MASARYKOVA
UNIVERZITA

PŘÍRODOVĚDECKÁ FAKULTA

Chemická derivatizace histonových proteoforem pro hmotnostně spektrometrickou analýzu

Diplomová práce

PALINA RYZHAYA

Vedoucí práce: Mgr. Gabriela Lochmanová, Ph.D

Ústav biochemie
Program Biomolekulární chemie

Brno 2024

MUNI
SCI

Bibliografický záznam

Autor: Palina Ryzhaya
Přírodovědecká fakulta
Masarykova univerzita
Ústav biochemie

Název práce: Chemická derivatizace histonových proteoforem pro hmotnostně spektrometrickou analýzu

Studijní program: Biomolekulární chemie

Vedoucí práce: Mgr. Gabriela Lochmanová, Ph.D

Rok: 2024

Počet stran: 70

Klíčová slova: histony, posttranslační modifikace, chemická derivatizace, hmotnostní spektrometrie

Bibliographic record

Author: Palina Ryzhaya
Faculty of Science
Masaryk University
Department of Biochemistry

Title of Thesis: Chemical derivatization of histone proteoforms for mass spectrometric analysis

Degree Programme: Biomolecular Chemistry

Supervisor: Mgr. Gabriela Lochmanová, Ph.D

Year: 2024

Number of Pages: 70

Keywords: histones, post-translational modifications, chemical derivatization, mass spectrometry

Abstrakt

Posttranslační modifikace histonů jsou epigenetické značky spojené s regulací klíčových buněčných procesů, jako je transkripce, replikace a oprava DNA. Provádění kvalitativní a kvantitativní analýzy histonových proteoforem je zásadní pro pochopení mechanismů normálních a patologických procesů, protože jejich regulace zahrnuje změny ve struktuře a dynamice chromatinu. Hmotnostní spektrometrie je v současnosti nejrozšířenějším přístupem pro analýzu histonů. Navzdory vysoké přesnosti a citlivosti se histonová analýza pomocí LC-MS/MS stále potýká s problémy, z nichž mnohé souvisí s častou přítomností zbytků lysinu a argininu v aminokyselinových sekvencích histonů. Ačkoli chemická derivatizace amino skupin byla zavedena jako užitečná strategie pro přípravu histonových vzorků pro proteomiku z dola, nevýhody spojené se značením aktuálně používanými anhydridy omezují výsledky histonových analýz založených na hmotnostní spektrometrii.

V této diplomové práci bylo testováno několik anhydridů jako alternativních derivatizačních činidel k překonání nevýhod běžně používaného anhydridu kyseliny propionové. MALDI-TOF hmotnostní spektrometrie byla použita k výběru potenciálních kandidátů pro následné upřesnění podmínek značení, včetně koncentrace anhydridu a parametrů inkubace. Po určení anhydridu kyseliny trifluoropropionové jako nejslibnějšího alternativního derivatizačního činidla byla vyhodnocena účinnost a porovnána s deri-vatizací anhydridu kyseliny propionové. Získaný protokol prozatím neumožňuje jeho použití jako alternativu ke stávajícím metodám z důvodu výrazně nižší účinnosti značení. Pokud se v budoucnu podaří protokol vylepšit, může se stát anhydrid kyseliny trifluoropropionové vzhledem k velkému zvýšení hydrofobicity peptidů a absenci nespecifických reakcí slibnou alternativou pro derivatizaci histonů.

Abstract

Histone post-translational modifications are epigenetic marks associated with the regulation of key cellular processes such as transcription, replication, and DNA repair. Conducting qualitative and quantitative analysis of histone proteoforms is crucial for understanding the mechanisms of normal and pathological processes, as their regulation involves changes in chromatin structure and dynamics. Mass spectrometry is currently the most widely used approach for histone analysis. Despite its high accuracy and performance, histone analysis by LC-MS/MS still faces challenges, many of which are related to the frequent presence of lysine and arginine residues in histone amino acid sequences. Although chemical derivatization of amine groups has been introduced as a useful strategy to prepare histone samples for bottom-up proteomics, drawbacks associated with labelling with currently used anhydrides limit the outcome of mass spectrometry-based histone analyses.

