resin (loading capacity: 0.38 mmol/g), the Fmoc-L-Glu(OtBu)-tritylchloride-polystyrene-(TCP) resin (loading capacity: 0.35 mmol/g) and the H-L-Lys(Boc)-2-ClTrt resin (loading capacity: 0.79 mmol/g). The scale of the synthesis ranged from 12.5 µmol over 25 µmol to 50 µmol per petide for the resin-bound syntheses and was at ~12,5 nmol per spot for the membrane-bound syntheses. The following amino acid derivatives were used in the automated peptide syntheses: 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-L-Gly-OH, Fmoc-L-His(Trt)-OH, Fmoc-L-Ile-OH, Fmoc-L-Leu-OH, Fmoc-L-Lys(Boc)-OH, Fmoc-L-Met-OH, Fmoc-L-Phe-OH, Fmoc-L-Pro-OH, Fmoc-L-Ser(tBu)-OH, Fmoc-L-Thr(tBu)-OH, Fmoc-L-Trp(Boc)-OH, Fmoc-L-Tyr(tBu)-OH, Fmoc-L-Val-OH, Fmoc-Ahx-OH, Fmoc-L-Lys(Ac)-OH, Fmoc-D-Arg(Pbf)-OH, Fmoc-D-His(Trt)-OH, Fmoc-D-Leu-OH, Fmoc-D-Phe-OH, Fmoc-D-Trp(Boc)-OH, Fmoc-D-Tyr(tBu)-OH, Fmoc-D-Val-OH, Fmoc-L-*N*-Me-Arg(Pbf)-OH, Fmoc-L-*N*-Me-His(Trt)-OH, Fmoc-L-*N*-Me-Leu-OH, Fmoc-L-*N*-Me-Phe-OH, Fmoc-L-*N*-Me-Trp(Boc)-OH, Fmoc-L-*N*-Me-Tyr(tBu)-OH, Fmoc-L-*N*-Me-Val-OH and Fmoc-2,4-dimethoxybenzyl-phenoxyacetic acid (Rink Linker).

4.2.3.1 Automated peptide synthesis on resin

The resin based peptide synthesis was conducted on a Syro I system (MultiSynTech) using a double-coupling method. All amino acid derivatives were dissolved in *N*-methyl-2-pyrrolidone (NMP) and all other reagents were dissolved in DMF. Each cycle of the synthesis consisted of two steps: the double coupling of the building block and the Fmoc-deprotection. For the first coupling, the respective amino acid derivative (3 eq in relation to the resin) was activated with *N*,*N*′-diisopropylcarbodiimide (DIC) (3 eq) and ethyl cyanohy-

droxyiminoacetate (Oxyma) (3 eq) and coupling took place for 40 min while shaking. The coupling was repeated using the building block (3 eq), 1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxide hexafluorophosphate (HATU) (2.7 eq) as the activation reagent and *N*-methylmorpholine (NMM) (5.4 eq) as a base while shaking for 30 min. Fmoc-deprotection was achieved by treating the resin with 40 % piperidine in DMF for 3 min and afterward again for 12 min. Between each step, the resin was thoroughly washed with DMF.

4.2.3.2 Automated peptide synthesis on membrane (SPOT synthesis)

Membrane bound peptides were synthesized using the ResPep SL Automated Multiple Peptide Synthesizer (Intavis). As in the resin based SPPS, the amino acid derivatives were dissolved in NMP and all other reagents in DMF. Because the size and therefore the amount of peptide differs from spot to spot, the volume instead of the equivalent will be used to describe the process of the SPOT synthesis. In the first coupling cycle, Oxyma (30 nL, 1.1 M), DIC (30 nL, 1.1 M) and the respective amino acid derivative (60 nL, 0.5 M) were pre-activated for 10 min and afterward coupled for 10 min. Coupling was repeated once for 15 min. All following cycles were conducted using HATU (59 nL, 0.5 M), NMM (15 nL, 4 M) and the respective amino acid derivative (60 nL, 0.5 M). The coupling lasted for 10 min and was repeated once for an additional 10 min. After each coupling cycle, the membrane was soaked with a capping mixture (5 % acetic anhydride, 6 % 2,6-lutidine in DMF) for 5 min to acetylate all remaining free N-termini. The Fmoc-group was removed by treating the membrane twice with 20 % piperidine in DMF for 5 min each. Between each step, the membrane was washed thoroughly with DMF and ethanol.

4.2.3.3 Manual peptide synthesis

Manual coupling of amino acids

All manual couplings were conducted at room temperature on a shaker. After the coupling, the resin was washed with DMF (3 × 2 mL for 5 min each) and, if required, washed with DCM (3 × 2 mL) and dried for 10 min under vacuum. Synthesized Fmoc-L-Apm(3-methyl-1,2,4-triazole)-OH was manually coupled using 2 eq of the building block together with HATU (1.5 eq) in 400 mM NMM in DMF (4 mL) for 1 h.

Coupling of the commercially available building blocks Fmoc-L-propargylglycine-OH (Fmoc-Pra), Fmoc-4-azido-phenylalanine-OH (Fmoc-F_{AZ}) and 3-mercaptopropionic acid (3-MPA) was performed by adding [ethyl-cyano-(hydroxyimino)-acetato-O^2]-tri-1-pyrrolidinylphos-

phonium hexafluorophosphate (PyOxim) (3 eq, in 2 mL 400 mM NMM in DMF) to the dissolved building block (3 eq, in 2 mL 400 mM NMM in DMF). The mixture was preactivated for 10 min, followed by an incubation with the resin for 1 h. Coupling of Fmoc-NH-PEG_3-OH was done in a similar fashion, but 2 instead of 3 eq of the building block and PyOxim were used.

The amino acid derivatives Fmoc-*N*-ε-1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl-L-lysine-OH (Fmoc-Lys(Dde)), Fmoc-L-Lys(Biotin)-OH and Fmoc-L-Lys(5-carboxyfluorescein)-OH (Fmoc-Lys(5-FAM)) were coupled by incubating the resin with the building block (3 eq, in 2 mL 400 mM NMM in DMF) and HATU (2.7 eq, in 2 mL 400 mM NMM in DMF) for 1 h.

The Fmoc-group was removed by treating the resin-bound peptide with 40 % piperidine in DMF (3 × 2 mL for 5 min each). Afterward, the resin was washed with DMF (3 × 2 mL for 5 min each).

Acetylation of N-terminus

Except for the peptides that required further modification at the N-terminus, all peptides were acetylated at the $N\alpha$-group of the first amino acid. The acetylation was performed by incubating the resin with a capping mixture (5 % acetic anhydride, 6 % 2,6-lutidine in DMF) (2 × 20 min) while shaking. Afterward, the resin was washed with DMF (3 × 2 mL for 5 min each), then with DCM (3 × 2 mL) and lastly dried for 10 min under vacuum.

Coupling of 2-bromoacetic acid to N-terminus

An electrophilic linker in the form of 2-bromoacetic acid was used to couple coenzyme A to the N-terminus of the peptide probe. The coupling of this linker was performed under water- and O_2-free conditions under argon atmosphere. Reaction vessels used in the synthesis were evacuated for 5 min and subsequently flushed with argon. This process was repeated 2 times. Syringes and cannulas used to transfer liquids were flushed with argon 3 times. 2-bromoacetic anhydride (5 eq in relation to the peptide) was dissolved in 5 mL anhydrous DMF and transferred to the dry resin. To minimize the amount of water and O_2 that entered the vessel containing the resin after transferring the dissolved 2-bromoacetic anhydride, a constant flow of argon gas was passed through the container for 5 minutes. The coupling took place on a shaker at room temperature for 30 minutes and was repeated once. After the coupling, the resin was washed with DMF (3 × 2 min) and with DCM (3 × 30 s), subsequently dried under reduced pressure and stored at 4 °C.

Removal of Dde-group from lysine side-chain

The Dde-group was used as an orthogonal protection group for coupling of the $N\varepsilon$-group of lysine to the C-terminus of the peptide. The Lys(Dde) deprotection was conducted by treating the resin 3 × 2 min with 4 mL 2 % hydrazine in DMF. Subsequently, the resin was washed with DMF (3 × 2 mL for 5 min each).

4.2.3.4 Cleavage of resin-bound peptides

Before resin-bound peptides were subjected to cleavage, they were dried *in vacuo*. Cleavage off the resin and parallel removal of all acid-labile protection groups was achieved by treatment with 10 mL of a cleavage cocktail (CC) containing TFA, water, triisopropyl silane (TIPS) and phenol (85:5:5:5). Firstly, the resin was incubated with 7 mL CC for 2 h on a shaker. The solution was then collected and the resin was incubated twice with 1.5 mL fresh CC for 30 min each and afterward washed with DCM (2 mL). All solutions were combined and concentrated on a rotary evaporator. Precipitation of the peptide was achieved by transferring the concentrated solution to -80 °C cold diethyl ether followed by centrifugation (10 min, 4,000 × g, -4 °C). After removing the supernatant, the pellet was dissolved in water (10 mL) and subsequently lyophilized. RP-HPLC was used to purify the crude peptide.

4.2.3.5 Cleavage of fully protected peptides from TCP-resin

For cleavage without simultaneous removal of the acid-labile protection groups, a different protocol was applied to the TCP-resin-bound peptides. At first, the resin was subjected to 3 mL 20 % hexafluoroisopropanol (HFIP) in DCM for 2 h, followed by washing the resin twice with fresh cleavage solution (3 mL each) for 10 min. For achieving complete cleavage, the resin was incubated with 3 mL of a second mixture consisting of DCM, 2,2,2-trifluoroethanol (TFE) and acetic acid (8:1:1) for 2 h, again followed by washing twice with fresh solution (3 mL each) for 10 min. All solutions were combined and concentrated under reduced pressure. The crude peptide was then washed once with 20 mL 5 % $NaHCO_3$ in water, followed by 3 × 20 mL water to remove the salts. Finally, the precipitate was dissolved in 10 mL water/tBuOH (50:50) and lyophilized.

4.2.3.6 Deprotection and cleavage of membrane-bound peptide

Prior to the deprotection of the membrane-bound peptides, the dry membrane was illuminated using UV-light (λ = 366 nm) and the spots containing rink linker-bound

peptide were marked with a pencil, cut out and separately stored. The membrane was then deprotected for 1.5 h using 10 mL of the same CC as described before (TFA, water, TIPS, phenol (85:5:5:5)). Afterward, the membrane was rinsed and washed with DCM (10 mL, 4 × 30 s), DMF (10 mL, 4 × 2 min) and ethanol (10 mL, 4 × 5 min). The membrane was subsequently cut into desired dimensions, shrink-wrapped and stored at -20 °C.
The cut-out spots with peptides containing a rink linker were cleaved from the membrane and analyzed via LC-MS. Therefore, the spot was treated with 50 µL CC for 15 min at 50 °C. Afterward, the solution was put in 2 mL of -80 °C cold diethyl ether and centrifuged for 10 min at 14,100 rpm. The supernatant was discarded and the peptide pellet was dissolved in 100 µL of LC-MS buffer A and analyzed by LC-MS.

4.2.3.7 Cyclization of cyclic cell-penetrating peptides (cCPPs)

Three different cyclization strategies were used in this work. In the head-to-side-chain cyclization, a bromoacetic acid was introduced to the N-terminus of the peptide (4.2.3.3) which reacted in an intramolecular fashion with the thiol of a cysteine after deprotection. In the side-chain-to-tail and the head-to-tail cyclization, the coupling points were selectively deprotected and subsequently coupled in high dilution using standard SPPS coupling reagents. For these latter two cyclization strategies, the peptide was dissolved in DMF (100 mL, equals a concentration of 0.25 mM for 50 % yield of a 50 µmol peptide synthesis). HBTU (56.9 mg, 150 µmol, 6 eq relative to 50 % yield of a 50 µmol peptide synthesis), Oxyma (21.3 mg, 150 µmol, 6 eq) and DIPEA (52.3 µL, 300 µmol, 12 eq) were added and the solution was stirred at room temperature for 6 h. Additional 6 eq HBTU and 12 eq DIPEA were added and the reaction mixture was left stirring overnight. Afterward, the solvent was removed using a rotary evaporator, water was added to the residual and it was subsequently freeze-dried. To remove impurities, the crude product was dissolved in ethyl acetate and washed with brine (3 × 100 mL) and $NaHCO_3$-solution (3 × 100 mL, saturated in water). The organic phase was then dried under reduced pressure and freeze-dried. Deprotection followed as described in (4.2.3.4), except that the peptide was only incubated once with 30 mL CC for 3 h at room temperature.

Spectroscopic test for free thiols

To test if the head-to-side-chain cyclization worked as indicated or if the $N\varepsilon$ of lysine unintentionally acted as the nucleophile in the cyclization, a test for free thiols was conducted. A plastic cuvette was filled with 50 µL 5,5′-Dithiobis(2-nitrobenzoic acid)

(DTNB) solution (50 mM sodium acetate, 2 mM DTNB in water), 100 µL Tris solution (1 M, pH 8.0), filled with water to 990 µL and measured as a blank at 412 nm. 10 µL of the sample peptide were directly added to the cuvette, the solution was homogenized and the absorbance was recorded after 5 min of incubation. To calculate the amount of free thiols in the sample, the extinction coefficient of the TNB^{2-}-anion was used: $\varepsilon(TNB^{2-})$= 14,150 $M^{-1}cm^{-1}$ [278]. The absorbance of the sample A_S was yielded according to

$$A_S = \frac{V}{V_S} \cdot A_{412}$$

where V is the total volume, V_S is the volume of the sample and A_{412} is the recorded absorbance at 412 nm. By measuring triplicates, an average absorbance of the sample was used to calculate the concentration of free thiols in the sample, using the equation

$$c(thiol) = \frac{\bar{A}_S}{\varepsilon(TNB^{2-})}$$

where $c(thiol)$ is the concentration of free thiols and $\bar{A}_S$ is the average absorbance of sample A.

4.2.3.8 N-terminal coupling of coenzyme A (CoA)

For the coupling of CoA, the purified bromoacetylated peptides were dissolved in triethylammonium bicarbonate buffer (TEAB) (0.5 mL, 1 M in water, pH 8.5) and CoA trilithium salt (5 eq) was separately dissolved in the same amount of buffer. The two solutions were combined and the reaction mixture was incubated for 24 h under argon at room temperature. After verifying the complete conversion via LC-MS, the reaction mixture was directly introduced to the HPLC and subsequently purified. Finally, all fractions containing the product were combined, lyophilized and the dry peptide-conjugate was stored at -20 °C.

4.2.3.9 Copper-catalyzed azide-alkyne Huisgen cycloaddition

A peptide ligation strategy employed in this work is the copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC). For this reaction, stock solutions for the azide-containing peptide (20 mM in water), the alkyne-containing peptide (20 mM in water), $CuSO_4$ (100 mM in water), tris((1-benzyl-4-triazolyl)methyl)amine (TBTA) (100 mM in dimethyl sulfoxide (DMSO)) and sodium ascorbate (1.8 M in water) were prepared. The aqueous solutions of

the azide-containing peptide (3 µmol), the alkyne-containing peptide (4 µmol), $CuSO_4$ (12 µmol), and sodium ascorbate (2.2 mmol) were combined and DMSO was added until a ratio of water to DMSO of 7:3 was reached. Following that, the TBTA solution (12 µmol) was added and the reaction was incubated overnight at room temperature. After the success of the reaction was verified by LC-MS, the reaction mixture was directly introduced to the HPLC and subsequently purified. For the conjugation of the CoA-peptide conjugates with CuAAC, TCEP at a final concentration of 0.5 mM was added to the HPLC fractions prior to lyophilization.

4.2.3.10 Immobilization of thiol-containing peptides

For the immobilization on SulfoLink™ agarose beads (Thermo Fisher), the peptides were dissolved in water to yield a 10 mM or 100 mM stock solution, depending on the solubility. To ensure that even amounts of peptides were immobilized, the concentration of the stock solutions were verified via analytical HPLC, for each set of peptides separately. The stock solutions were diluted with HPLC A to a concentration of 1 mM peptide and equal volumes were injected. The adjusted volume V_{exact} for the immobilization was yielded according to

$$V_{exact} = V \cdot F = V \cdot \frac{\bar{A}}{A_i}$$

where F is the correction factor for the volume of stock solution employed in the immobilization V, which was calculated for each peptide by dividing the average peak area $\bar{A}$ by the individual peak area A_i, both recorded at 218 nm or 280 nm.
The respective V_{exact} was used to dilute the individual solutions to a concentration of 1 mM in coupling buffer (310 µL, 50 mM Tris-HCl, 5 mM ethylenediaminetetraacetic acid (EDTA) in water, pH 8.5). For each peptide 300 µL of SulfoLink™ bead slurry (50 % beads in 10 mM EDTA-Na, 0.05 % NaN3, 50 % glycerol) was used, which was equilibrated with coupling buffer (6 × 800 µL) prior to coupling. The equilibrated SulfoLink™ beads were incubated with 300 µL of the prepared 1 mM peptide solution for 15 min on a shaker and for an additional 45 min without shaking at room temperature. After washing the beads with coupling buffer (3 × 500 µL), the matrix was blocked by incubating the beads with 500 µL blocking buffer (3.49 µL β-mercaptoethanol in coupling buffer) for 1 hour, the first 15 mins on a shaker, the latter 45 mins without shaking at room temperature. Finally, the beads were washed with 1 M NaCl (6 × 1 mL), water (2 × 1 mL) and lastly with storage buffer (4 × 1 mL 50 % ACN in water). The peptide-conjugated beads were stored as a 25 % slurry (450 µL storage buffer + 150 µL beads) in 40 µL aliquots at -20 °C.

4.2.4 Biochemical methods

4.2.4.1 Cultivation of *E. coli* strains

If not stated otherwise, *E. coli* strains were cultivated in LB medium containing the appropriate antibiotic (ampicillin: 100 µg/mL, kanamycin: 30 µg/mL) and cultivation took place under shaking (170 rpm) at 37 °C. New strains were stored at -80 °C as cryo cultures, which were prepared by adding 800 µL of a sterile glycerine solution (40 % in water) to 800 µL of an overnight culture.

4.2.4.2 General methods for cloning

Transformation of electrocompetent bacteria cells

If not already in solution, the plasmid to be transformed was dissolved with water to reach a concentration of 250 µg/mL. A small volume was then further diluted to a final concentration of 5 µg/mL. An aliquot (50 µL) of electrocompetent *E. coli* (either *E. coli* DE3 (BL21) or *E. coli* XL1-blue strain) was thawed on ice and mixed with 1 µL of the plasmid solution (5 µg/mL, 5 ng). The mixture was pipetted into an electroporation cuvette (2 mm, 450 µL) (Carl Roth) and the cells were pulsed (U = 2.5 kV, 5 ms) with a MicroPulser™ electroporator (Bio-Rad). The cells were immediately mixed with 1 mL of pre-warmed LB medium and incubated at 37 °C for 1 h while shaking at 400 rpm. Afterward, the cells were plated on LB agar Petri dishes containing the appropriate antibiotic and incubated overnight at 37 °C.

Preparation of plasmid DNA

Single colonies of transformed *E. coli* XL1-blue cells were picked, transferred into 5 mL fresh LB medium containing the appropriate antibiotic and incubated overnight. Plasmid isolation was conducted using the peqGOLD Plasmid Miniprep Kit 1 (VWR) following the kit's instruction manual. The concentration of the purified plasmid was determined using the NanoDrop™ One microvolume UV-Vis spectrophotometer (Thermo Fisher).

Restriction enzyme digestion

Purified plasmids were digested utilizing FastDigest (FD) restriction enzymes (Thermo Fisher). One double digestion reaction consisted of 10× FD-buffer (3 µL, Thermo Fisher), 1-2 µg DNA and 0.5-1.5 µL of each restriction enzyme (volume depending on their activity), filled up with water to a volume of 30 µL. The digestion took place at 37 °C for 3.5 h

followed by a heat deactivation at 80 °C for 10 min, after which the reaction mixture was stored at -20 °C.

Agarose gel electrophoresis

For the separation of the digestion products, a 1 % agarose gel was cast, consisting of 0.5 g agarose and 50 mL tris-acetate-EDTA (TAE) buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.5). Both substances were mixed and heated to dissolve the agarose, then the mixture was let cool down to approximately 40 °C and Midori Green Advance (50 µL of a 1:10 dilution of the stock) was added before pouring the solution into the cast. After the gel solidified, the DNA samples were mixed 5:1 with 6× Loading Dye and filled into the gel pockets. Gene Ruler DNA Ladder Mix SM0333 (3 µL) (Thermo Fisher) was used as a size marker on the gel. The chamber containing the gel was filled with 1× TAE buffer and the electrophoresis was run for 1 h at 80 mA and 100 V. After the run was finished, the bands containing the DNA of interest were cut out using UV light as an illumination source and were stored at -20 °C for further use.

DNA extraction from agarose gels

The GeneJET Extraction Kit (Thermo Fisher) was used to extract the DNA from the gel following the kit's instruction manual. The concentration of the extracted DNA was measured via UV/Vis spectrophotometry.

Ligation of DNA fragments

Ligation of digested and purified DNA fragments was conducted using 1 eq of the plasmid and 10 eq of the insert. The plasmid and the insert as well as T4 DNA Ligase (1 U, 1 µL) (Thermo Fisher) and 10× T4 DNA Ligase buffer (1 µL) (Thermo Fisher) were combined and filled up with water to a volume of 10 µL. After the ligation took place at room temperature for 30 min, 1 µL of the mix was directly used for cell transformation, the remaining 9 µL were stored at -20 °C for further use.

DNA sequencing

The DNA sequence was verified by sequencing. Therefore, 5 µL of plasmid DNA with a concentration of 80-100 ng/mL was mixed with 5 µL of the desired primer (c = 5 µM). The mixture was then sent to the company GATC, where the sequencing was conducted using the didesoxy method according to Sanger [279]. Analysis of the results was done with the help of the ApE software.

4.2.4.3 General methods for recombinant protein expression and purification

Recombinant bacterial protein expression

Pre-cultures were prepared by adding kanamycin (30 µg/mL) to 20 mL LB medium and inoculating it with *E. coli* BL21 (DE3) strains containing a pET28a(+) expression vector with the appropriate insert. After incubating the pre-cultures overnight, a small sample was taken and diluted with fresh LB medium. The diluted sample was used to measure the OD_{600} and calculate the amount of pre-culture necessary for achieving an initial OD_{600} of 0.1 in the main culture (700 mL of LB medium supplied with kanamycin (30 µg/mL)). The calculated amount of pre-culture was transferred to the main culture which was incubated until an OD_{600} of 0.6 was reached, at which point the protein expression was initiated by adding 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Expression took place at 25 °C overnight while shaking (170 rpm).
The next morning, the cells were harvested by centrifugation (5,000 × g, 10 min, 4 °C), the pellet was frozen in liquid nitrogen and stored at -80 °C.

Cell lysis by high-pressure homogenization

The cell pellet was thawed on ice in 30 mL lysis buffer (4.6) and 2 mM dithiothreitol (DTT) was added to the mixture. A douncer (Carl Roth) was used to achieve a homogeneous suspension. Lysis took place in an Emulsiflex-C5 high-pressure homogenizer (Avestin) and the lysates were centrifuged (15,000 × g, 30 min, 4 °C). The clear supernatant was subsequently exposed to chromatographic purification.

Table 4.6: Buffer composition for cell lysis and protein purification.

Buffer	Contents
HisTrap running buffer	300 mM NaCl, 50 mM Tris, 10 % glycerine, pH 7.5
HisTrap elution buffer	HisTrap running buffer + 500 mM imidazole
Dialysis buffer	100 mM KCl, 20 mM HEPES, 20 % glycerine, pH 7.9
Lysis buffer	HisTrap running buffer, 1 mM Ca^{2+}, 5 mM Mg^{2+}, 2 µL/30 mL DNAse, 600 µL/30 mL protease inhibitor*
*Protease inhibitor	cOmplete™, EDTA-free protease inhibitor cocktail (Roche), 50× stock solution: one tablet dissolved in 1 mL water

Immobilized metal ion affinity chromatography (IMAC)

Supernatant from cell lysis was filtered (0.45 µm syringe filter (Sarstedt)) and introduced to a 5 mL HisTrap High Performance column (Cytiva) via an ÄKTAprime plus system (Cytiva). Ni-NTA column purification took place at 4 °C and at a flowrate of 5 mL/min using the running and elution buffers described in 4.6. Elution took place at a gradient from 8-100 % elution buffer in 16 min and the fractions (4 mL) were afterward analyzed via coomassie-stained sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) (4.2.4.6).

Dialysis, concentration and storage

Pure fractions were combined and dialyzed using a Spectra/Por® 1 dialysis tube with a cut-off of 6-8 kDa (Spectrum) against dialysis buffer (4.6) overnight at 4 °C. Amicon® Ultra-4 centrifugal filters (30 kDa cut-off, Merck) were employed to concentrate the protein solutions by repeating short centrifuge runs (5,000 × g, 10 min, 4 °C) until sufficient concentration was achieved. The concentration was measured using the NanoDrop™ One microvolume UV-Vis spectrophotometer (Thermo Fisher). Final protein solutions were aliquoted and stored at -20 °C.

4.2.4.4 Pull-down experiments

Pull-downs (PDs) were conducted using the recombinant proteins BRD3(2)-TurboYFP, BRD4(1)-TagGFP2, BRD4(2)-TurboYFP, NAA10-YFP (established by Dr. Sören Kirchgässner) and BRD3(1)-mKate2, BRD3-BFP, BRD4-BFP (established in this work together with Yandan Zhao) as well as HeLa whole cell extract (WCE) (established by Dr. Julia Sindlinger).

Pull-down experiments with resin-bound peptides and recombinant protein

Pull-down experiments using resin-bound peptide included equilibrating the beads, incubation with protein solution, washing of the resin and finally elution of the bound protein. Since BRDs showed degradation on microcentrifuge filter units, PDs with these proteins were done using Eppendorf tubes (1.5 mL, Eppendorf). Between every equilibration and washing step, the beads were centrifuged (5,000 rpm, 2 min) and the supernatant was discarded. The PD-buffers used included 20 mM HEPES, 20 % glycerine at pH 7.9 and a varying concentration of KCl (4.8).

One aliquot of immobilized peptide (40 µL) was equilibrated 3 × with 200 µL PD-buffer on a shaker (700 rpm) at room temperature. The beads were subsequently incubated

with 200 µL incubation buffer (respective protein in PD-buffer + 10 mg/mL bovine serum albumin (BSA)) at 700 rpm. Afterward, they were washed with 200 µL of pull-down buffer with varying conditions (4.8). The beads were then eluted by adding 20 µL 3× SDS sample buffer (250 mM Tris, 10 % SDS, 30 % glycerine, 0.5 M DTT, 0.012 % bromophenol blue) and heating at 95 °C for 10 min. Following that, the samples were centrifuged (5,000 rpm, 2 min) and the supernatant was stored at -20 °C or directly analyzed by SDS-PAGE and Coomassie staining. Inputs for SDS-PAGE were created by incubating 2 µg protein with 3 µL 5× SDS sample buffer at 95 °C for 10 min. The specific conditions for equilibration, incubation and wash for each protein are listed in 4.8.

Table 4.8: Pull-down conditions for different proteins.

Protein	Equilibration	Incubation	Wash
BD1	3 × 2 min, 100 mM KCl	21 °C, 1 h, 100 mM KCl, 6.8 µM protein	4 × 2 min, 300 mM KCl, 2 × 2 min 100 mM KCl
BD2	3 × 2 min, 100 mM KCl	21 °C, 1 h, 100 mM KCl, 10.5 µM protein	4 × 2 min, 300 mM KCl, 2 × 2 min 100 mM KCl
BRD3/4	3 × 2 min, 300 mM KCl	4 °C, 2 h, 300 mM KCl, 1.05 µM protein	6 × 5 min, 600 mM KCl, 2 × 2 min 100 mM KCl
NAA10	3 × 2 min, 100 mM KCl	21 °C, 1 h, 100 mM KCl, 5.0 µM protein	4 × 2 min, 300 mM KCl, 2 × 2 min 100 mM KCl

Pull-down experiments with resin-bound peptides and HeLa WCE

For pull-downs with HeLa WCE, one aliquot (40 µL) of the resin-bound peptide was transferred into a microcentrifuge filter unit (Ultrafree-MC-HV, Merck) using 200 µL PD-buffer (100 mM KCl, 20 mM HEPES, 20 % glycerine, pH 7.9). The filter units were centrifuged (1,700 × g, 2 min), the filtrate was discarded and the beads were equilibrated with 200 µL PD-buffer by incubating the beads at first at 550 rpm and room temperature for 2 min, followed by a centrifugation at 1,700 × g for 2 min. This step was repeated two times. The HeLa WCE (c = 3.13 µg/µL) was diluted to a concentration of 1 µg/µL using PD-buffer. The resin was incubated with 200 µL of the diluted WCE at room temperature while shaking (550 rpm) for 1 h. After the incubation, the filter units were again centrifuged, the flow-through was discarded and the resin was washed with pull-down

wash buffer (300 mM KCl, 20 mM HEPES, 20 % glycerine, pH 7.9, 6 × 500 µL) under agitation (550 rpm) at room temperature for 5 min and subsequent centrifugation at 1,700 × g for 2 min. For each peptide, three biological replicates were created by performing pull-downs independently with three aliquots.
Following that, elution for label-free proteomics was conducted by adding 20 µL 3× SDS sample buffer without bromophenol blue (150 mM Tris, 6 % SDS, 20 % glycerine, 0.3 M DTT) to each sample. Then, the mixtures were incubated at 95 °C while agitating at 550 rpm for 10 min. The input samples were created by mixing 20 µL of the diluted HeLa WCE (c = 1 µg/µL) with 30 µL of 5× SDS sample buffer without bromophenol blue (250 mM Tris, 10 % SDS, 30 % glycerine, 0.5 M DTT) and treating them the same way as described above. The samples were then eluted by centrifuging the filter units at 9,400 × g for 5 min and the flow-through as well as the inputs were stored at 4 °C.

Pull-down experiments with membrane-bound peptides

Membranes with immobilized peptides were equilibrated 4 × 2 min with PD-buffer (100 mM KCl, 20 mM HEPES, 20 % glycerine, pH 7.9). The volume used each time was approximately 10 mL, enough to cover the membrane. Afterward, the membrane was shrink-wrapped and incubated with the respective protein in PD-buffer + 10 mg/mL BSA. BRD3(2) and BRD4(2) were incubated for 1 h at room temperature at a concentration of 10.5 µM, while BRD3(1) and BRD4(1) were incubated overnight at 4 °C at a concentration of 6.8 µL. The membrane was subsequently rinsed 3 × 2 min with a washing buffer (PD-buffer with 300 mM KCl) and the protein spots were visualized using the ChemiDoc MP Imager (BioRad). The program Alexa 488 Blot, 532/28, Blue Epi was used for visualization of the fluorophores.

4.2.4.5 Membrane regeneration

With the goal of being able to test bindings of several BRDs for one membrane, the proteins were removed after each pull-down by applying a recycling protocol. Two recycling mixtures were prepared, recycling mixture A (SM-A) (8 M urea, 1 % SDS in phosphate-buffered saline (PBS)(10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl), pH 7.0) and recycling mixture mix B (SM-B) (40 % water, 50 % ethanol, 10 % acetic acid). 0.5 % 2-mercaptoethanol was added to SM-A prior to use.
Directly after the incubation of the membrane with the protein solution and the following fluorescence measurement, the membrane was put in water and stored at room temperature overnight. The next day, the membrane was washed twice with 20 mL of DMF for 10 min

and three times with 20 mL of water for 10 min. Afterward, it was washed three times with 20 mL SM-A in a sonication bath at 40 °C for 10 min. SM-A was discarded and washing continued with 20 mL SM-B (3 × 10 min) and 20 mL ethanol (3 × 10 min). After the last washing step with ethanol, the membrane was left to air dry for 15 min and stored in a freezer at -20 °C.

4.2.4.6 SDS-PAGE and gel staining

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis () was performed on 12 % gels (percentage of acrylamide in the separation gel), consisting of a stacking gel and a separation gel (4.10). If not denatured already, samples and inputs (2 µg) were subjected to 10 min incubation at 95 °C with 5× SDS sample buffer(4.10). Marker (PageRuler™ Prestained Protein Ladder #SM0671, Thermo Fisher), denatured input and denatured samples were loaded onto polyacrylamide gels. The gel cassette containing the gels was put into the electrophoresis chamber and both were filled with SDS-PAGE running buffer(4.10). The separation was carried out at 140 V for 60 min.

Table 4.10: SDS-PAGE components and their contents.

Component	Contents
5× SDS sample buffer	250 mM Tris, 10 % SDS, 30 % glycerine, 0.5 M DTT, 0.02 % bromophenol blue
Polyacrylamide stacking gel	250 mM Tris pH 6.8, 4 % acrylamide/bisacrylamide (37.5:1), 0.1 % SDS, 0.1 % TEMED*, 0.04 % APS†
Polyacrylamide separation gel	375 mM Tris pH 8.8, 4 % acrylamide/bisacrylamide (37.5:1), 0.1 % SDS, 0.1 % TEMED, 0.04 % APS
SDS-PAGE running buffer	25 mM Tris pH 8.8, 192 mM glycine, 0.1 % SDS

*Tetramethylethylenediamine (TEMED), †Ammonium persulfate (APS)

All gels that were not used for Western blotting 4.2.4.7 were stained with Coomassie. After the gel electrophoresis, the gels were put in water, heated to boiling using a microwave and put on a shaker for 5 min. This step was repeated once and the gels were subsequently treated with Coomassie Brilliant Blue staining solution (80 mg Coomassie Brilliant Blue G250 (Sigma Aldrich), HCl (37 % in water, 3 mL) in 1 L water), heated until boiling and incubated for 30 min while shaking. Destaining was achieved by incubating the gels with

water. The ChemiDoc MP system (application: Coomassie Blue, excitation source: White Trans Illumination) was then used to take an image of the gels.

4.2.4.7 Western blotting

Western blotting was performed to transfer the proteins from SDS-PAGE gels to polyvinylidene fluoride (PVDF) membranes using the Mini Trans-Blot® system from Bio-Rad. All steps were done at room temperature if not stated otherwise. At first, the PVDF membrane was incubated with methanol (1 × 15 s), water (1 × 2 min) and blotting buffer (25 mM Tris, 192 mM glycine, 0.05 % SDS, 10 % methanol in water) (1 × 15 min) and the SDS-PAGE gels were equilibrated in blotting buffer for 30 min. The gel and the PVDF membrane were assembled between the plate electrodes, the electrophoresis chamber was filled with blotting buffer and the blotting was conducted at a current of 25 mA overnight at 4 °C. Afterward, the PVDF membrane was blocked by incubating it with blocking buffer (5 % low fat powdered milk in PBS (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4)) for 30 min and washed with PBST (0.1 % Tween 20 (v/v) in PBS) twice for 10 s each. Primary and secondary antibody solutions were created by diluting the antibodies (see 4.12) separately in primary antibody buffer (2.5 % low fat powdered milk in 50 mL PBST) and secondary antibody buffer (5 % low fat powdered milk in 50 mL PBST, 0.01 % SDS), respectively.

Table 4.12: Antibodies used in Western blotting.

Primary antibodies	**Isotype**	**Dilution**	**Item No.**	**Supplier**
NAA10	rabbit-poly-lgG	1:500	14803-1-AP	Proteintech
NAA25	rabbit-poly-lgG	1:250	ab79490	Abcam
NAA50	rabbit-poly-lgG	1:500	ab241487	Abcam
NAA80	rabbit-poly-lgG	1:500	PA5-21588	Invitrogen
Secondary antibody	**Isotype**	**Dilution**	**Item No.**	**Supplier**
donkey anti-rabbit	IgG-IRDye® 800 CW conjugate	1:10000	925-32213	Li-cor

Membranes were incubated with primary antibody solutions for 1 h at room temperature while shaking and then washed with PBST (3 × 10 s then 3 × 5 min). Incubation with secondary antibody solution was performed in the same way as with primary antibody

solutions. After washing again with PBST (3 × 10 s then 3 × 5 min), the membrane was washed with PBS (2 × 10 s) and fluorescence was recorded using the ChemiDoc MP Imaging system (Bio-Rad).

4.2.4.8 Filter aided sample preparation (FASP)

Eluates for label-free proteomics (4.2.4.4) were diluted with 450 µL U_A buffer (8 M urea, 100 mM Tris, pH 8.5) and transferred to a Microcon YM-10 centrifugal filter (Merck). The filter units were centrifuged at 13,900 × g for 10 min and the filtrate was discarded. Afterward, the samples were washed 3 × with 450 µL U_A buffer, each time followed by centrifugation (13,900 × g, 10 min) and discarding of the filtrate. 100 µL 55 mM chloroacetamide in U_A buffer was added to the samples and the filter units were incubated in the dark for 20 min. The chloroacetamide was removed by centrifugation (13,900 × g, 7.5 min) and the filter units were again washed with 100 µL U_A buffer and centrifuged at 13,900 × g for 5 min. The washing step was repeated twice and lastly new collection tubes were installed.
For the enzymatic digestion 1 µg lysyl endopeptidase (Lys-C, Wako Chemicals GmbH) (0.5 µg/µL Lys-C in 50 mM ammonium bicarbonate (ABC) buffer) in 40 µL U_A buffer was added to the concentrated protein solution. Incubation took place on a shaker (300 rpm) at room temperature for 1 h. Lys-C digestion was repeated by adding 2 µL of the Lys-C stock solution (0.5 µg/µL in 50 mM ABC buffer) and incubating the mixture overnight at room temperature while shaking (300 rpm). The protein solution was diluted with ABC buffer (300 µL) and 0.2 µg trypsin (1 µg/µL trypsin stock solution in 1 mM HCl) was added. Incubation took place at 37 °C for 3.5 h while shaking (300 rpm). Trypsination was repeated under the same conditions by adding the same amount of trypsin. TFA was added to the collection tubes so that the final concentration was 1 % and the digested peptide solution was centrifuged at 13,900 × g for 10 min. Then, 50 µL ABC buffer was added to the filter unit and centrifugation was repeated. The acidified eluates were directly subjected to desalting and concentrating using StageTips.