In this thesis, several anhydrides were tested as alternative derivatization agents to overcome the drawbacks of commonly used propionic anhydride. MALDI-TOF mass spectrometry was used to select potential candidates for subsequent refinement of the labelling conditions, including anhydride concentration and incubation parameters. After determining trifluoropropionic anhydride as the most promising alternative derivatization agent, the performance was evaluated and compared with propionic anhydride derivatization. The protocol obtained does not yet allow its use as an alternative to existing methods because of significantly lower labelling efficiency. When improved, trifluoropropionic anhydride can be a promising candidate for histone derivatization because of the large increase in peptides' hydrophobicity and the absence of non-specific reactions.

MUNI
SCI

MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
KOTLÁŘSKÁ 2, 611 37 BRNO
IČ: 00216224
DIČ: CZ00216224

ZADÁNÍ
DIPLOMOVÉ PRÁCE

Akademický rok: 2023/2024

Ústav:	Ústav biochemie
Studentka:	Palina Ryzhaya
Program:	Biochemie
Specializace:	Biomolekulární chemie

Ředitel *ústavu* PřF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje diplomovou práci s názvem:

Název práce:	Chemická derivatizace histonových proteoforem pro hmotnostně spektrometrickou analýzu
Název práce anglicky:	Chemical derivatization of histone proteoforms for mass spectrometric analysis
Jazyk závěrečné práce:	angličtina

Oficiální zadání:
Histone post-translational modifications are epigenetic marks that strongly affect numerous processes, including cell cycle and protein interactions. Currently, mass spectrometry analysis represents a preferred approach for histone characterization. The diversity of histones and the complexity of their modifications arising from high contents of reactive amine groups in their amino acid sequences make histone characterization challenging. One of the key steps in sample preparation prior to LC-MS/MS analysis is the chemical derivatization of histones. Even though commonly used derivatization agents improve the chromatographic properties and detection of post-translationally modified histone peptides, they still possess certain limits. Within the frame of the master thesis, the student will test alternative derivatization reagents for histone labeling to improve chromatographic separation of histone peptide forms, facilitate identification and localization of natural acylations, and ensure effective ionization of derivatized peptides. The topic covers chemical derivatization of amine groups, LC-MS/MS analysis of histone peptides, and data processing. The experimental part of the study will be done in the laboratories of Zbyněk Zdráhal Research Group, CEITEC-MU.

Literatura:
KUCHAŘÍKOVÁ, Hana, Pavlína DOBROVOLNÁ, Gabriela LOCHMANOVÁ a Zbyněk ZDRÁHAL. Trimethylacetic Anhydride-Based Derivatization Facilitates Quantification of Histone Marks at the MS1 Level. *Molecular and Cellurar Proteomic*. American Society for Biochemistry and Molecular Biology, 2021, roč. 20, JAN 01, s. ”100114”, 14 s. ISSN 1535-9476. doi:10.1016/j.mcpro.2021.100114.

Vedoucí práce:	Mgr. Gabriela Lochmanová, Ph.D.
Konzultant:	Mgr. Pavlína Pírek
Datum zadání práce:	13. 5. 2022
V Brně dne:	3. 1. 2024

Zadání bylo schváleno prostřednictvím IS MU.

Palina Ryzhaya, 2. 1. 2024
Mgr. Gabriela Lochmanová, Ph.D., 2. 1. 2024
doc. Mgr. Jan Havliš, Dr., 2. 1. 2024

Declaration

I declare that I have prepared my thesis independently under the mentorship of the thesis supervisor and using the information sources provided in the thesis.

Brno December 28, 2023 ……..

Palina Ryzhaya

Acknowledgements

I would like to express my gratitude to my supervisor, Mgr. Gabriela Lochmanová, Ph.D., and my consultant, Mgr. Pavlína Pírek, for their assistance, patience, and valuable advice and encouragement during my experiments. I am also grateful to Prof. RNDr. Zbyněk Zdráhal, Dr., for providing me with the opportunity to work in their laboratory and be a part of an exceptional and skilled research team.