4.2.4.9 StageTips

To desalt and concentrate the peptide solutions, stop-and-go-extraction tips (StageTips) were employed. Three layers of two different Empore™ extraction disk material 47 mm (3M) were cut out and transferred to a 200 µL pipette tip using a Hamilton syringe. For pull-down samples, C_{18} material (C_{18} (octadecyl-bonded silica) item number: 2215) was

used and for input samples, SDB material (Poly(styrene-divenylbenzene) copolymer, item number: 2240) was used. Different protocols were employed for each material.
SDB StageTips were successively equilibrated with 100 µL ACN, 100 µL buffer C (30 % methanol, 1 % TFA in water), and 100 µL buffer D (0.2 % TFA in water). After the addition of each buffer, the StageTips were centrifuged at 1,300 × g for 2 min using a tip-to-reaction tube adapter. The respective input samples were loaded onto the tips by centrifugation at 500 × g for 3 min and 600 × g for 1 min. The filter tips were then washed by adding 100 µL buffer D followed by centrifugation at 1,300 × g for 2 min. Eppendorf reaction tubes for collecting the samples were changed and elution took place in three consecutive steps. For the first elution, 20 µL buffer 1 (100 mM NH_4HCO_2, 40 % ACN, 0.5 % FA in water) was added, for the second elution 20 µL buffer 2 (150 mM NH_4HCO_2, 60 % ACN, 0.5 % FA in water) was added and for the third elution 20 µL buffer 3 (5 % ammonia, 80 % ACN in water) was used. Between each step, the fraction was eluted into a new collection vial by centrifuging for 5 min at 500 × g. All three fractions were diluted with 20 µL water and lyophilised. C_{18} StageTips were successively equilibrated with 60 µL methanol, 60 µL buffer B (80 % ACN, 0.5 % FA in water) and 60 µL buffer A (0.5 % FA in water). After the addition of each buffer, the StageTips were centrifuged at 1,300 × g for 2 min using a tip-to-reaction tube adapter. The respective pull-down samples were loaded onto the tips by centrifugation at 500 × g for 3 min and 600 × g for 1 min. The filter tips were then washed twice with 30 µL buffer A and subsequent centrifugation at 1,300 × g for 2 min. Pull-down samples were eluted into new collection tubes by adding 2 × 20 µL buffer B and centrifuging at 500 × g for 5 min each.

4.2.4.10 Nano-LC-MS/MS

Peptides from digested pull-down and input eluates were separated on an EASY-nlc™ 1200 UHPLC system (Thermo Fisher) equipped with a 20 cm column encasing a ReproSil-Pur C_{18}-AQ 1.9 µm resin (Dr. Maisch). The UHPLC system was coupled to a nano-electrospray source leading to an Exploris 480 mass spectrometer (Thermo Fisher).
Total peptides (0.5 µg) were separated over a segmented linear gradient from 3-80 % buffer B (0.5 % formic acid in ACN) at 50 °C for 120 min. The survey scans were acquired in a mass range of 380 – 1500 m/z, at a resolution of 120,000 at 200 m/z and setting the automatic gain control target value to 3×10^6 in the orbitrap, before acquiring MS/MS scans in data-dependent mode. Fragmentation was induced by selecting the 20 most intense ions with an isolation width of 1.6 m/z and fragmenting them at a collision energy of 30 EV. Molecules with a charge of `+1` or higher than `+5` were excluded from fragmentation.

The corresponding spectra were recorded at a target value of 5×10^4 and a resolution of 15,000. Nano-LC-MS/MS experiments were carried out by Dr. Jürgen Eirich (group of Prof. Dr. Iris Finkemeier, University of Münster).

4.2.4.11 Data processing and quantification

The MaxQuant software package (Version 2.0.3.0, `https://www.maxquant.org/`) was used to process the MS/MS raw data utilizing the human reference proteome UP000005649_9606 (`https://www.uniprot.org/`) and an internal database containing standard contaminants. Search against the reference proteome was performed with full trypsin specificity, the threshold for missed cleavages was set to a maximum of two and the protein false discovery rate was set to 1 %. Due to the treatment with chloroacetamide during the FASP protocol (4.2.4.8), the carbamidomethylation of cysteine residues was set as fixed, whereas oxidation of methionine and N-terminal acetylation was set as variable. Minimal peptide length was set to seven amino acids, all other search parameters remained unchanged. For the label-free quantification (LFQ), the MaxLFQ algorithm integral to MaxQuant was applied with match between runs enabled and the LFQ minimum ratio count was set to 1. Raw data processing using MaxQuant was carried out by Dr. Jürgen Eirich (group of Prof. Dr. Iris Finkemeier, University of Münster).
The processed data was analyzed using the Perseus software package (Version 1.16.15.0). Potential contaminants, proteins only identified including modified amino acids and reverse sequences were removed and the LFQ intensity of remaining proteins was $\log_2$ transformed. Hits were then grouped and proteins that were only found in one of the three replicates were filtered and discarded. Remaining missing values were imputed for each column separately from the normal distribution, using a width of 0.4 as well as a down shift of 1.6. Before subjecting the data to statistical analysis using R, it was normalized by subtracting the median from each value. The matrices were exported and the p-value as well as the $\log_2$-fold change of the average of the three replicates of each protein was determined using the limma package in R (Version 3.5.0). Finally, volcano plots were created with Microsoft Excel by plotting the -$\log_{10}$(p-value) against the $\log_2$-fold change.

Acknowledgements

I would like to express my deepest gratitude to my thesis supervisor, Prof. Dr. Dirk Schwarzer, for giving me the opportunity to conduct my research in his laboratory and the interesting topic that comprises so many areas of organic chemistry and biochemistry. His expertise and insights have been truly helpful and never failed to help stir my work into the right direction.

I would also like to thank Prof. Dr. Gabriele Dodt for being my second supervisor.

Many thanks to Prof Dr. Iris Finkemeier and Dr. Jürgen Eirich from the Westfälische Wilhelms-Universität Münster for the good collaboration. I want to especially thank Jürgen who was always eager to help me with the analysis of the proteomics data and who always gave valuable information and insights.

Without the heart of the group, my colleagues, the time I have spent in the lab would not have been as exciting and fun as it was. I am very thankful to Luisa for all the good talks and time we had and especially for proofreading my thesis, thank you so much! Thank you Sören for all the valuable help in the lab, if it were methods, tools or devices, you always shared your knowledge and I am very grateful for that. Thank you Yandan for expanding our culinary expertise and thank you for the good talks on various topics. I could have lived without getting to know the Durian fruit though. Thank you Julia for the amazing time and the wonderful conversations as well as the move from the old IFIB to the new IFIB, which was truly a unique experience. I am also thankful to Julian who has provided me with valuable feedback and was always there to answer my questions, even if it meant distracting him from writing his own thesis. Thank you Katharina for all the help in the lab and for introducing me to the art of cell counting. Vitasta, thank you for the great time we had, the unconditional help, the encouragement in times of great need and for all the other things that do not fit on one page.

Lastly, I would like to thank my family for their never-ending support and for always believing in me. Thank you, from the bottom of my heart.

Bibliography

1. Brown, T. A. *Genomes. 2nd edition.* (2002).
2. Piovesan, A. *et al.* On the length, weight and GC content of the human genome. *BMC Research Notes* **12** (Feb. 2019).
3. Alberts, B. *et al. Molecular biology of the cell. (6 ed.)* (2015).
4. Luger, K. *et al.* Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389,** S. 251–260 (1997).
5. Richmond, T. J. *et al.* The structure of DNA in the nucleosome core. *Nature* **423,** 145–150 (May 2003).
6. Maeshima, K. *et al.* Chromatin structure: does the 30-nm fibre exist in vivo? *Current Opinion in Cell Biology* **22,** 291–297 (June 2010).
7. Tremethick, D. J. Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell* **128,** 651–654 (Feb. 2007).
8. Chen, P. *et al.* New insights into the helical structure of 30-nm chromatin fibers. *Protein & Cell* **5,** 489–491 (June 2014).
9. Poonperm, R. *et al.* Chromosome scaffold is a double-stranded assembly of scaffold proteins. *Scientific reports* **5,** 1–10 (2015).
10. Babu, A. *et al.* in *International Review of Cytology* 1–60 (Elsevier, 1987).
11. Saksouk, N. *et al.* Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin* **8** (Jan. 2015).
12. Oberdoerffer, P. *et al.* The role of nuclear architecture in genomic instability and ageing. *Nature reviews Molecular cell biology* **8,** 692–702 (2007).
13. Felsenfeld, G. *et al.* Controlling the double helix. *Nature* **421,** 448–453 (2003).
14. Davey, C. A. *et al.* Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 Å Resolution. *Journal of Molecular Biology* **319,** 1097–1113 (June 2002).
15. Bannister, A. J. *et al.* Regulation of chromatin by histone modifications. *Cell Research* **21,** 381–395 (Feb. 2011).

16. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics* **39,** 457–466 (Mar. 2007).

17. Bowman, G. D. *et al.* Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chemical Reviews* **115,** 2274–2295 (Nov. 2014).

18. Huang, H. *et al.* SnapShot: histone modifications. *Cell* **159,** 458–458 (2014).

19. Rodríguez-Paredes, M. *et al.* Cancer epigenetics reaches mainstream oncology. *Nature medicine* **17,** 330–339 (2011).

20. Hassa, P. O. *et al.* Nuclear ADP-Ribosylation Reactions in Mammalian Cells: Where Are We Today and Where Are We Going? *Microbiology and Molecular Biology Reviews* **70,** 789–829 (Sept. 2006).

21. Spange, S. *et al.* Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *The International Journal of Biochemistry & Cell Biology* **41,** 185–198 (Jan. 2009).

22. Watson, J. D. *Molecular Biology of the Gene* ISBN: 9780321762436 (Pearson, 2014).

23. Rossetto, D. *et al.* Histone phosphorylation. *Epigenetics* **7,** 1098–1108 (Oct. 2012).

24. Udugama, M. *et al.* Histone H3.3 phosphorylation promotes heterochromatin formation by inhibiting H3K9/K36 histone demethylase. *Nucleic Acids Research* **50,** 4500–4514 (Apr. 2022).

25. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129,** 823–837 (May 2007).

26. Strahl, B. D. *et al.* The language of covalent histone modifications. *Nature* **403,** 41–45 (Jan. 2000).

27. Bauer, U.-M. *et al.* Methylation at arginine 17 of histone H3 is linked to gene activation. *EMBO reports* **3,** 39–44 (Jan. 2002).

28. Itzen, F. *et al.* Brd4 activates P-TEFb for RNA polymerase II CTD phosphorylation. *Nucleic Acids Research* **42,** 7577–7590 (May 2014).

29. Wood, A. *et al.* in *Advances in Protein Chemistry* 201–222 (Elsevier, 2004).

30. Pedersen, M. T. *et al.* Histone demethylases in development and disease. *Trends in Cell Biology* **20,** 662–671 (Nov. 2010).

31. Tamaru, H. *et al.* Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa. *Nature Genetics* **34,** 75–79 (Apr. 2003).

32. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389,** 349–352 (1997).

33. Hollebeke, J. *et al.* N-terminal acetylation and other functions of N^{α}-acetyltransferases. *bchm* **393,** 291–298 (Apr. 2012).

34. Starheim, K. K. *et al.* Knockdown of Human N^{α}-Terminal Acetyltransferase Complex C Leads to p53-Dependent Apoptosis and Aberrant Human Arl8b Localization. *Molecular and Cellular Biology* **29,** 3569–3581 (July 2009).

35. Scott, D. C. *et al.* N-Terminal Acetylation Acts as an Avidity Enhancer Within an Interconnected Multiprotein Complex. *Science* **334,** 674–678 (Nov. 2011).

36. Drazic, A. *et al.* The world of protein acetylation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1864,** 1372–1401 (Oct. 2016).

37. Seto, E. *et al.* Erasers of Histone Acetylation: The Histone Deacetylase Enzymes. *Cold Spring Harbor Perspectives in Biology* **6,** a018713–a018713 (Apr. 2014).

38. Graf, L. G. *et al.* Assays to Study Enzymatic and Non-Enzymatic Protein Lysine Acetylation *in vitro*. *Current Protocols* **1** (Nov. 2021).

39. Damme, P. V. *et al.* NatF Contributes to an Evolutionary Shift in Protein N-Terminal Acetylation and Is Important for Normal Chromosome Segregation. *PLoS Genetics* **7** (ed Snyder, M.) e1002169 (July 2011).

40. Damme, P. V. *et al.* Proteome-derived Peptide Libraries Allow Detailed Analysis of the Substrate Specificities of Nα-acetyltransferases and Point to hNaa10p as the Post-translational Actin Nα-acetyltransferase. *Molecular & Cellular Proteomics* **10,** M110.004580 (May 2011).

41. Helsens, K. *et al.* Bioinformatics Analysis of a *Saccharomyces cerevisiae* N-Terminal Proteome Provides Evidence of Alternative Translation Initiation and Post-Translational N-Terminal Acetylation. *Journal of Proteome Research* **10,** 3578–3589 (June 2011).

42. Aksnes, H. *et al.* in *International Review of Cell and Molecular Biology* 267–305 (Elsevier, 2015).

43. Allfrey, V. G. *et al.* National Academy of Sciences: Abstracts of Papers Presented at the Annual Meeting, Washington, D.C., 27-29 April 1964. *Science* **144,** 559–569 (May 1964).

44. Inoue, A. *et al.* Enzymatic deacetylation of histone. *Biochemical and Biophysical Research Communications* **36,** 146–150 (July 1969).

45. Allis, C. D. *et al.* New Nomenclature for Chromatin-Modifying Enzymes. *Cell* **131,** 633–636 (Nov. 2007).

46. Saha, R. N. *et al.* HATs and HDACs in neurodegeneration: a tale of disconcerted acetylation homeostasis. *Cell Death & Differentiation* **13,** 539–550 (Oct. 2005).

47. Haberland, M. *et al.* Genetic dissection of histone deacetylase requirement in tumor cells. *Proceedings of the National Academy of Sciences* **106,** 7751–7755 (May 2009).

48. Miyagi, M. in *Methods in Enzymology* 85–95 (Elsevier, 2017).

49. Dutta, R. *et al.* CBP/p300 acetyltransferase activity in hematologic malignancies. *Molecular Genetics and Metabolism* **119,** 37–43 (Sept. 2016).

50. Roth, S. Y. *et al.* Histone Acetyltransferases. *Annual Review of Biochemistry* **70,** 81–120 (June 2001).

51. Yuan, H. *et al.* Histone acetyltransferases: Rising ancient counterparts to protein kinases. *Biopolymers* **99,** 98–111 (Nov. 2012).

52. Berndsen, C. E. *et al.* Catalytic Mechanism of a MYST Family Histone Acetyltransferase. *Biochemistry* **46,** 623–629 (Dec. 2006).

53. Yan, Y. *et al.* The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. *Nature Structural Biology* (Sept. 2002).

54. Sterner, D. E. *et al.* Acetylation of Histones and Transcription-Related Factors. *Microbiology and Molecular Biology Reviews* **64,** 435–459 (June 2000).

55. Kimura, A. *et al.* A Decade of Histone Acetylation: Marking Eukaryotic Chromosomes with Specific Codes. *The Journal of Biochemistry* **138,** 647–662 (Dec. 2005).

56. Lee, K. K. *et al.* Histone acetyltransferase complexes: one size doesn't fit all. *Nature Reviews Molecular Cell Biology* **8,** 284–295 (Apr. 2007).

57. Kleff, S. *et al.* Identification of a Gene Encoding a Yeast Histone H4 Acetyltransferase. *Journal of Biological Chemistry* **270,** 24674–24677 (Oct. 1995).

58. Guan, K.-L. *et al.* Regulation of intermediary metabolism by protein acetylation. *Trends in Biochemical Sciences* **36,** 108–116 (Feb. 2011).

59. Yanginlar, C. *et al.* HDAC11 is a regulator of diverse immune functions. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1861,** 54–59 (Jan. 2018).

60. Tamkun, J. W. *et al.* brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2SWI2. *Cell* **68,** 561–572 (Feb. 1992).

61. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399,** 491–496 (June 1999).

62. Zeng, L. *et al.* Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466,** 258–262 (July 2010).

63. Su, D. *et al.* Structural basis for recognition of H3K56-acetylated histone H3–H4 by the chaperone Rtt106. *Nature* **483,** 104–107 (Feb. 2012).

64. Li, Y. *et al.* AF9 YEATS Domain Links Histone Acetylation to DOT1L-Mediated H3K79 Methylation. *Cell* **159,** 558–571 (Oct. 2014).

65. Jeanmougin, F. The bromodomain revisited. *Trends in Biochemical Sciences* **22,** 151–153 (May 1997).

66. Filippakopoulos, P. *et al.* Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **149,** 214–231 (Mar. 2012).

67. Muller, S. *et al.* Bromodomains as therapeutic targets. *Expert Reviews in Molecular Medicine* **13** (Sept. 2011).

68. Filippakopoulos, P. *et al.* The bromodomain interaction module. *FEBS Letters* **586,** 2692–2704 (May 2012).

69. Goodwin, G. H. *et al.* The BAH domain, polybromo and the RSC chromatin remodelling complex. *Gene* **268,** 1–7 (May 2001).

70. Fujisawa, T. *et al.* Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nature Reviews Molecular Cell Biology* **18,** 246–262 (Jan. 2017).

71. Yang, X.-J. *et al.* A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* **382,** 319–324 (July 1996).

72. Fairbridge, N. A. *et al.* Cecr2 mutations causing exencephaly trigger misregulation of mesenchymal/ectodermal transcription factors. *Birth Defects Research Part A: Clinical and Molecular Teratology* **88,** 619–625 (Aug. 2010).

73. Zeng, L. *et al.* Structural Basis of Site-Specific Histone Recognition by the Bromodomains of Human Coactivators PCAF and CBP/p300. *Structure* **16,** 643–652 (Apr. 2008).

74. Hudson, B. P. *et al.* Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. *Journal of Molecular Biology* **304,** 355–370 (Dec. 2000).

75. Hassan, A. H. *et al.* Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. *Biochemical Journal* **402,** 125–133 (Jan. 2007).

76. Kalkhoven, E. CBP and p300: HATs for different occasions. *Biochemical Pharmacology* **68,** 1145–1155 (Sept. 2004).

77. Sun, H. *et al.* Solution structure of BRD7 bromodomain and its interaction with acetylated peptides from histone H3 and H4. *Biochemical and Biophysical Research Communications* **358,** 435–441 (June 2007).

78. Ciró, M. *et al.* ATAD2 Is a Novel Cofactor for MYC, Overexpressed and Amplified in Aggressive Tumors. *Cancer Research* **69,** 8491–8498 (Oct. 2009).

79. Revenko, A. S. *et al.* Chromatin Loading of E2F-MLL Complex by Cancer-Associated Coregulator ANCCA via Reading a Specific Histone Mark. *Molecular and Cellular Biology* **30,** 5260–5272 (Nov. 2010).

80. Oppikofer, M. *et al.* Expansion of the ISWI chromatin remodeler family with new active complexes. *EMBO reports* **18,** 1697–1706 (Aug. 2017).

81. Dou, Y. *et al.* Physical Association and Coordinate Function of the H3 K4 Methyltransferase MLL1 and the H4 K16 Acetyltransferase MOF. *Cell* **121,** 873–885 (June 2005).

82. Patel, A. *et al.* On the Mechanism of Multiple Lysine Methylation by the Human Mixed Lineage Leukemia Protein-1 (MLL1) Core Complex. *Journal of Biological Chemistry* **284,** 24242–24256 (Sept. 2009).

83. Schultz, D. C. *et al.* SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development* **16,** 919–932 (Apr. 2002).

84. Schultz, D. C. *et al.* Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2αsubunit of NuRD. *Genes & Development* **15,** 428–443 (Feb. 2001).

85. Masselink, H. *et al.* The adenovirus E1A binding protein BS69 is a corepressor of transcription through recruitment of N-CoR. *Oncogene* **19,** 1538–1546 (Mar. 2000).

86. Wassarman, D. A. *et al.* TAFII250. *Journal of Cell Science* **114,** 2895–2902 (Aug. 2001).

87. Wang, P. J. Functional substitution for TAFII250 by a retroposed homolog that is expressed in human spermatogenesis. *Human Molecular Genetics* **11,** 2341–2346 (Sept. 2002).