Additionally, I would like to thank Ing. Renata Holánková for providing me with valuable laboratory techniques, guiding me through numerous experiments, and offering insightful advice for my projects.

I would also like to extend special appreciation to Mgr. Ondrej Šedo, Ph.D., for his guidance in MALDI-TOF analysis.

Finally, I am grateful to my family and friends for their unwavering support and encouragement.

Table of Contents

Glossary **15**

1 Introduction **17**

2 Theoretical part **18**

2.1 The role of epigenetics in physiological processes 18

2.2 Nucleosome subunit organization 20

2.3 Histone PTMs 23

2.4 MS-based approaches of histone analysis 26

2.5 Histone chemical derivatization 32

3 The goal of the work **38**

4 Materials and Methods **39**

4.1 Materials 39

4.2 Methods 39

5 Results **45**

5.1 Primary selection of derivatization agents using a synthetic peptide 45

5.2 Comparison of the best candidates' performances 49

5.3 Optimization of TFPA derivatization protocol 50

6 Discussion **58**

7 Conclusion **62**

Bibliography **63**

Glossary

ACN	–	Acetonitrile
B-CLL	–	B chronic lymphocytic leukemia
CENPA	–	Centromere protein A
DDA	–	Data-Dependened Acquisition
DIA	–	Data Independent Acquisition
ESI	–	Electrospray ionization
HCCA	–	α-Cyano-4-hydroxytrans-cinnamic acid
KAT	–	Lysine acetyltransferases
KDAC	–	Lysine deacetylases
KMT	–	Lysine methyltransferases
LSD	–	Lysine-specific demethylase
MALDI	–	Matrix-assisted laser desorption/ionization
PIC	–	Phenyl isocyanate
PRMT	–	Protein arginine methyltransferases
PTM	–	Post-translational modification
Prop	–	Propionic anhydride
RP-HPLC	–	Reverse-phase high-performance liquid chromatographic
TFPA	–	3,3,3-Trifluoropropionic anhydride
TOF	–	Time-of-Flight detector
TMA	–	Trimethylacetic anhydride

1 Introduction

Qualitative and quantitative analysis of post-translational modifications of histones is an important component for understanding the role of epigenetic regulation in cellular processes. Since histones are associated with nucleosome, a structural unit of DNA packaging in eukaryotic cells, studying modification patterns is crucial for understanding many pathological processes and disorders. This understanding can further be used in the search for biomarkers and the development of drugs that target these mechanisms.

The most applicable among all modern methods for histone modifications analysis is mass spectrometry using the bottom-up approach, which involves sample preparation using trypsin and further analysis of short peptides. The application of this method is also associated with challenges during LC-MS/MS analysis and subsequent data processing, which can be addressed by introducing an additional step in the sample preparation process. This involves the chemical modification of N-terminal and free amine groups of lysines using organic anhydrides.

Among the numerous tested anhydrides, propionic anhydride is the most commonly used at the moment. However, it has a set of negative qualities such as insufficient increase in the hydrophobicity of short hydrophilic peptides, side reactions at hydroxyl residues, *etc*. The significance of epigenetics has increased in recent years, leading to a demand for a more precise and comprehensive understanding of epigenetic modifications. This requires the development of advanced analytical methods that can meet the current and future demands of epigenetic studies.

2 Theoretical part

2.1 The role of epigenetics in physiological processes

The first suggestion that histone modifications can have a functional effect on the regulation of transcription was made by Vincent Allfrey in 1964. Currently, epigenetics includes events associated with changes in chromatin, which regulates DNA template processes without changing the DNA sequence. Such regulation includes specific DNA and histone modifications. All of them affect such key processes as transcription, replication and DNA repair (Table I) [1,2].