88. Huang, H. *et al.* Expression of theWdr9 gene and protein products during mouse development. *Developmental Dynamics* **227,** 608–614 (July 2003).

89. Chen, X. *et al.* Structural insights into preinitiation complex assembly on core promoters. *Science* **372** (Apr. 2021).

90. Wang, H. *et al.* Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. *Cell Research* **24,** 1433–1444 (Nov. 2014).

91. Hilton, T. L. *et al.* TAF1 Histone Acetyltransferase Activity in Sp1 Activation of the Cyclin D1 Promoter. *Molecular and Cellular Biology* **25,** 4321–4332 (May 2005).

92. Jacobson, R. H. *et al.* Structure and Function of a Human TAFII250 Double Bromodomain Module. *Science* **288,** 1422–1425 (May 2000).

93. Martinez, E. *et al.* Novel Cofactors and TFIIA Mediate Functional Core Promoter Selectivity by the Human TAFII150-Containing TFIID Complex. *Molecular and Cellular Biology* **18,** 6571–6583 (Nov. 1998).

94. Li, A. G. *et al.* An Acetylation Switch in p53 Mediates Holo-TFIID Recruitment. *Molecular Cell* **28,** 408–421 (Nov. 2007).

95. Shen, W. *et al.* Solution Structure of Human Brg1 Bromodomain and Its Specific Binding to Acetylated Histone Tails. *Biochemistry* **46,** 2100–2110 (Feb. 2007).

96. Singh, M. *et al.* Structural Ramification for Acetyl-Lysine Recognition by the Bromodomain of Human BRG1 Protein, a Central ATPase of the SWI/SNF Remodeling Complex. *ChemBioChem* **8,** 1308–1316 (July 2007).

97. Choudhary, C. *et al.* Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **325,** 834–840 (Aug. 2009).

98. Polesskaya, A. *et al.* Interaction between Acetylated MyoD and the Bromodomain of CBP and/or p300. *Molecular and Cellular Biology* **21,** 5312–5320 (Aug. 2001).

99. Taniguchi, Y. The Bromodomain and Extra-Terminal Domain (BET) Family: Functional Anatomy of BET Paralogous Proteins. *International Journal of Molecular Sciences* **17,** 1849 (Nov. 2016).

100. Shang, E. *et al.* Identification of unique, differentiation stage-specific patterns of expression of the bromodomain-containing genes Brd2, Brd3, Brd4, and Brdt in the mouse testis. *Gene Expression Patterns* **4,** 513–519 (Sept. 2004).

101. Bisgrove, D. A. *et al.* Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. *Proceedings of the National Academy of Sciences* **104,** 13690–13695 (Aug. 2007).

102. Zhou, Q. *et al.* The Yin and Yang of P-TEFb Regulation: Implications for Human Immunodeficiency Virus Gene Expression and Global Control of Cell Growth and Differentiation. *Microbiology and Molecular Biology Reviews* **70,** 646–659 (Sept. 2006).

103. Rahman, S. *et al.* The Brd4 Extraterminal Domain Confers Transcription Activation Independent of pTEFb by Recruiting Multiple Proteins, Including NSD3. *Molecular and Cellular Biology* **31,** 2641–2652 (July 2011).

104. Aksnes, H. *et al.* Co-translational, Post-translational, and Non-catalytic Roles of N-Terminal Acetyltransferases. *Molecular Cell* **73,** 1097–1114 (Mar. 2019).

105. Favrot, L. *et al.* Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation. *Biochemistry* **55,** 989–1002 (Feb. 2016).

106. Liszczak, G. *et al.* Molecular basis for N-terminal acetylation by the heterodimeric NatA complex. *Nature Structural & Molecular Biology* **20,** 1098–1105 (Aug. 2013).

107. Aksnes, H. *et al.* An Organellar Nα-Acetyltransferase, Naa60, Acetylates Cytosolic N Termini of Transmembrane Proteins and Maintains Golgi Integrity. *Cell Reports* **10,** 1362–1374 (Mar. 2015).

108. Aksnes, H. *et al.* First Things First: Vital Protein Marks by N-Terminal Acetyltransferases. *Trends in Biochemical Sciences* **41,** 746–760 (Sept. 2016).

109. Drazic, A. *et al.* NAA80 is actin's N-terminal acetyltransferase and regulates cytoskeleton assembly and cell motility. *Proceedings of the National Academy of Sciences* **115,** 4399–4404 (Mar. 2018).

110. Frottin, F. *et al.* The Proteomics of N-terminal Methionine Cleavage. *Molecular & Cellular Proteomics* **5,** 2336–2349 (Dec. 2006).

111. Damme, P. V. *et al.* N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. *Proceedings of the National Academy of Sciences* **109,** 12449–12454 (July 2012).

112. Deng, S. *et al.* Protein N-Terminal Acetylation: Structural Basis, Mechanism, Versatility, and Regulation. *Trends in Biochemical Sciences* **46,** 15–27 (Jan. 2021).

113. Aksnes, H. *et al.* Molecular determinants of the N-terminal acetyltransferase Naa60 anchoring to the Golgi membrane. *Journal of Biological Chemistry* **292,** 6821–6837 (Apr. 2017).

114. Dinh, T. V. *et al.* Molecular identification and functional characterization of the first Nα-acetyltransferase in plastids by global acetylome profiling. *PROTEOMICS* **15,** 2426–2435 (June 2015).

115. Knorr, A. G. *et al.* Ribosome-NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation. *Nature Structural & Molecular Biology* **26,** 35–39 (Dec. 2018).

116. Hole, K. *et al.* The Human N-Alpha-Acetyltransferase 40 (hNaa40p/hNatD) Is Conserved from Yeast and N-Terminally Acetylates Histones H2A and H4. *PLoS ONE* **6** (ed Imhof, A.) e24713 (Sept. 2011).

117. Gautschi, M. *et al.* The Yeast Nα-Acetyltransferase NatA Is Quantitatively Anchored to the Ribosome and Interacts with Nascent Polypeptides. *Molecular and Cellular Biology* **23,** 7403–7414 (Oct. 2003).

118. Arnesen, T. *et al.* The Chaperone-Like Protein HYPK Acts Together with NatA in Cotranslational N-Terminal Acetylation and Prevention of Huntingtin Aggregation. *Molecular and Cellular Biology* **30,** 1898–1909 (Apr. 2010).

119. Damme, P. V. *et al.* N-terminal acetylome analysis reveals the specificity of Naa50 (Nat5) and suggests a kinetic competition between N-terminal acetyltransferases and methionine aminopeptidases. *PROTEOMICS* **15,** 2436–2446 (June 2015).

120. Polevoda, B. *et al.* Nat3p and Mdm20p Are Required for Function of Yeast NatB Nα-terminal Acetyltransferase and of Actin and Tropomyosin. *Journal of Biological Chemistry* **278,** 30686–30697 (Aug. 2003).

121. Polevoda, B. *et al.* NatC Nα-terminal Acetyltransferase of Yeast Contains Three Subunits, Mak3p, Mak10p, and Mak31p. *Journal of Biological Chemistry* **276,** 20154–20159 (Jan. 2001).

122. Magin, R. S. *et al.* The Molecular Basis for Histone H4- and H2A-Specific Amino-Terminal Acetylation by NatD. *Structure* **23,** 332–341 (Feb. 2015).

123. Deng, S. *et al.* Molecular basis for N-terminal acetylation by human NatE and its modulation by HYPK. *Nature Communications* **11** (Feb. 2020).

124. Wiame, E. *et al.* NAT6 acetylates the N-terminus of different forms of actin. *The FEBS Journal* **285,** 3299–3316 (Aug. 2018).

125. Dikiy, I. *et al.* N-terminal Acetylation Stabilizes N-terminal Helicity in Lipid- and Micelle-bound α-Synuclein and Increases Its Affinity for Physiological Membranes. *Journal of Biological Chemistry* **289,** 3652–3665 (Feb. 2014).

126. Stefanis, L. α-Synuclein in Parkinson's Disease. *Cold Spring Harbor Perspectives in Medicine* **2,** a009399–a009399 (Dec. 2011).

127. Kang, L. *et al.* N-terminal acetylation of α-synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer. *Protein Science* **21,** 911–917 (June 2012).

128. Hershko, A. *et al.* Role of the alpha-amino group of protein in ubiquitin-mediated protein breakdown. *Proceedings of the National Academy of Sciences* **81,** 7021–7025 (Nov. 1984).

129. Hwang, C.-S. *et al.* N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. *Science* **327,** 973–977 (Feb. 2010).

130. Bachmair, A. *et al.* In Vivo Half-Life of a Protein Is a Function of Its Amino-Terminal Residue. *Science* **234,** 179–186 (Oct. 1986).

131. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Science* **20,** 1298–1345 (July 2011).

132. Shemorry, A. *et al.* Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. *Molecular Cell* **50,** 540–551 (May 2013).

133. Croft, T. *et al.* A functional link between NAD^+-homeostasis and N-terminal protein acetylation in Saccharomyces cerevisiae. *Journal of Biological Chemistry* **293,** 2927–2938 (Feb. 2018).

134. Myklebust, L. M. *et al.* Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. *Human Molecular Genetics* **24,** 1956–1976 (Dec. 2014).

135. Saunier, C. *et al.* Expanding the Phenotype Associated with NAA10-Related N-Terminal Acetylation Deficiency. *Human Mutation* **37,** 755–764 (May 2016).

136. Cheng, H. *et al.* Truncating Variants in NAA15 Are Associated with Variable Levels of Intellectual Disability, Autism Spectrum Disorder, and Congenital Anomalies. *The American Journal of Human Genetics* **102,** 985–994 (May 2018).

137. Neri, L. *et al.* NatB-mediated protein N-α-terminal acetylation is a potential therapeutic target in hepatocellular carcinoma. *Oncotarget* **8,** 40967–40981 (Apr. 2017).

138. Varland, S. *et al.* Identification of an alternatively spliced nuclear isoform of human N-terminal acetyltransferase Naa30. *Gene* **644,** 27–37 (Feb. 2018).

139. Pavlou, D. *et al.* Depletion of histone N-terminal-acetyltransferase Naa40 induces p53-independent apoptosis in colorectal cancer cells via the mitochondrial pathway. *Apoptosis* **21,** 298–311 (Dec. 2015).

140. Midorikawa, Y. *et al.* Identification of Genes Associated with Dedifferentiation of Hepatocellular Carcinoma with Expression Profiling Analysis. *Japanese Journal of Cancer Research* **93,** 636–643 (June 2002).

141. Wang, Z. *et al.* Inactivation of androgen-induced regulator ARD1 inhibits androgen receptor acetylation and prostate tumorigenesis. *Proceedings of the National Academy of Sciences* **109,** 3053–3058 (Feb. 2012).

142. Zeng, Y. *et al.* High expression of Naa10p associates with lymph node metastasis and predicts favorable prognosis of oral squamous cell carcinoma. *Tumor Biology* **37,** 6719–6728 (Dec. 2015).

143. Yi, C. H. *et al.* A genome-wide RNAi screen reveals multiple regulators of caspase activation. *Journal of Cell Biology* **179,** 619–626 (Nov. 2007).

144. Zeng, Y. *et al.* Inhibition of STAT5a by Naa10p contributes to decreased breast cancer metastasis. *Carcinogenesis* **35,** 2244–2253 (June 2014).

145. Kuo, H.-P. *et al.* ARD1 Stabilization of TSC2 Suppresses Tumorigenesis Through the mTOR Signaling Pathway. *Science Signaling* **3** (Feb. 2010).

146. Hua, K.-T. *et al.* N-α-Acetyltransferase 10 Protein Suppresses Cancer Cell Metastasis by Binding PIX Proteins and Inhibiting Cdc42/Rac1 Activity. *Cancer Cell* **19,** 218–231 (Feb. 2011).

147. French, C. A. *et al.* BRD4-NUT fusion oncogene: a novel mechanism in aggressive carcinoma. *Cancer research* **63,** 304–307 (Jan. 2003).

148. French, C. A. *et al.* BRD–NUT oncoproteins: a family of closely related nuclear proteins that block epithelial differentiation and maintain the growth of carcinoma cells. *Oncogene* **27,** 2237–2242 (Oct. 2007).

149. French, C. A. *et al.* BRD4 Bromodomain Gene Rearrangement in Aggressive Carcinoma with Translocation t(15;19). *The American Journal of Pathology* **159,** 1987–1992 (Dec. 2001).

150. Bauer, D. E. *et al.* Clinicopathologic Features and Long-term Outcomes of NUT Midline Carcinoma. *Clinical Cancer Research* **18,** 5773–5779 (Oct. 2012).

151. Reynoird, N. *et al.* Oncogenesis by sequestration of CBP/p300 in transcriptionally inactive hyperacetylated chromatin domains. *The EMBO Journal* **29,** 2943–2952 (July 2010).

152. Eagen, K. P. *et al.* Supercharging BRD4 with NUT in carcinoma. *Oncogene* **40,** 1396–1408 (Jan. 2021).

153. Gu, B. *et al.* Challenges and Opportunities in NUT Carcinoma Research. *Genes* **12,** 235 (Feb. 2021).

154. Nevola, L. *et al.* Modulating protein–protein interactions: the potential of peptides. *Chemical Communications* **51,** 3302–3315 (2015).

155. Wang, S. *et al.* Uncovering post-translational modification-associated protein–protein interactions. *Current Opinion in Structural Biology* **74,** 102352 (June 2022).

156. Craik, D. J. *et al.* The Future of Peptide-based Drugs. *Chemical Biology & Drug Design* **81,** 136–147 (Dec. 2012).

157. Muttenthaler, M. *et al.* Trends in peptide drug discovery. *Nature Reviews Drug Discovery* **20,** 309–325 (Feb. 2021).

158. Yoshida, M. *et al.* Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *Journal of Biological Chemistry* **265,** 17174–17179 (1990).

159. Ho, T. C. S. *et al.* Thirty Years of HDAC Inhibitors: 2020 Insight and Hindsight. *Journal of Medicinal Chemistry* **63,** 12460–12484 (July 2020).

160. Marks, P. A. *et al.* Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nature Biotechnology* **25,** 84–90 (Jan. 2007).

161. Hahnen, E. *et al.* Histone deacetylase inhibitors: possible implications for neurodegenerative disorders. *Expert Opinion on Investigational Drugs* **17,** 169–184 (Jan. 2008).

162. Machado-Vieira, R. *et al.* Histone Deacetylases and Mood Disorders: Epigenetic Programming in Gene-Environment Interactions. *CNS Neuroscience & Therapeutics* **17,** 699–704 (Oct. 2010).

163. Adcock, I. M. HDAC inhibitors as anti-inflammatory agents. *British Journal of Pharmacology* **150,** 829–831 (Apr. 2007).

164. Hay, D. A. *et al.* Discovery and Optimization of Small-Molecule Ligands for the CBP/p300 Bromodomains. *Journal of the American Chemical Society* **136,** 9308–9319 (June 2014).

165. Martin, L. J. *et al.* Structure-Based Design of an in Vivo Active Selective BRD9 Inhibitor. *Journal of Medicinal Chemistry* **59,** 4462–4475 (Mar. 2016).

166. Drouin, L. *et al.* Structure Enabled Design of BAZ2-ICR, A Chemical Probe Targeting the Bromodomains of BAZ2A and BAZ2B. *Journal of Medicinal Chemistry* **58,** 2553–2559 (Feb. 2015).

167. Fedorov, O. *et al.* Selective targeting of the BRG/PB1 bromodomains impairs embryonic and trophoblast stem cell maintenance. *Science Advances* **1** (Nov. 2015).

168. Doroshow, D. B. *et al.* BET inhibitors: a novel epigenetic approach. *Annals of Oncology* **28,** 1776–1787 (Aug. 2017).

169. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–1073 (Sept. 2010).

170. Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468,** 1119–1123 (Nov. 2010).

171. Filippakopoulos, P. *et al.* Targeting bromodomains: epigenetic readers of lysine acetylation. *Nature Reviews Drug Discovery* **13,** 337–356 (Apr. 2014).

172. Moyer, M. W. First drugs found to inhibit elusive cancer target. *Nature Medicine* **17,** 1325–1325 (Nov. 2011).

173. Amorim, S. *et al.* Bromodomain inhibitor OTX015 in patients with lymphoma or multiple myeloma: a dose-escalation, open-label, pharmacokinetic, phase 1 study. *The Lancet Haematology* **3,** e196–e204 (Apr. 2016).

174. Stathis, A. *et al.* Clinical Response of Carcinomas Harboring the BRD4–NUT Oncoprotein to the Targeted Bromodomain Inhibitor OTX015/MK-8628. *Cancer Discovery* **6,** 492–500 (May 2016).

175. Wong, C. *et al.* The bromodomain and extra-terminal inhibitor CPI203 enhances the antiproliferative effects of rapamycin on human neuroendocrine tumors. *Cell Death & Disease* **5,** e1450–e1450 (Oct. 2014).

176. Bhadury, J. *et al.* BET and HDAC inhibitors induce similar genes and biological effects and synergize to kill in Myc-induced murine lymphoma. *Proceedings of the National Academy of Sciences* **111** (June 2014).

177. Tontsch-Grunt, U. *et al.* Therapeutic impact of BET inhibitor BI 894999 treatment: backtranslation from the clinic. *British Journal of Cancer* **127,** 577–586 (Apr. 2022).

178. Gil, J. *et al.* Lysine acetylation and cancer: A proteomics perspective. *Journal of Proteomics* **150,** 297–309 (Jan. 2017).

179. Lau, O. D. *et al.* HATs off. *Molecular Cell* **5,** 589–595 (Mar. 2000).

180. Foyn, H. *et al.* Design, Synthesis, and Kinetic Characterization of Protein N-Terminal Acetyltransferase Inhibitors. *ACS Chemical Biology* **8,** 1121–1127 (Apr. 2013).

181. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **85,** 2149–2154 (July 1963).

182. Recio, C. *et al.* The Potential Therapeutic Application of Peptides and Peptidomimetics in Cardiovascular Disease. *Frontiers in Pharmacology* **7** (Jan. 2017).

183. Wang, X. *et al.* Rational Design of Peptide-Based Inhibitors Disrupting Protein-Protein Interactions. *Frontiers in Chemistry* **9** (May 2021).

184. Wang, L. *et al.* Therapeutic peptides: current applications and future directions. *Signal Transduction and Targeted Therapy* **7** (Feb. 2022).

185. Maack, T. *et al.* Renal filtration, transport, and metabolism of low-molecular-weight proteins: A review. *Kidney International* **16,** 251–270 (Sept. 1979).

186. Morrison, K. L. *et al.* Combinatorial alanine-scanning. *Current Opinion in Chemical Biology* **5,** 302–307 (June 2001).

187. Levin, A. M. *et al.* Optimizing the affinity and specificity of proteins with molecular display. *Mol. BioSyst.* **2,** 49–57 (2006).

188. Goldflam, M. *et al.* Recent Advances Toward the Discovery of Drug-Like Peptides De novo. *Frontiers in Chemistry* **3** (Dec. 2015).

189. Menegatti, S. *et al.* De Novo Design of Skin-Penetrating Peptides for Enhanced Transdermal Delivery of Peptide Drugs. *Advanced Healthcare Materials* **5,** 602–609 (Jan. 2016).

190. Fosgerau, K. *et al.* Peptide therapeutics: current status and future directions. *Drug Discovery Today* **20,** 122–128 (Jan. 2015).

191. Miranda, L. P. *et al.* Identification of Potent, Selective, and Metabolically Stable Peptide Antagonists to the Calcitonin Gene-Related Peptide (CGRP) Receptor. *Journal of Medicinal Chemistry* **51,** 7889–7897 (Nov. 2008).