Table I: The impact of certain modifications on key cellular processes

Histone modification	**Chromatin-Reader Motif**
Acetylation (Kac)	Bromodomain Tandem, PHD fingers
Methylation (Kme1, Kme2, Kme3)	Chromodomain, Tudor domain, MBT domain, PWWP domain, PHD fingers, WD40/β propeller
Methylation (Rme1, Rme2s, Rme2a)	Tudor domain
Phosphorylation (Sph, Tph)	14-3-3, BRCT
Phosphorylation (Yph)	SH2[a]
Ubiquitylation (Kub)	UIM, IUIM
SUMOylation (Ksu)	SIM[a]

Me1 – monomethylation; Me2 – dimethylation; me3 – trimethylation; me2s – symmetrical dimethylation; me2a – asymmetrical dimethylation; PHD – plant homeodomain; MBT – malignant brain tumour domain; PWWP – proline-tryptophan-proline domain; BRCT – Breast Cancer Susceptibility Gene 1 C terminus domain; UIM, – ubiquitin interaction motif; IUIM – inverted ubiquitin interaction motif; SIM – sumo interaction motif.

Epigenetic changes are a common hallmark of cancer, characterized by DNA hypomethylation and/or loss or increase in acetylation or methylation of histone proteins [3]. In light of this, the dynamic and reversible nature of epigenetic modulation is an attractive feature for new methods of clinical treatment, which contributes to a more detailed study of these processes.

For instance, global histone H3 methylation and H2A acetylation, among other factors, are observed in all types of lung cancer which is responsible for almost 30% of cancer-related deaths worldwide. H3 histone mutation, involving the replacement of lysine by methionine at position 27, is associated with approximately 80 % of paediatric high-grade glioma cases, which tragically holds a survival rate of less than 1 %. Histone acetylation and deacetylation are pivotal factors contributing to abnormal

chromatin condensation and remodelling during malignant tumour development in the pathogenesis of acute myeloid leukaemia. These epigenetic modifications are essential in understanding and addressing the complexities of cancer progression[4].

Despite the not fully understood ethology of autoimmune diseases, they are currently associated with violations of DNA methylation profiles and histone modifications. Patients with active systemic lupus erythematosus show a decrease in global H3 and H4 acetylation in CD4+ T cells, with the level of acetylation negatively associated with disease activity [5]. Rheumatoid arthritis is also an example of an autoimmune disease associated with epigenetic aberration that affects multiple inflammatory and matrix-related pathways and contributes to the pathogenesis [6].

Methylation of certain DNA regions, along with the effect of H3 methylation and H4 histone acetylation, as well as miRNA dysregulation, affect the diversity of the phenotypes and are associated with the development of asthma [7]. Other respiratory diseases, the pathogenesis of which is associated with impaired phosphorylation of histones H3 and H4, also include chronic obstructive pulmonary disease [8].

In recent years paternal epigenetics has gained prominence in understanding male infertility and its reproductive impacts. During sperm maturation, a majority of histones (around 85–95 %) are removed to compact DNA within the sperm nucleus, a process termed spermiogenesis. Proper replacement of histones by protamines is crucial for normal spermiogenesis and disruptions in this process are associated with male infertility. Identifying distinct post-translational modifications (PTMs) signatures linked to clinical violations holds significant translational potential as it can shed light on how abnormal chromatin may be carried into the oocyte by sperm with phenotypic abnormalities [9].

A large number of published works account for the search for a connection between epigenetics and neurodegenerative diseases. Alzheimer's disease development is associated with many processes, the role of which is still not completely clear. Various *in vitro* experiments have shown a relationship between the development of the disease and the effect on histone modifications pattern: a general decrease of acetylation in both mice and human brain cells [10], a significant difference in the level of H3K9 methylation [11], and others. Since there is currently no effective treatment for Alzheimer's disease, the accumulated knowledge can be used to develop one. Clinical research has already explored the testing of KDAC repressors [12].

The accumulation of knowledge about the influence of epigenetics on the development of pathologies is important for the creation of effective methods of diagnosis and treatment, including the selection of highly specific biomarkers and the development of effective drugs. Significant advantages of biomarkers are that they do not depend solely on the DNA sequence, they can be measured in various body fluids, tissues, and most of them can be diagnosed at very early stages of disease development [13]. Also, by the year 2015, three KDAC inhibitors had been approved by the US Food and Drug Administration (FDA) for the treatment of various types of T-cell lymphoma. The first

drug, Vorinostat (SAHA, Zolina), developed by Merck & Co. Inc., was approved in October 2006. The second drug, romidepsin (Istodax, FK228, FR901228, depsipeptide), developed by Gloucester Pharmaceuticals, was approved at the end of 2009, and the third drug, belinostat (Beleodaq, PXD-101), developed by Spectrum Pharmaceuticals, was approved on July 3, 2014 [14].