192. Ricardo, M. G. *et al.* Multicomponent Peptide Stapling as a Diversity-Driven Tool for the Development of Inhibitors of Protein–Protein Interactions. *Angewandte Chemie International Edition* **59,** 5235–5241 (Mar. 2020).

193. Drucker, D. J. Advances in oral peptide therapeutics. *Nature Reviews Drug Discovery* **19,** 277–289 (Dec. 2019).

194. Sindlinger, J. *et al.* Investigating Peptide-Coenzyme A Conjugates as Chemical Probes for Proteomic Profiling of N-Terminal and Lysine Acetyltransferases. *ChemBioChem* **23** (July 2022).

195. Patel, K. *et al.* Cyclic peptides can engage a single binding pocket through highly divergent modes. *Proceedings of the National Academy of Sciences* **117,** 26728–26738 (Oct. 2020).

196. Kirchgäßner, S. *et al.* Synthesis, Biochemical Characterization, and Genetic Encoding of a 1,2,4-Triazole Amino Acid as an Acetyllysine Mimic for Bromodomains of the BET Family. *Angewandte Chemie International Edition* **62** (Feb. 2023).

197. Kotz, J. Phenotypic screening, take two. *Science-Business eXchange* **5,** 380–380 (Apr. 2012).

198. Bozovičar, K. *et al.* Evolving a Peptide: Library Platforms and Diversification Strategies. *International Journal of Molecular Sciences* **21,** 215 (Dec. 2019).

199. Mersich, C. *et al.* Generation of bioactive peptides by biological libraries. *Journal of Chromatography B* **861,** 160–170 (Jan. 2008).

200. Galán, A. *et al.* Library-based display technologies: where do we stand? *Molecular BioSystems* **12,** 2342–2358 (2016).

201. Boder, E. T. *et al.* Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology* **15,** 553–557 (June 1997).

202. Fukuda, I. *et al.* In vitro evolution of single-chain antibodies using mRNA display. *Nucleic Acids Research* **34,** e127–e127 (Sept. 2006).

203. Wang, H. *et al.* Advantages of mRNA display selections over other selection techniques for investigation of protein–protein interactions. *Expert Review of Proteomics* **8,** 335–346 (June 2011).

204. Lin, C.-W. *et al.* A Selection of Macrocyclic Peptides That Bind STING From an mRNA-Display Library With Split Degenerate Codons. *Angewandte Chemie International Edition* **60,** 22640–22645 (Sept. 2021).

205. Furka, Á. *et al.* General method for rapid synthesis of multicomponent peptide mixtures. *International Journal of Peptide and Protein Research* **37,** 487–493 (1991).

206. Gallop, M. A. *et al.* Applications of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries. *Journal of Medicinal Chemistry* **37,** 1233–1251 (Apr. 1994).

207. Ostresh, J. M. *et al.* Peptide libraries: Determination of relative reaction rates of protected amino acids in competitive couplings. *Biopolymers* **34,** 1681–1689 (Dec. 1994).

208. Houghten, R. A. General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proceedings of the National Academy of Sciences* **82,** 5131–5135 (Aug. 1985).

209. Pinilla, C. *et al.* Rapid identification of high affinity peptide ligands using positional scanning synthetic peptide combinatorial libraries. *Biotechniques* **13,** 901–905 (1992).

210. Gray, B. P. *et al.* Combinatorial Peptide Libraries: Mining for Cell-Binding Peptides. *Chemical Reviews* **114,** 1020–1081 (Dec. 2013).

211. Frank, R. Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* **48,** 9217–9232 (Jan. 1992).

212. Eichler, J. Synthetic Peptide Arrays and Peptide Combinatorial Libraries for the Exploration of Protein-Ligand Interactions and the Design of Protein Inhibitors. *Combinatorial Chemistry & High Throughput Screening* **8,** 135–143 (Mar. 2005).

213. Anderson, N. L. *et al.* Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis* **19,** 1853–1861 (Aug. 1998).

214. Bantscheff, M. *et al.* Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and Bioanalytical Chemistry* **404,** 939–965 (July 2012).

215. Bantscheff, M. *et al.* Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry* **389,** 1017–1031 (Aug. 2007).

216. Anand, S. *et al.* in *Methods in Molecular Biology* 31–43 (Springer New York, Dec. 2016).

217. Cox, J. *et al.* MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26,** 1367–1372 (Nov. 2008).

218. Cox, J. *et al.* Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics* **13,** 2513–2526 (Sept. 2014).

219. Conway, L. P. *et al.* Chemoproteomic-enabled phenotypic screening. *Cell Chemical Biology* **28,** 371–393 (Mar. 2021).

220. Berger, A. B. *et al.* Activity-Based Protein Profiling. *American Journal of PharmacoGenomics* **4,** 371–381 (2004).

221. Van de Merbel, N. C. Protein quantification by LC–MS: a decade of progress through the pages of Bioanalysis. *Bioanalysis* **11,** 629–644 (Apr. 2019).

222. Mann, M. Proteomics for biomedicine: a half-completed journey. *EMBO Molecular Medicine* **4,** 75–77 (Jan. 2012).

223. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–1073 (2010).

224. Stewart, K. M. *et al.* Cell-penetrating peptides as delivery vehicles for biology and medicine. *Organic & Biomolecular Chemistry* **6,** 2242 (2008).

225. Ueda, Y. *et al.* Induction of autophagic cell death of glioma-initiating cells by cell-penetrating d-isomer peptides consisting of Pas and the p53 C-terminus. *Biomaterials* **33,** 9061–9069 (Dec. 2012).

226. Mahato, R. I. *et al.* in *Advances in Genetics* 95–156 (Elsevier, 1999).

227. Brunner, K. Eine neue Darstellungsweise von Triazolen. *Monatshefte für Chemie* **36,** 509–534 (July 1915).

228. Subach, O. M. *et al.* An Enhanced Monomeric Blue Fluorescent Protein with the High Chemical Stability of the Chromophore. *PLoS ONE* **6** (ed Rao, J.) e28674 (Dec. 2011).

229. Pletneva, N. V. *et al.* Yellow fluorescent protein phiYFPv (Phialidium): structure and structure-based mutagenesis. *Acta Crystallographica Section D Biological Crystallography* **69,** 1005–1012 (May 2013).

230. Shcherbo, D. *et al.* Far-red fluorescent tags for protein imaging in living tissues. *Biochemical Journal* **418,** 567–574 (Feb. 2009).

231. Subach, O. M. *et al.* Conversion of Red Fluorescent Protein into a Bright Blue Probe. *Chemistry & Biology* **15,** 1116–1124 (Oct. 2008).

232. Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Research* **50,** D20–D26 (Dec. 2021).

233. Adessi, C. *et al.* Pharmacological Profiles of Peptide Drug Candidates for the Treatment of Alzheimer's Disease. *Journal of Biological Chemistry* **278,** 13905–13911 (Apr. 2003).

234. Yang, N. J. *et al.* in *Site-Specific Protein Labeling* 29–53 (Springer New York, Dec. 2014).

235. Derakhshankhah, H. *et al.* Cell penetrating peptides: A concise review with emphasis on biomedical applications. *Biomedicine & Pharmacotherapy* **108,** 1090–1096 (Dec. 2018).

236. Feni, L. *et al.* Cell-penetrating peptides containing 2,5-diketopiperazine (DKP) scaffolds as shuttles for anti-cancer drugs: conformational studies and biological activity. *Chemical Communications* **56,** 5685–5688 (2020).

237. Schneider, A. F. L. *et al.* Cellular uptake of large biomolecules enabled by cell-surface-reactive cell-penetrating peptide additives. *Nature Chemistry* **13,** 530–539 (Apr. 2021).

238. Herce, H. D. *et al.* Cell-permeable nanobodies for targeted immunolabelling and antigen manipulation in living cells. *Nature Chemistry* **9,** 762–771 (July 2017).

239. Szabó, I. *et al.* Cell-penetrating conjugates of pentaglutamylated methotrexate as potential anticancer drugs against resistant tumor cells. *European Journal of Medicinal Chemistry* **115,** 361–368 (June 2016).

240. Puckett, C. A. *et al.* Targeting a ruthenium complex to the nucleus with short peptides. *Bioorganic & Medicinal Chemistry* **18,** 3564–3569 (May 2010).

241. Ree, R. *et al.* Spotlight on protein N-terminal acetylation. *Experimental & molecular medicine* **50,** 1–13 (July 2018).

242. Haynes, W. M. *CRC Handbook of Chemistry and Physics* 97th ed., 5–89 (CRC Press, 2016).

243. Tian, F. *et al.* A Phage Display System with Unnatural Amino Acids. *Journal of the American Chemical Society* **126,** 15962–15963 (Nov. 2004).

244. Li, S. *et al.* In Vitro Selection of mRNA Display Libraries Containing an Unnatural Amino Acid. *Journal of the American Chemical Society* **124,** 9972–9973 (July 2002).

245. Passioura, T. *et al.* A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chemical Communications* **53,** 1931–1940 (2017).

246. Chatterjee, J. *et al.* N-Methylation of Peptides and Proteins: An Important Element for Modulating Biological Functions. *Angewandte Chemie International Edition* **52,** 254–269 (Nov. 2012).

247. Feng, Z. *et al.* Inspiration from the mirror: D-amino acid containing peptides in biomedical approaches. *Biomolecular Concepts* **7,** 179–187 (June 2016).

248. Joo, S.-H. Cyclic Peptides as Therapeutic Agents and Biochemical Tools. *Biomolecules and Therapeutics* **20,** 19–26 (Jan. 2012).

249. Gao, Y. *et al.* Direct Comparison of Linear and Macrocyclic Compound Libraries as a Source of Protein Ligands. *ACS Combinatorial Science* **17,** 190–195 (Feb. 2015).

250. Hobert, E. M. *et al.* Effective Molarity: Proximity as a Guiding Force in Chemistry and Biology. *Israel Journal of Chemistry* **53,** 567–576 (Aug. 2013).

251. Hunter, C. A. *et al.* What is Cooperativity? *Angewandte Chemie International Edition* **48,** 7488–7499 (Sept. 2009).

252. Krishnamurthy, V. M. *et al.* Dependence of Effective Molarity on Linker Length for an Intramolecular Protein-Ligand System. *Journal of the American Chemical Society* **129,** 1312–1320 (Jan. 2007).

253. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (July 2021).

254. Martin, R. M. *et al.* Principles of protein targeting to the nucleolus. *Nucleus* **6,** 314–325 (July 2015).

255. Dang, C. V. *et al.* Nuclear and nucleolar targeting sequences of c-erb-A, c-myb, N-myc, p53, HSP70, and HIV tat proteins. *Journal of Biological Chemistry* **264,** 18019–18023 (Oct. 1989).

256. López-Mirabal, H. R. *et al.* Redox characteristics of the eukaryotic cytosol. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1783,** 629–640 (Apr. 2008).

257. Dougherty, P. G. *et al.* Understanding Cell Penetration of Cyclic Peptides. *Chemical Reviews* **119,** 10241–10287 (May 2019).

258. Pillow, T. H. *et al.* Decoupling stability and release in disulfide bonds with antibody-small molecule conjugates. *Chemical Science* **8,** 366–370 (2017).

259. Dose, A. *et al.* Interrogating Substrate Selectivity and Composition of Endogenous Histone Deacetylase Complexes with Chemical Probes. *Angewandte Chemie International Edition* **55,** 1192–1195 (Dec. 2015).

260. Han, Y. *et al.* Multifunctional oral delivery systems for enhanced bioavailability of therapeutic peptides/proteins. *Acta Pharmaceutica Sinica B* **9,** 902–922 (Sept. 2019).

261. Räder, A. F. B. *et al.* Improving oral bioavailability of cyclic peptides by N-methylation. *Bioorganic & Medicinal Chemistry* **26,** 2766–2773 (June 2018).

262. Wei, L. *et al.* ToxIBTL: prediction of peptide toxicity based on information bottleneck and transfer learning. *Bioinformatics* **38** (ed Martelli, P. L.) 1514–1524 (Jan. 2022).

263. Gupta, S. *et al.* In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS ONE* **8** (ed Patterson, R. L.) e73957 (Sept. 2013).

264. Uchide, N. *et al.* Lactate Dehydrogenase Leakage as a Marker for Apoptotic Cell Degradation Induced by Influenza Virus Infection in Human Fetal Membrane Cells. *Intervirology* **52,** 164–173 (2009).

265. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *Journal of Immunological Methods* **65,** 55–63 (Dec. 1983).

266. Gupta, S. *et al.* in *Methods in Molecular Biology* 143–157 (Springer New York, Dec. 2014).

267. Moutal, A. *et al.* Differential neuroprotective potential of CRMP2 peptide aptamers conjugated to cationic, hydrophobic, and amphipathic cell penetrating peptides. *Frontiers in Cellular Neuroscience* **8** (Jan. 2015).

268. Xie, J. *et al.* Cell-Penetrating Peptides in Diagnosis and Treatment of Human Diseases: From Preclinical Research to Clinical Application. *Frontiers in Pharmacology* **11** (May 2020).

269. Goetze, S. *et al.* Identification and Functional Characterization of N-Terminally Acetylated Proteins in Drosophila melanogaster. *PLoS Biology* **7** (ed MacCoss, M. J.) e1000236 (Nov. 2009).

270. Arnesen, T. *et al.* Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proceedings of the National Academy of Sciences* **106,** 8157–8162 (May 2009).

271. Hong, B. S. *et al.* Crystal Structures of Human Pantothenate Kinases. *Journal of Biological Chemistry* **282,** 27984–27993 (Sept. 2007).

272. Singleton, R. S. *et al.* OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation. *Proceedings of the National Academy of Sciences* **111,** 4031–4036 (Feb. 2014).

273. Goto, Y. *et al.* Translation Initiation with Initiator tRNA Charged with Exotic Peptides. *Journal of the American Chemical Society* **131,** 5040–5041 (Mar. 2009).

274. Khalil, E. M. *et al.* Mechanism-based inhibition of the melatonin rhythm enzyme: Pharmacologic exploitation of active site functional plasticity. *Proceedings of the National Academy of Sciences* **96,** 12418–12423 (Oct. 1999).

275. Deng, Y. *et al.* Novel Bisubstrate Inhibitors for Protein N-Terminal Acetyltransferase D. *Journal of Medicinal Chemistry* **64,** 8263–8271 (June 2021).

276. Iskar, M. *et al.* Drug-Induced Regulation of Target Expression. *PLoS Computational Biology* **6** (ed Kitano, H.) e1000925 (Sept. 2010).

277. Burke, M. R. *et al.* Overcoming Cancer Drug Resistance Utilizing PROTAC Technology. *Frontiers in Cell and Developmental Biology* **10** (Apr. 2022).

278. Riddles, P. W. *et al.* in *Enzyme Structure Part I* 49–60 (Elsevier, 1983).

279. Sanger, F. *et al.* DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74,** 5463–5467 (Dec. 1977).

Appendix

Table 4.1: Sequences of recombinant proteins expressed in *E. coli*

His_6-Thrombin-BRD3(1)-TEV-mKate2-Strep

```
  1 MGSSHHHHHH SSGLVPRGSH MPEVSNPSKP GRKTNQLQYM QNVVVKTLWK HQFAWPFYQP
 61 VDAIKLNLPD YHKIIKNPMD MGTIKKRLEN NYYWSASECM QDFNTMFTNC YIYNKPTDDI
121 VLMAQALEKI FLQKVAQMPQ EEGGENLYFQ SGGGSMVSEL IKENMHMKLY MEGTVNNHHF
181 KCTSEGEGKP YEGTQTMRIK AVEGGPLPFA FDILATSFMY GSKTFINHTQ GIPDFFKQSF
241 PEGFTWERVT TYEDGGVLTA TQDTSLQDGC LIYNVKIRGV NFPSNGPVMQ KKTLGWEAST
301 ETLYPADGGL EGRADMALKL VGGGHLICNL KTTYRSKKPA KNLKMPGVYY VDRRLERIKE
361 ADKETYVEQH EVAVARYCDL PSKLGHRGGG WSHPQFEK
```

His_6-Thrombin-BRD3(2)-TEV-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-Thrombin-BRD4(1)-TEV-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-Thrombin-BRD4(2)-TEV-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-Thrombin-BRD3-TEV-mTagBFP2

```
1    MGSSHHHHHH SSGLVPRGSH MPEVSNPSKP GRKTNQLQYM QNVVVKTLWK HQFAWPFYQP
61   VDAIKLNLPD YHKIIKNPMD MGTIKKRLEN NYYWSASECM QDFNTMFTNC YIYNKPTDDI
121  VLMAQALEKI FLQKVAQMPQ EEVELLPPAP KGKGRKPAAG AQSAGTQQVA AVSSVSPATP
181  FQSVPPTVSQ TPVIAATPVP TITANVTSVP VPPAAAPPPP ATPIVPVVPP TPPVVKKKGV
241  KRKADTTTPT TSAITASRSE SPPPLSDPKQ AKVVARRESG GRPIKPPKKD LEDGEVPQHA
301  GKKGKLSEHL RYCDSILREM LSKKHAAYAW PFYKPVDAEA LELHDYHDII KHPMDLSTVK
361  RKMDGREYPD AQGFAADVRL MFSNCYKYNP PDHEVVAMAR KLQDVFEMRF AKMPGGENLY
421  FQSGGGSVSK GEELIKENMH MKLYMEGTVD NHHFKCTSEG EGKPYEGTQT MRIKVVEGGP
481  LPFAFDILAT SFLYGSKTFI NHTQGIPDFF KQSFPEGFTW ERVTTYEDGG VLTATQDTSL
541  QDGCLIYNVK IRGVNFTSNG PVMQKKTLGW EAFTETLYPA DGGLEGRNDM ALKLVGGSHL
601  IANAKTTYRS KKPAKNLKMP GVYYVDYRLE RIKEANNETY VEQHEVAVAR YCDLPSKLGH
661  KLN
```

His_6-Thrombin-BRD4-TEV-mTagBFP2

```
  1 MGSSHHHHHH SSGLVPRGSH MNPPPPETSN PNKPKRQTNQ LQYLLRVVLK TLWKHQFAWP
 61 FQQPVDAVKL NLPDYYKIIK TPMDMGTIKK RLENNYYWNA QECIQDFNTM FTNCYIYNKP
121 GDDIVLMAEA LEKLFLQKIN ELPTEETEIM IVQAKGRGRG RKETGTAKPG VSTVPNTTQA
181 STPPQTQTPQ PNPPPVQATP HPFPAVTPDL IVQTPVMTVV PPQPLQTPPP VPPQPQPPPA
241 PAPQPVQSHP PIIAATPQPV KTKKGVKRKA DTTTPTTIDP IHEPPSLPPE PKTTKLGQRR
301 ESSRPVKPPK KDVPDSQQHP APEKSSKVSE QLKCCSGILK EMFAKKHAAY AWPFYKPVDV
361 EALGLHDYCD IIKHPMDMST IKSKLEAREY RDAQEFGADV RLMFSNCYKY NPPDHEVVAM
421 ARKLQDVFEM RFAKMPDEGG ENLYFQSGGG SVSKGEELIK ENMHMKLYME GTVDNHHFKC
481 TSEGEGKPYE GTQTMRIKVV EGGPLPFAFD ILATSFLYGS KTFINHTQGI PDFFKQSFPE
541 GFTWERVTTY EDGGVLTATQ DTSLQDGCLI YNVKIRGVNF TSNGPVMQKK TLGWEAFTET
601 LYPADGGLEG RNDMALKLVG GSHLIANAKT TYRSKKPAKN LKMPGVYYVD YRLERIKEAN
661 NETYVEQHEV AVARYCDLPS KLGHKLN
```

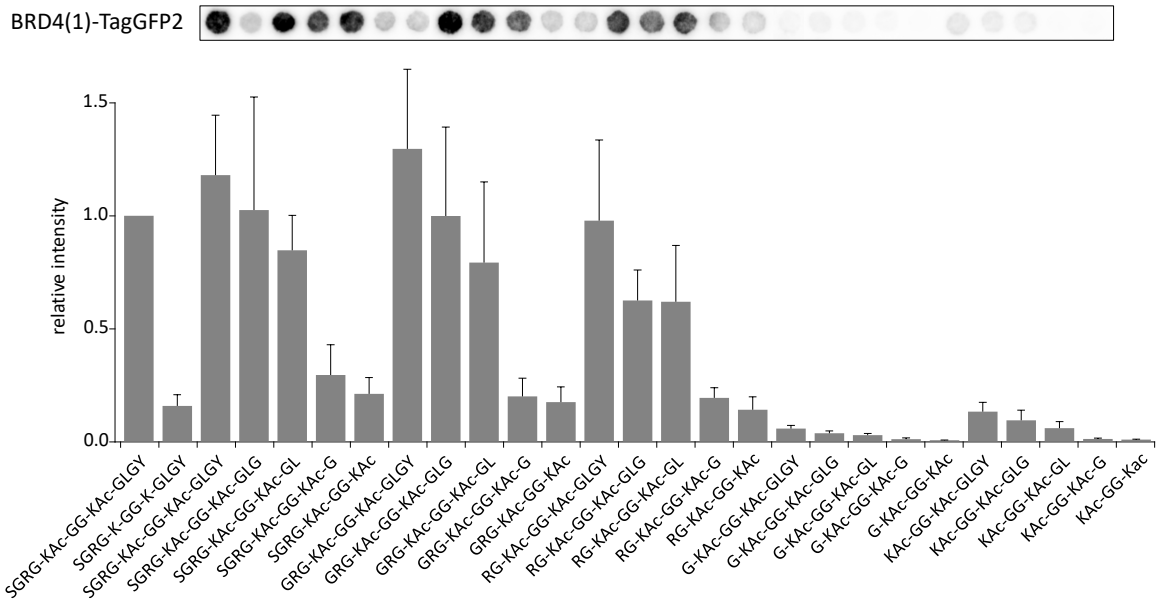


Figure 4.1: Results from pull-down experiments with truncated H4K5/8Ac and BRD4(1) fusion proteins. The image of the membrane after incubation with BRD4(1)-TagGFP2 is shown at the top. The quantified fluorescence intensity is displayed as a bar graph on the bottom. Values are normalized to the intensity of the non-truncated H4K5/8Ac sequence serving as a positive control and standard deviation is indicated (n=4).

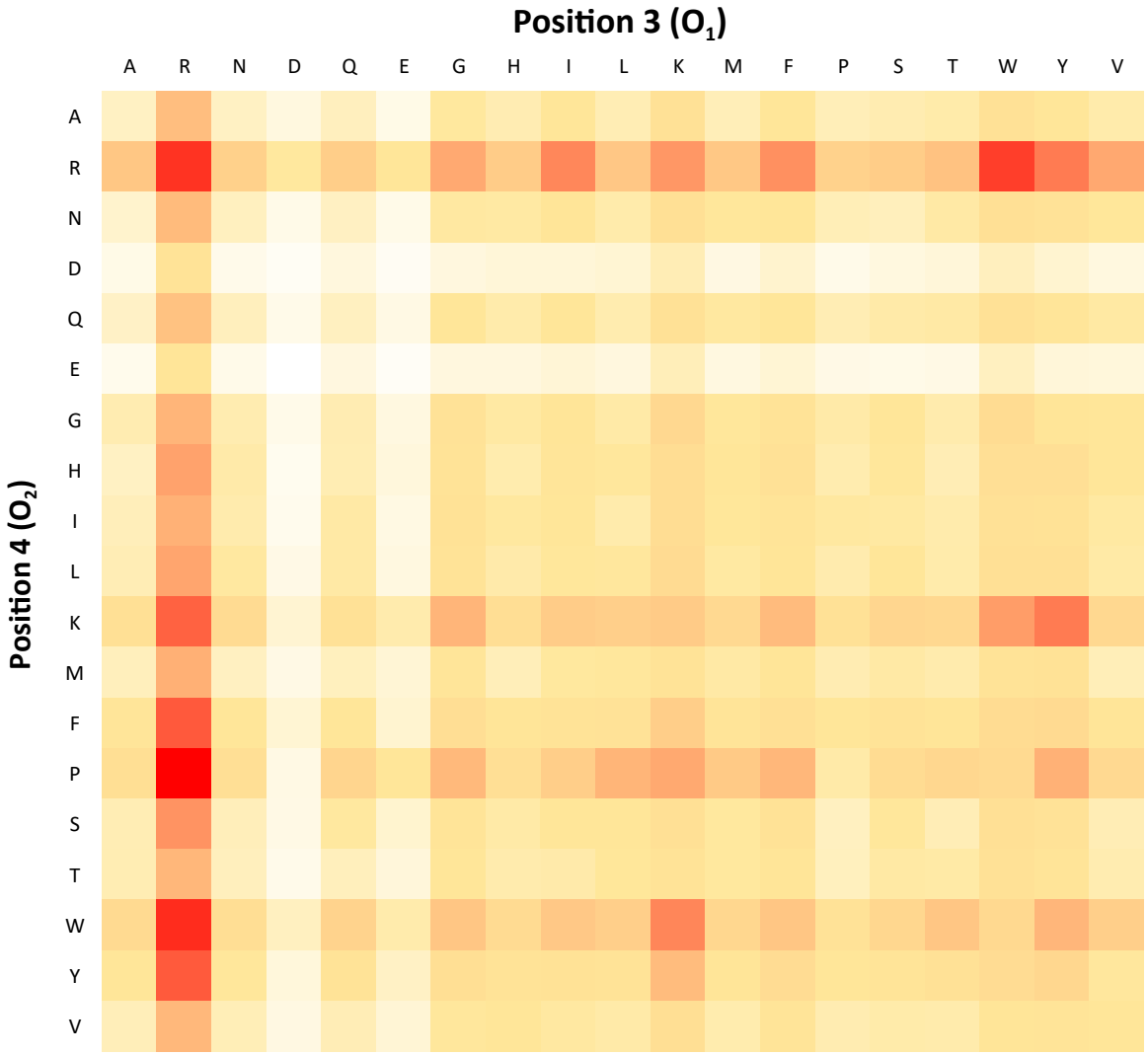


Figure 4.2: Average of fluorescence intensity of three pull-down experiment with BRD3(2)-TurboYFP and the mixture based SPCL for the peptide sequence XXO$_1$-KAc-O$_2$XX-Ahx normalized and displayed as a heat map. The amino acid for position O_1 is set for the indicated column, while the amino acid for position O_2 is set for the indicated row.

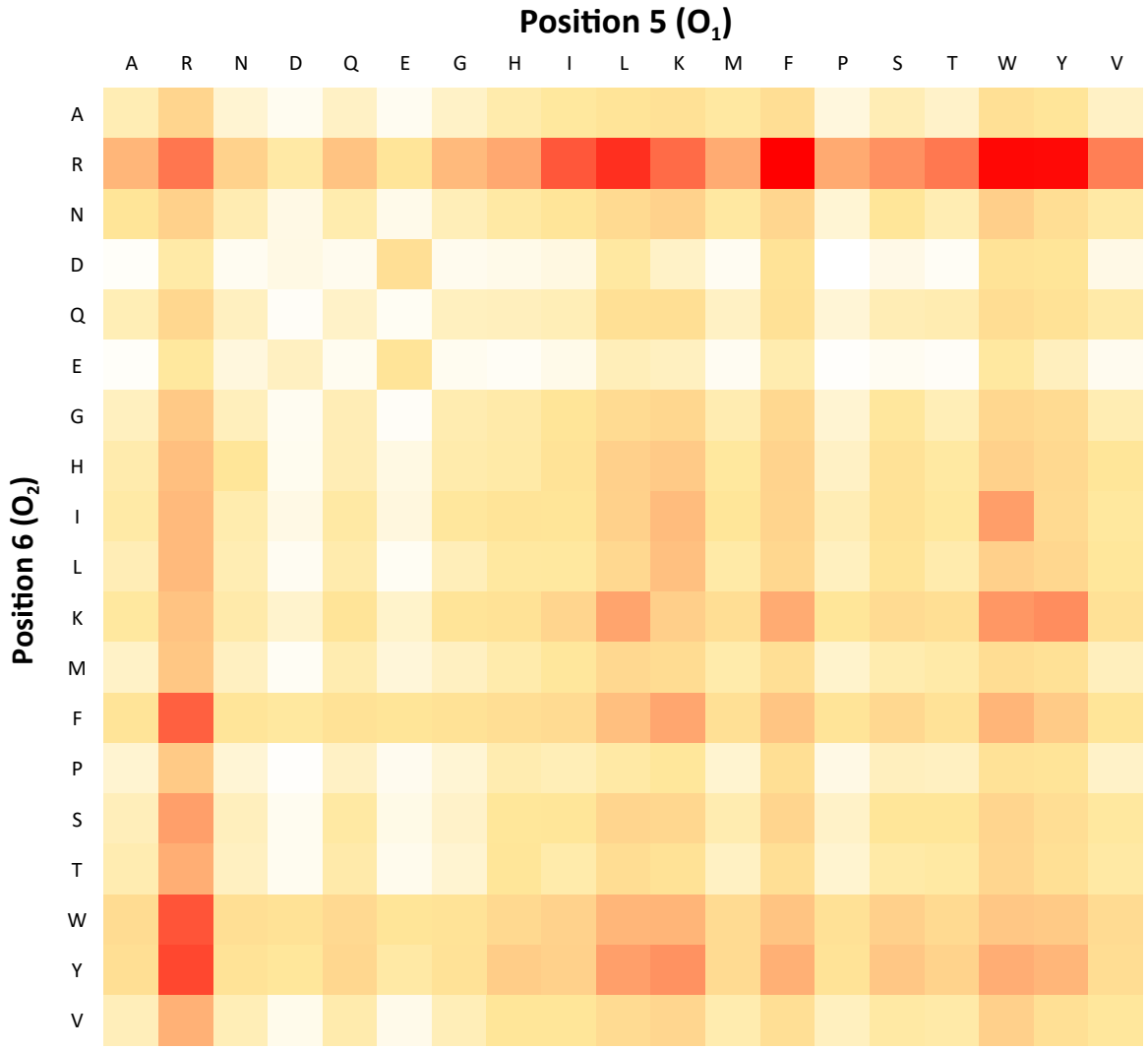


Figure 4.3: Average of fluorescence intensity of three pull-down experiment with BRD3(2)-TurboYFP and the mixture based SPCL for the peptide sequence XXX-KAc-XO_1O_2-Ahx normalized and displayed as a heat map. The amino acid for position O_1 is set for the indicated column, while the amino acid for position O_2 is set for the indicated row.

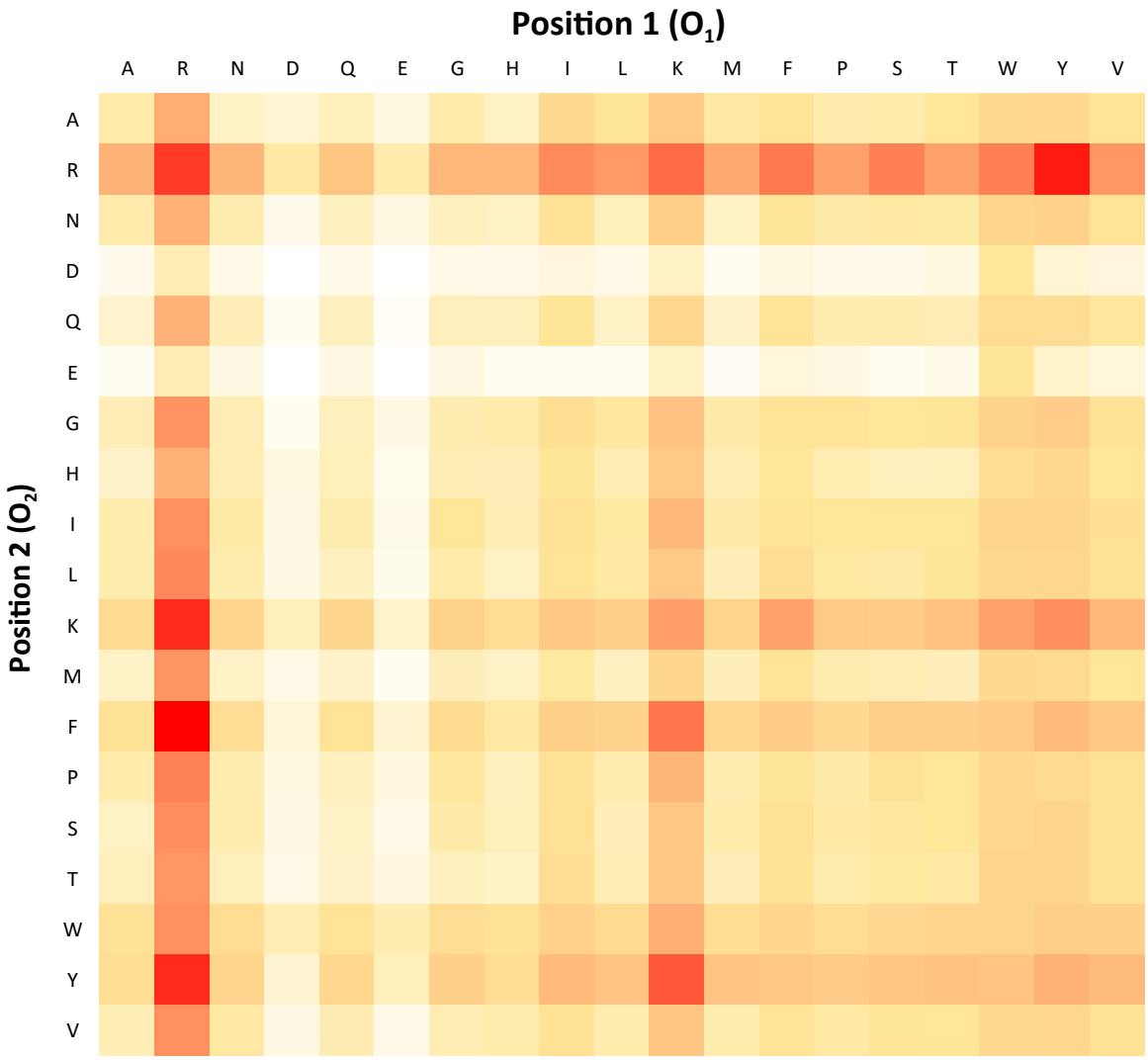


Figure 4.4: Average of fluorescence intensity of three pull-down experiment with BRD4(2)-TurboYFP and the mixture based SPCL for the peptide sequence O_1O_2X-KAc-XXX-Ahx normalized and displayed as a heat map. The amino acid for position O_1 is set for the indicated column, while the amino acid for position O_2 is set for the indicated row.

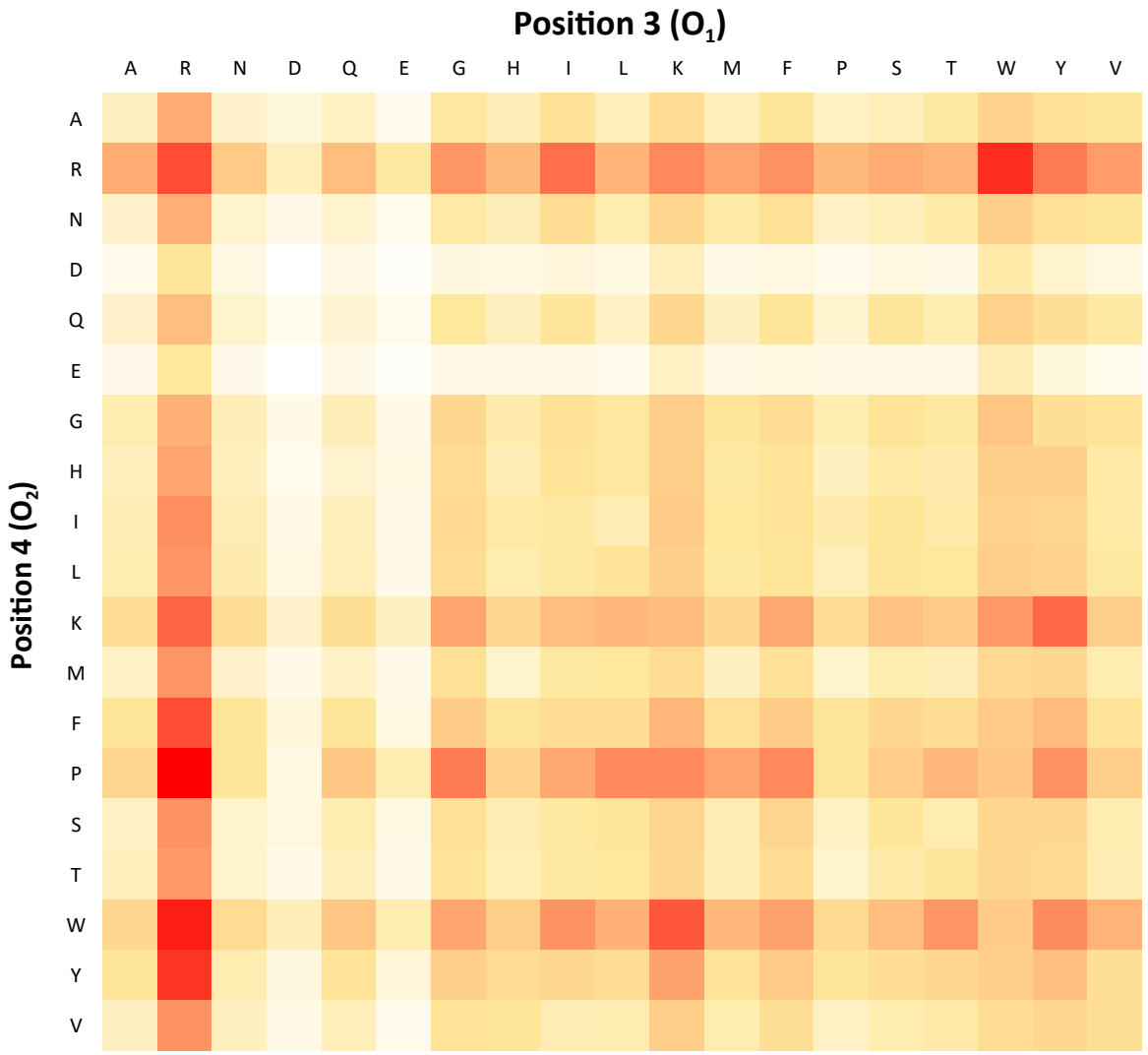


Figure 4.5: Average of fluorescence intensity of three pull-down experiment with BRD4(2)-TurboYFP and the mixture based SPCL for the peptide sequence XXO$_1$-KAc-O$_2$XX-Ahx normalized and displayed as a heat map. The amino acid for position O_1 is set for the indicated column, while the amino acid for position O_2 is set for the indicated row.