2.2 Nucleosome subunit organization

Nucleosome subunits organization in chromatin presents a multitude of functions. Its compact genomic DNA contributes to stable gene repression and restricts the binding of trans-acting factors to specific DNA sequences. However, nucleosomes can be readily disassembled and reassembled to facilitate rapid access to DNA during processes like transcription, replication, and DNA repair. Additionally, nucleosomes play a protective role against DNA damage and serve as a platform for the deposition of various epigenetic signals. They also form the backbone for the assembly of higher-order chromatin structures [15].

2.2.1 Nucleosome structure

Each nucleosome consists of a nucleosome core, linker DNA, and often a linker histone. The structure of the nucleosome core remains relatively constant from yeast to metazoans [16]. It comprises a 147 bp segment of DNA and two copies of four core histone proteins: H2A, H2B, H3, and H4 (Figure 1A). The core histones form a spool-like structure around which the core DNA is wrapped, completing approximately 1¾ left-handed superhelices turn. This results in a compact disc-like structure measuring about 5.5 nm in height and 11 nm in diameter [17].

The four core histones, which are highly conserved among eukaryotic species, are relatively small (11-15 kDa) and positively charged proteins. A significant portion of each core histone is composed of an unstructured N-terminal tail domain, while the bulk of the protein mass consists of a mainly α-helical C-terminal domain. This C-terminal domain is responsible for forming the histone-histone interactions necessary for the assembly of the octameric structure. Despite limited primary sequence similarity, all four core histones exhibit a conserved structural motif known as the histone fold [18]. The histone fold motif consists of two short α-helices (α1 and α3) flanking a longer central α-helix (α2), with connecting loop/β-segments (Figure 1B). This motif facilitates a specific protein-protein interaction referred to as a handshake interaction, leading to the formation of heterodimers between histones H2A with H2B and H3 with H4. The histone fold motif is not exclusive to histones but is also found in various protein complexes [19,20].

A B

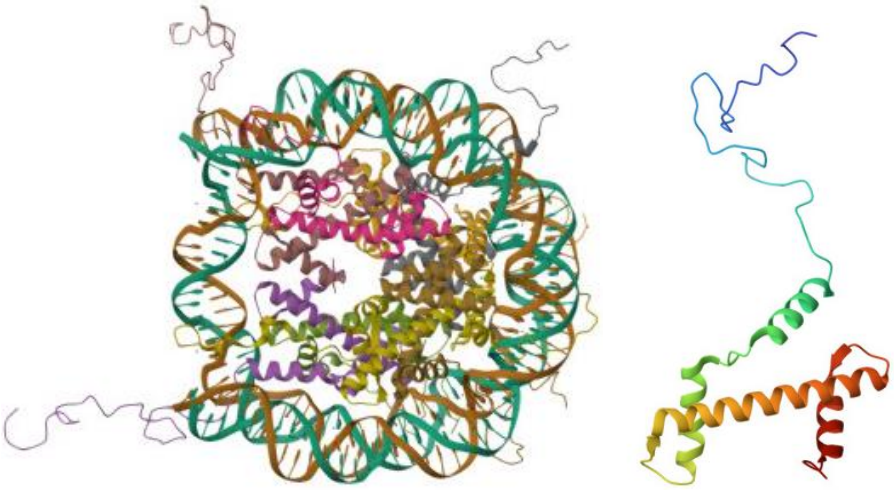


Figure 1 – X-Ray Structure (1.9 A Resolution): A) nucleosome core particle, B) H3 histone [21]

The linker histone, also known as H1, is one of the five major families of histone proteins found in eukaryotic cells. In metazoans, linker histones exhibit a conserved tripartite structure consisting of a flexible N-terminal tail, a globular domain with a winged helix fold, and a long, highly basic carboxy-terminal tail that lacks a defined structure. The globular domain of linker histones has a preference for binding to the nucleosome [22].