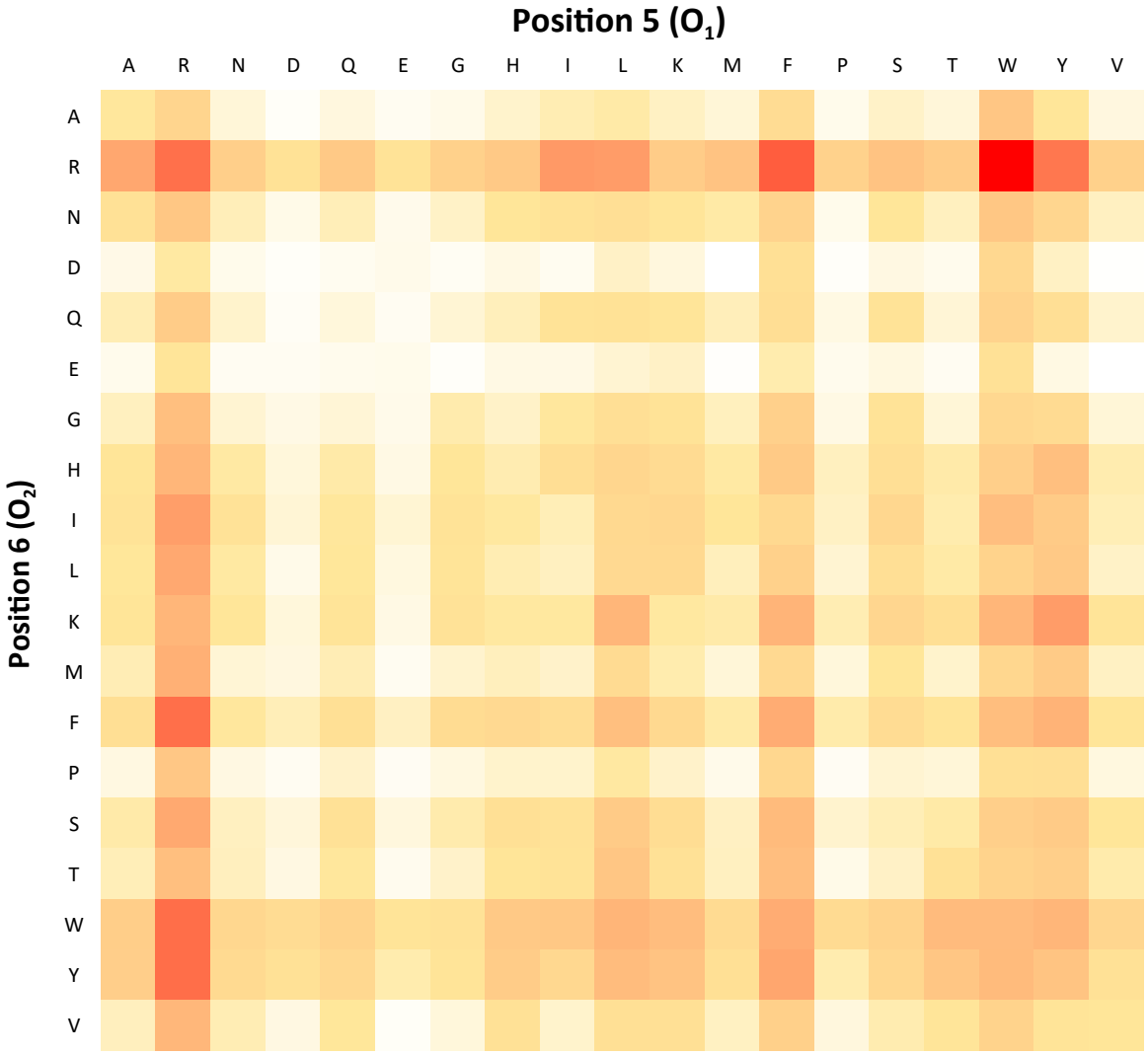


Figure 4.6: Average of fluorescence intensity of three pull-down experiment with BRD4(2)-TurboYFP and the mixture based SPCL for the peptide sequence XXX-KAc-XO$_1$O$_2$-Ahx normalized and displayed as a heat map. The amino acid for position O_1 is set for the indicated column, while the amino acid for position O_2 is set for the indicated row.

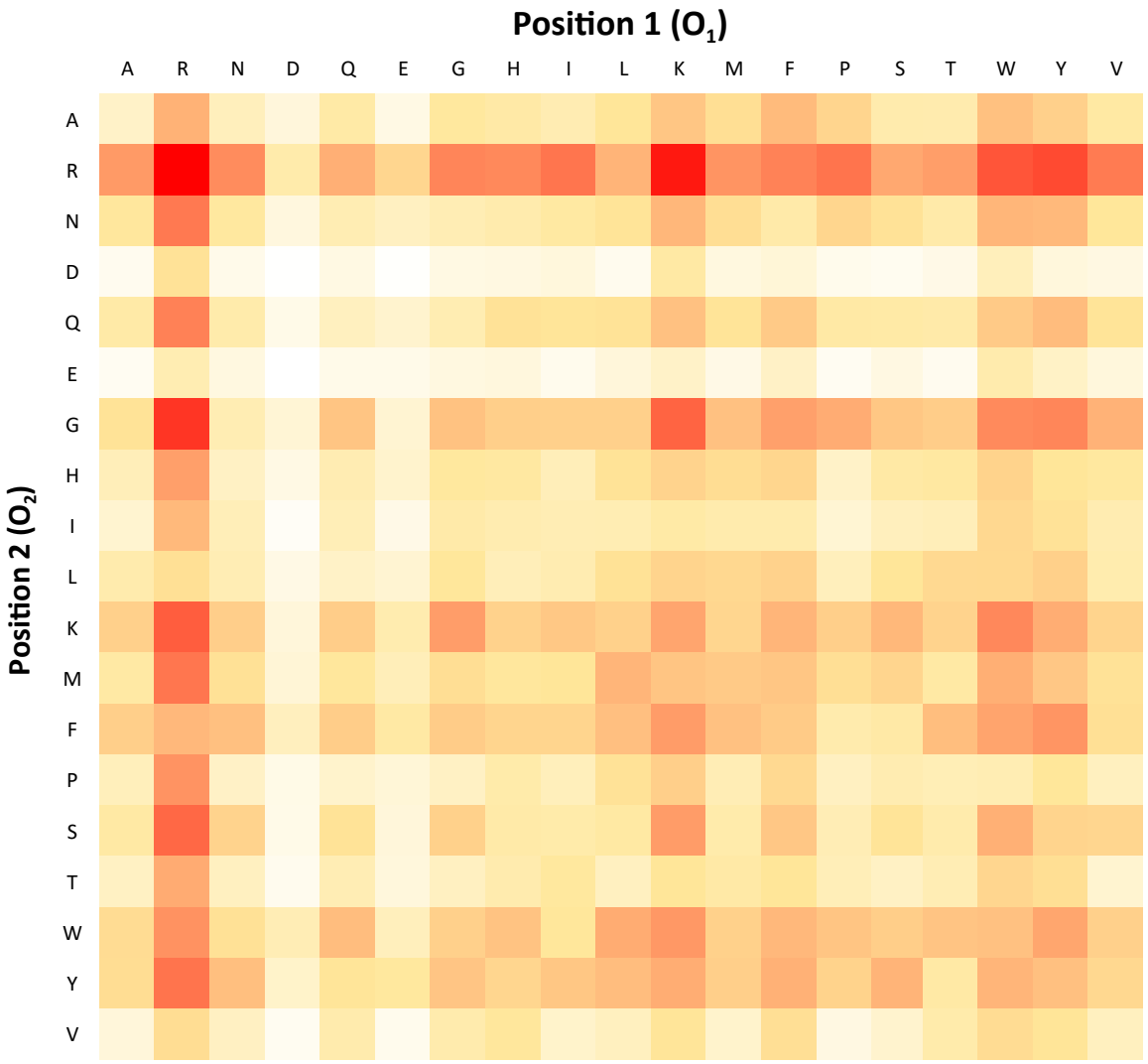


Figure 4.7: Average of fluorescence intensity of three pull-down experiment with BRD4(1)-GFP2 and the mixture based SPCL for the peptide sequence O_1O_2X-KAc-GG-KAc-XXX-Ahx normalized and displayed as a heat map. The amino acid for position O_1 is set for the indicated column, while the amino acid for position O_2 is set for the indicated row.

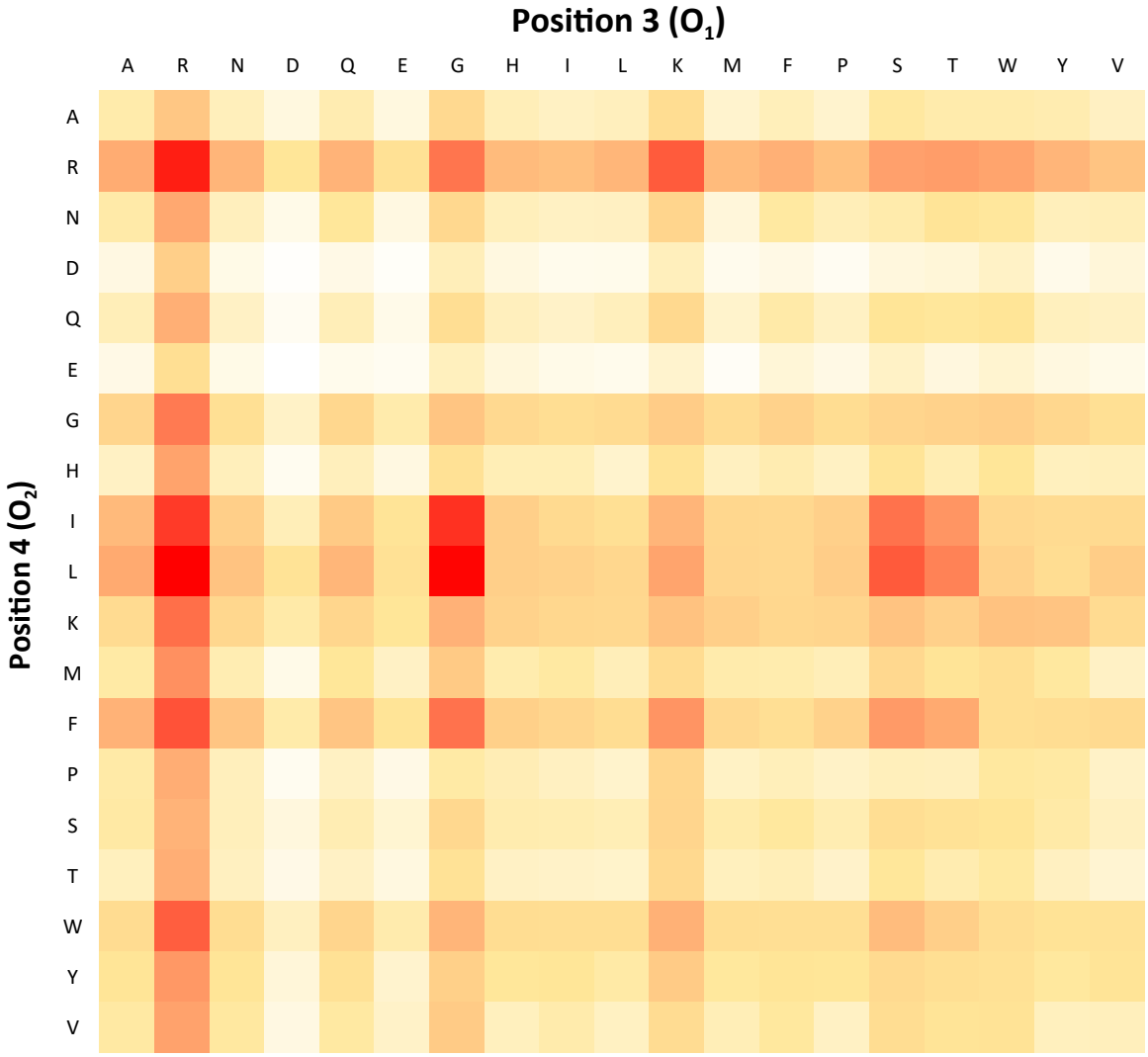


Figure 4.8: Average of fluorescence intensity of three pull-down experiment with BRD4(1)-GFP2 and the mixture based SPCL for the peptide sequence XXO$_1$-KAc-GG-KAc-O$_2$XX-Ahx normalized and displayed as a heat map. The amino acid for position O_1 is set for the indicated column, while the amino acid for position O_2 is set for the indicated row.

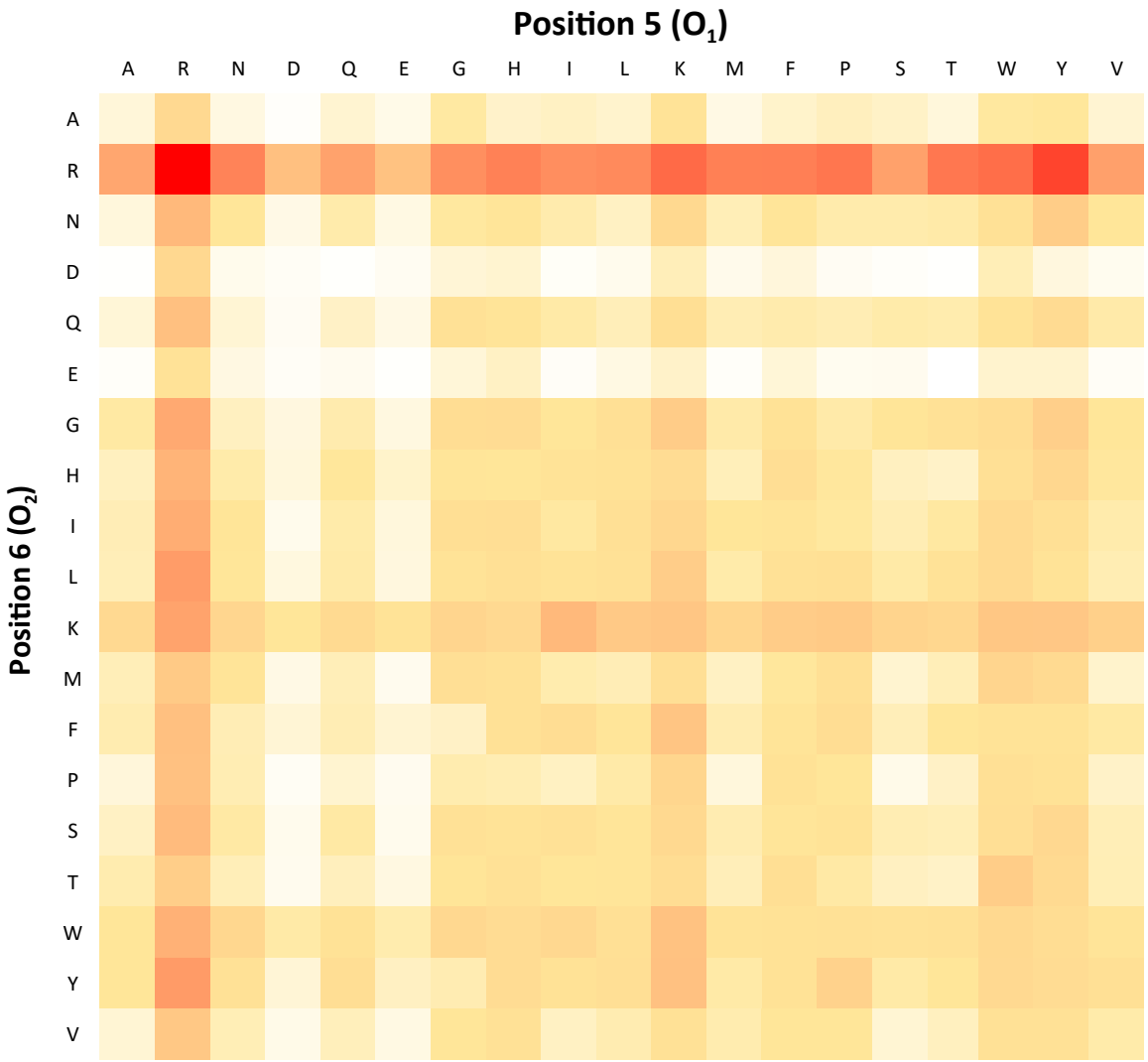


Figure 4.9: Average of fluorescence intensity of three pull-down experiment with BRD4(1)-GFP2 and the mixture based SPCL for the peptide sequence XXX-KAc-GG-KAc-XO_1O_2-Ahx normalized and displayed as a heat map. The amino acid for position O_1 is set for the indicated column, while the amino acid for position O_2 is set for the indicated row.

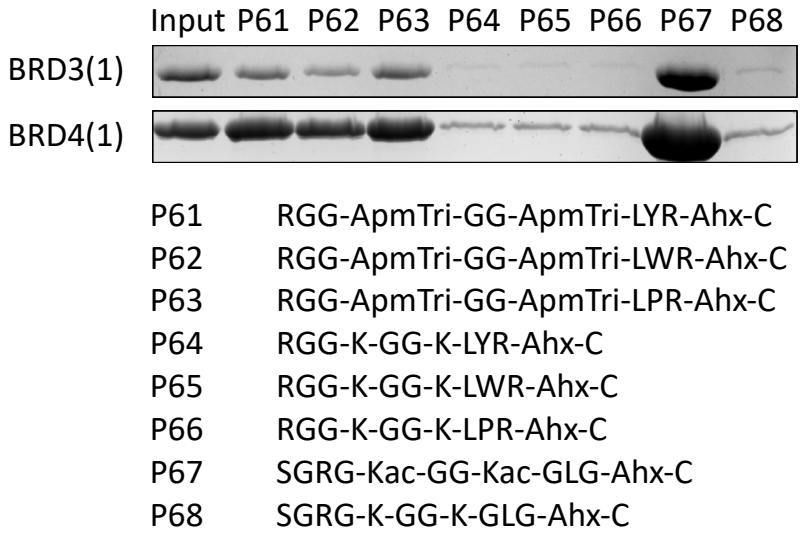


Figure 4.10: SDS-PAGE analysis of pull-down conducted with BRD3(1)-mKate2 and BRD4(1)-TagGFP2 as well as the peptide probes P061-P068. The bands of the SDS-PAGE are visualized on top while the sequences of the peptide are displayed at the bottom.

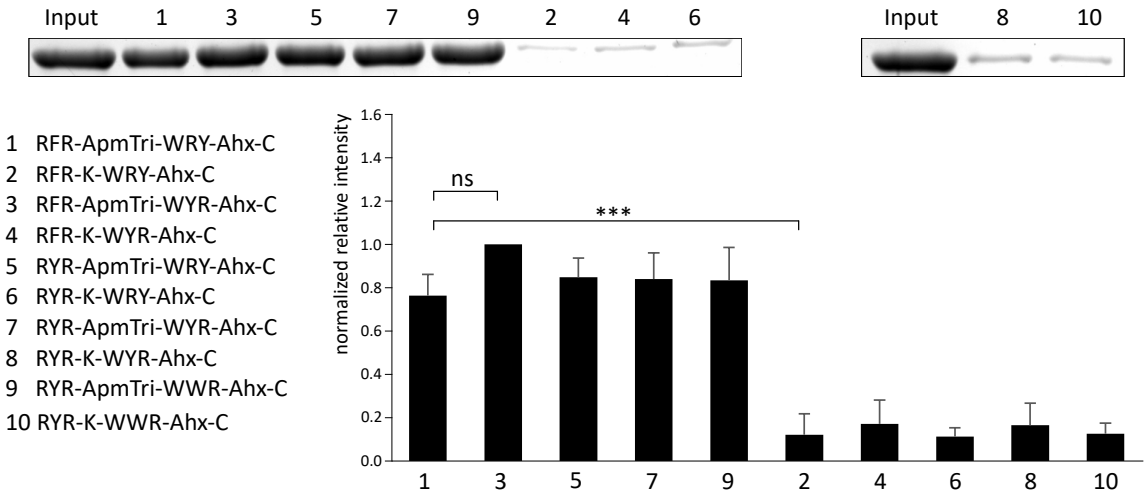


Figure 4.11: Pull-down results for BRD4(2)-TurboYFP with affinity mutants. Coomassie-stained SDS-PAGE bands with BRD4(2)-TurboYFP and the probes on the top with the sequences of each peptide displayed at the left. ApmTri is color-coded in yellow while all negative controls incorporate a red lysine. The quantified intensity was normalized to the intensity of the probe exhibiting the strongest band and is displayed as the mean (n=3). P-values were calculated using ANOVA and Tukey's HSD test for multiple comparisons. *=$p<0.05$; **=$p<0.01$; ***=$p<0.001$; ns=not significant.

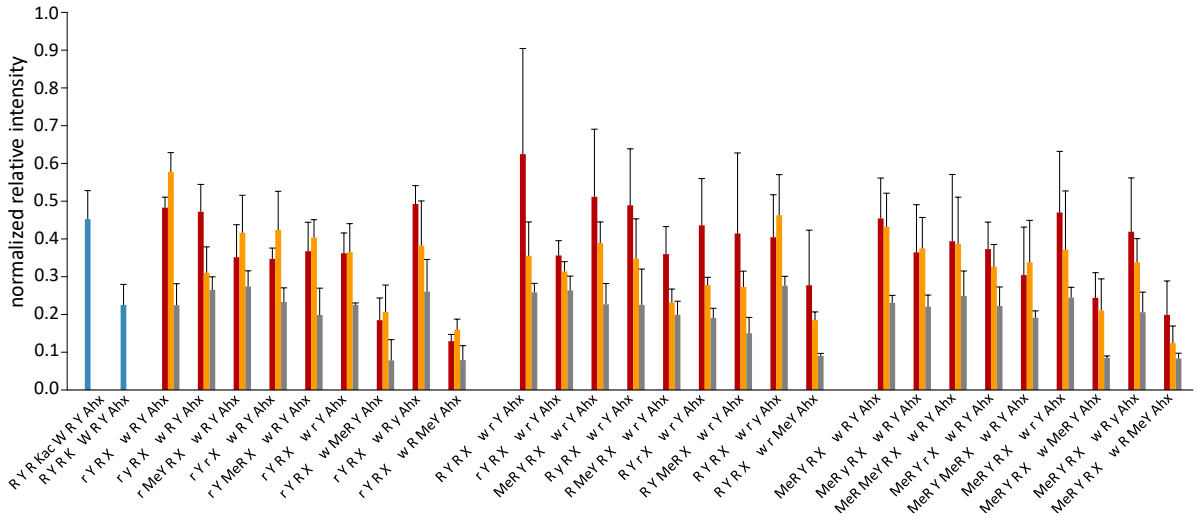


Figure 4.12: Normalized fluorescence intensity of pull-down results for BRD3(2)-TurboYFP with the second round of stability mutants of P5 (RYR-ApmTri-WRY-Ahx). The positive and negative control are displayed as blue bar graphs. Red bars represent the results of the peptides incorporating acetyllysine, orange graphs represent ApmTri peptides and grey graphs represent peptides containing Lys. Standard deviation is indicated.

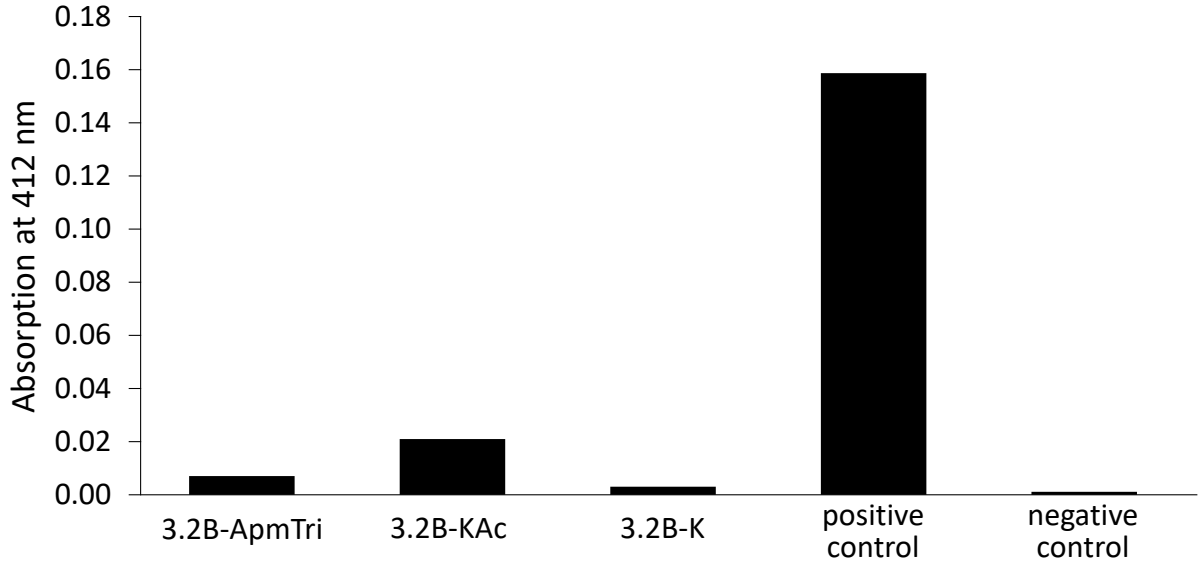


Figure 4.13: Absorbance measured of 10 µM solutions of peptides 3.2B-ApmTri, 3.2B-KAc, 3.2B-K as well as a positive and a negative control incubated for 5 min with a DTNB solution.

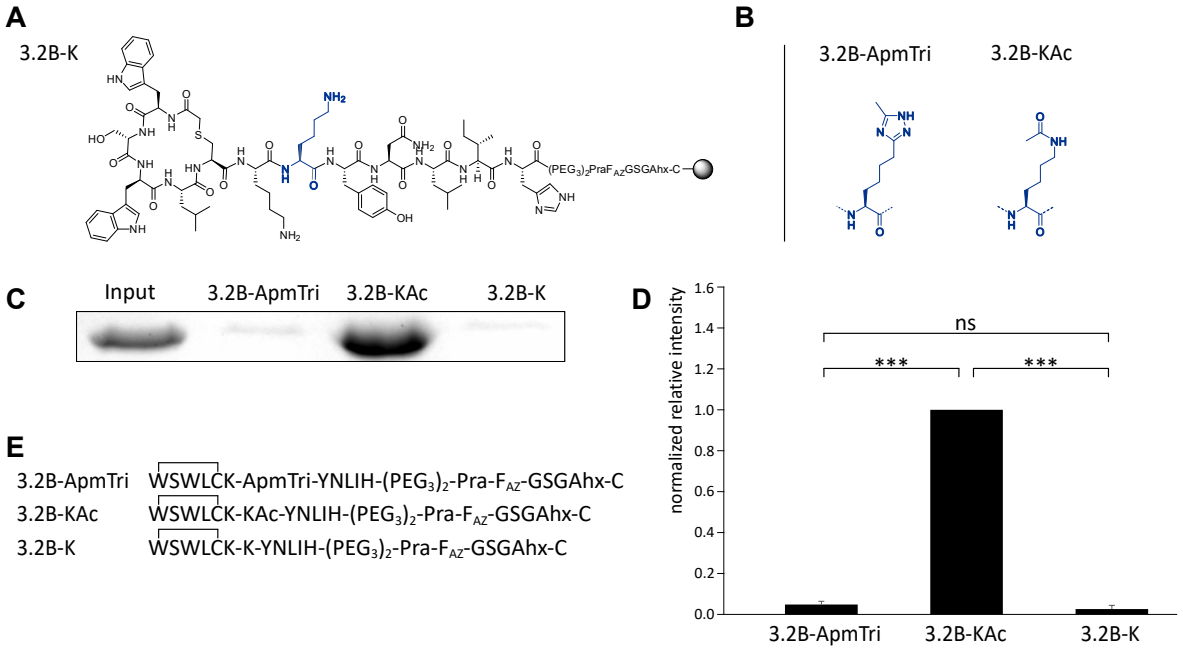


Figure 4.14: Pull-down results for BRD4(1)-TagGFP2 with 3.2B-KAc, 3.2B-ApmTri and 3.2B-K. A) Structure of 3.2B-K. B) ApmTri and KAc replacing lysine in the sequence of 3.2B-ApmTri and 3.2B-KAc, respectively. C) Coomassie-stained SDS-PAGE bands of pull-downs with BRD4(1)-TurboYFP and the probes. D) The quantified intensities of the pull-downs were normalized to the intensity of the probe exhibiting the strongest band and are displayed as the mean (n=3). P-values were calculated using ANOVA and Tukey's HSD test for multiple comparisons. *=$p<0.05$; **=$p<0.01$; ***=$p<0.001$; ns=not significant. E) The sequences of the three probes, cyclisation is indicated.

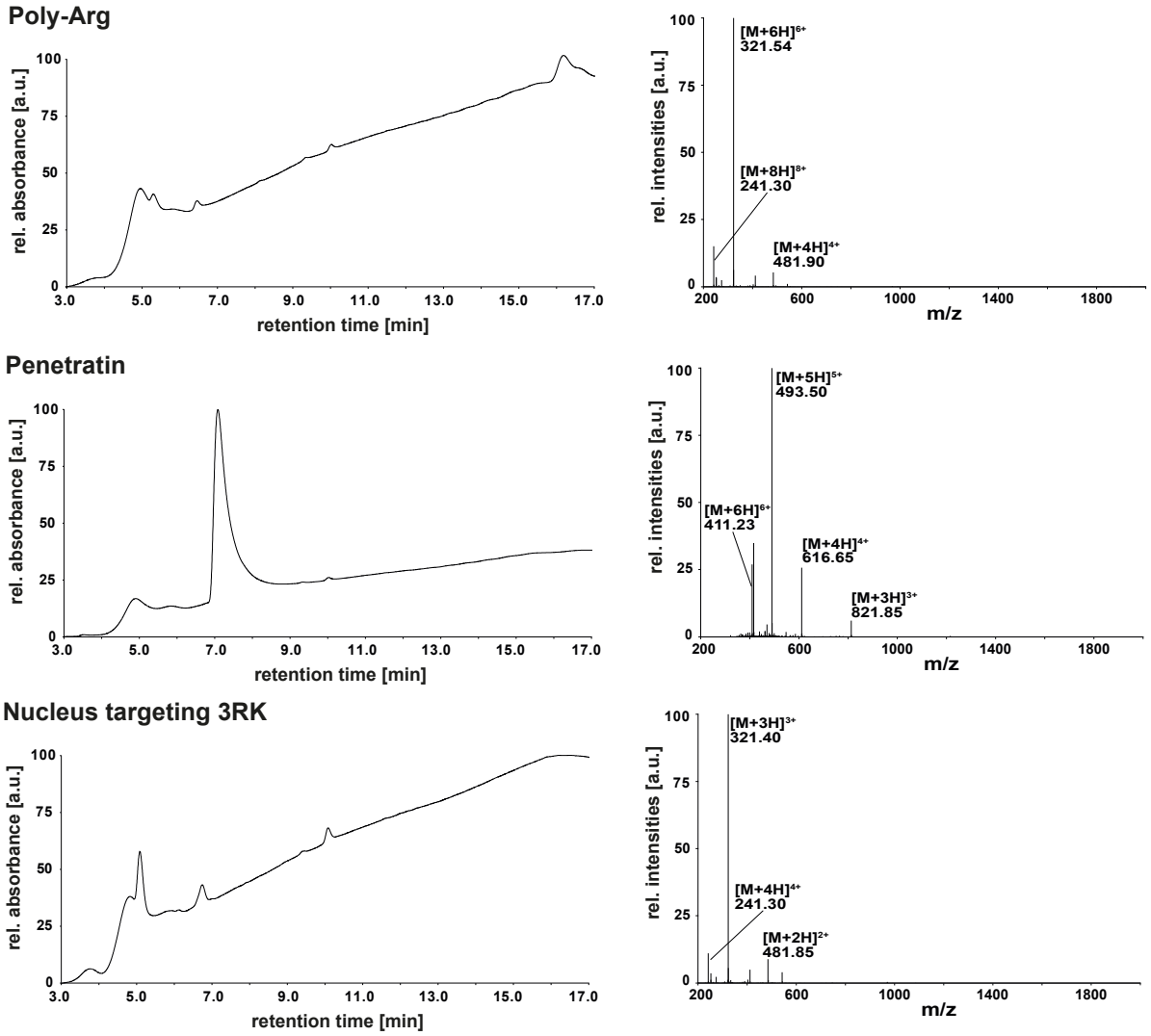


Figure 4.15: LC-MS analyses of linear CPPs poly-arg, penetratin and the nucleus targeting 3RK. Left side: UV-chromatograms, right side: MS spectra.

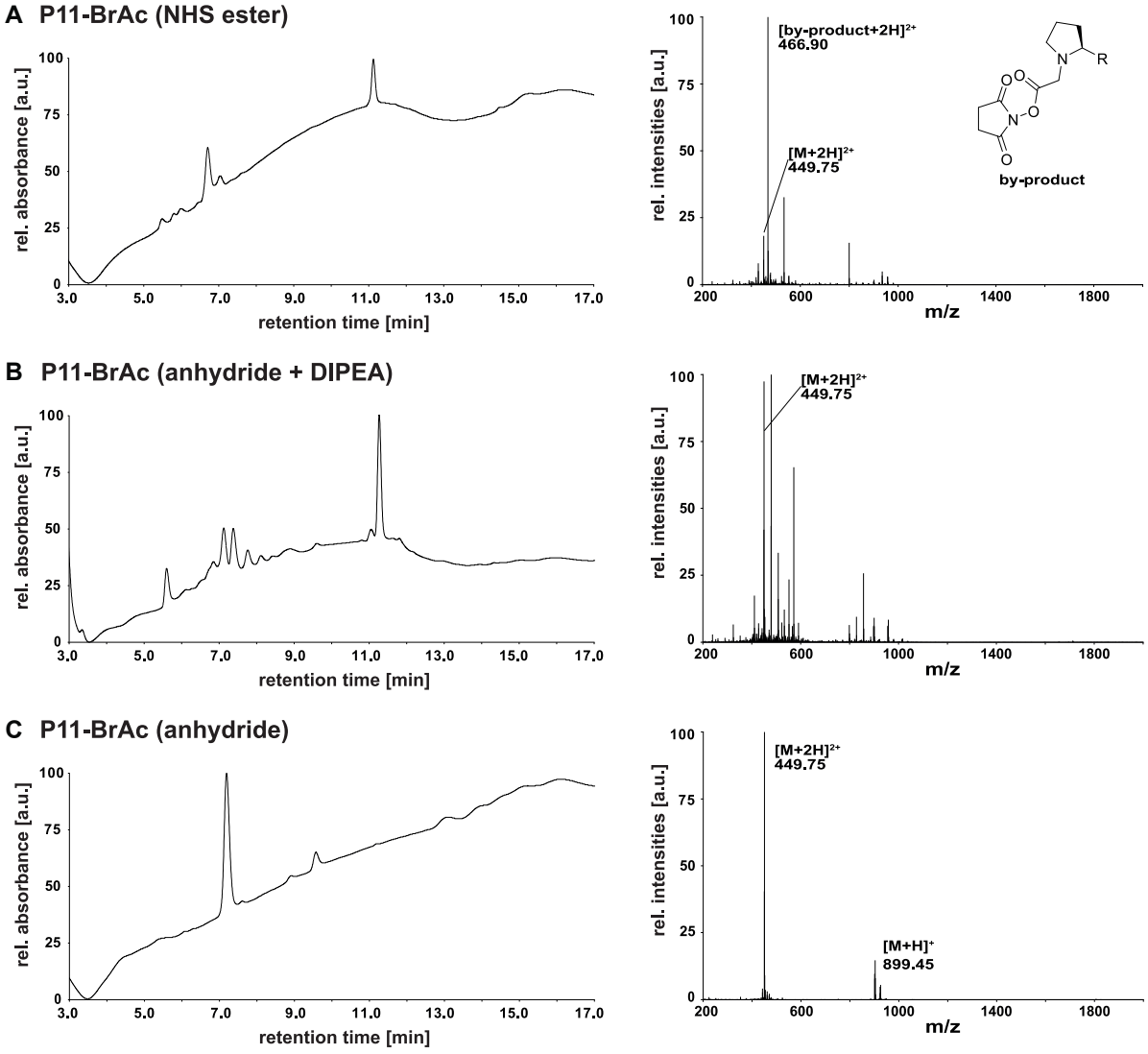


Figure 4.16: LC-MS analyses of the reactions described in 2.20 Left side: UV-chromatograms, right side: MS spectra, R (in the proposed by-product in a)): GKAcGG-AhxAhx-Pra.

Table 4.2: Overview of suppliers and the location of their headquarters (HQ).

Company	Short	HQ-location
3M	3M	Saint Paul, MN, United States
Abcam	Abcam	Cambridge, United Kingdom
Agilent Technologies, Inc.	Agilent	Santa Clara, CA, United States
AppliChem GmbH	AppliChem	Darmstadt, Germany
Avestin, Inc.	Avestin	Ottawa, Canada
Bachem AG	Bachem	Bubendorf, Switzerland
Bio-Rad Laboratories, Inc.	Bio-Rad	Hercules, CA, United States
Biosolve B.V.	Biosolve	Valkenswaard, Netherland
Biozym Scientific GmbH	Biozym	Hessisch Oldendorf, Germany
Bruker Corporation	Bruker	Billerica, MA, United States
Carbolution Chemicals GmbH	Carbolution	St. Ingbert, Germany
Cytiva Life Sciences	Cytiva	Marlborough, MA, United States
Dr. Maisch HPLC GmbH	Dr. Maisch	Ammerbuch, Germany
Eppendorf SE	Eppendorf	Hamburg, Germany
GL Biochem (Shanghai) Ltd.	GL Biochem	Shanghai, China
Infors AG	Infors	Bottmingen, Switzerland
Intavis GmbH & Co. KG	Intavis	Tübingen, Germany
Iris Biotech GmbH	Iris	Marktredwitz, Germany
LI-COR Biosciences	Li-cor	Lincoln, NE, United States
Lab Logistics Group GmbH	LLG Labware	Meckenheim, Germany
Lonza Group AG	Lonza	Basel, Switzerland
Martin Christ GmbH	Christ	Osterode am Harz, Germany
Merck KGaA	Merck	Darmstadt, Germany
MultiSynTech GmbH	MultiSynTech	Witten, Germany
Phenomenex Inc.	Phenomenex	Torrance, CA, United States
Proteintech Group, Inc.	Proteintech	Rosemont, IL, United States
Rapp Polymere GmbH	Rapp Polymere	Tübingen, Germany
Roche Holding AG	Roche	Basel, Switzerland
Carl Roth GmbH + Co. KG	Carl Roth	Karlsruhe, Germany
Sarstedt AG & Co. KG	Sarstedt	Nümbrecht, Germany
Shimadzu Corporation	Shimadzu	Kyoto, Japan
Sigma Aldrich	Sigma Aldrich	St. Louis, MO, United States
Th. Geyer GmbH & Co. KG	Th. Geyer	Renningen, Germany
Thermo Fisher Scientific Inc.	Thermo Fisher	Waltham, MA, United States
Tokyo Chemical Industry	TCI	Tokio, Japan
VWR International	VWR	Radnor, PA, United States
FUJIFILM Wako Europe GmbH	Wako	Neuss, Germany