The H3/H4 dimers self-associate to form a stable tetramer, while a single H2A/H2B dimer associates with each end of the H4/H3:H3/H4 tetramer. This arrangement creates a symmetrical string of four tetramers, which generates a helical ramp of DNA contact sites. The stability of the histone octamer is dependent on DNA wrapping or high salt concentrations due to the positive charges of the histone proteins. The contacts between the DNA and the histones involve interactions between phosphate residues and certain amino acids such as lysine and arginine, as well as main chain amide nitrogens. Each of the fourteen main backbone contact points along the 147 bp length of the nucleosome core DNA is a highly conserved arginine residue interacting with the minor groove. This arginine plays a crucial role in precisely positioning the DNA, facilitating overall DNA bending, and shaping the superhelical structure. They form hydrogen bonds and non-polar interactions with deoxyriboses. Mutations in these specific positions, known as SIN mutations in yeast, have been associated with increased nucleosome mobility and accessibility [23].

2.2.2 Histone variants

Unlike the standard histone proteins, which are encoded by multiple gene copies and expressed during the S phase of the cell cycle, mammals have evolved specific histone variants for each histone family, including somatic variants (such as H2A.X, H2A.Z, Macro-H2A, H2A.B, and H3.3) and testis-specific variants (such as TS H2A.1, H2A.L, TS H2B.1, H2B.L, subH2B, and H3T). Histone chaperones are responsible for depositing these variants onto chromatin, and they also interact with other chromatin modifiers [24]. In contrast to canonical histones, histone variants are usually represented by single-copy genes and exhibit expression throughout the cell cycle [25].

Three important concepts arise when considering the functional role of histone variant deposition into chromatin:

1) variants can confer distinct physical properties to nucleosomes due to sequence differences;
2) they facilitate the formation of unique chromatin-associated complexes at appropriate genomic regions through interactions with specific domains and chaperone systems;
3) nonreplicative histone variants often exhibit specific PTMs or carry unique PTMs patterns compared to their replication-deposited counterparts [26].

H3 variants can be categorized into two subgroups: replicative and replacement variants, and their abundance varies across different species. Histones involved in DNA replication exhibit a peak in expression during the S phase, constituting the primary histone source during this process. They are deposited in a DNA synthesis–coupled manner, represented by H3.1 and H3.2 in humans. Both of these variants must be hardly distinguishable by their sequence with 99 % identity. On the other hand, replacement variants, typically expressed regardless of the S phase, are incorporated independently of DNA synthesis. In humans, one of the well-characterized replacement variants is H3.3 [27]. Apart from H3.3, there are several other replication-independent H3 variants: variants with tissue-specific expression, like H3.4 and H3.5 found in the testis; unique to certain species, such as H3.Y.1, H3.Y.2; H3.5 specific to primates (hominid lineage); and associated with centromeres cenH3-CENPA[28]. Among all known variants of histone H3, there are indeed distinct variants, such as cenH3-CENPA, but the majority are still practically indistinguishable from each other. Nevertheless, histone chaperones possess the ability to discriminate between H3.1/H3.2 and H3.3 variants based on an amino acid motif located in the core of the protein and in the N-termini tail, that differs in only a few amino acids [26].

The H2A family exhibits the highest sequence diversity among identified variants. In humans, four classes of replication-independent H2A variants have been discovered: H2A.X, H2A.Z, macroH2A, and testis-specific short H2A variants (H2A.B, H2A.L, H2A.P, and H2A.Q). While certain H2A variants are known to associate with dedicated chaperones and ATP-dependent chromatin remodelling, the mechanisms governing the localization of many H2A variants to specific genomic regions remain unclear [29].

While both H3 and H2A are involved in octamer formation, H4 and H2B primarily interact with other histones, suggesting that the evolutionary forces acting on core histones have resulted in significant diversification for H2A and H3, but not to the same extent for H2B and H4. However, minor sequence differences are observed in H2B variants: H2B, H2BE, TSH2B, H2B.W and others. They can influence the histone-fold domain of these proteins, destabilizing histone–histone interactions in a similar way to that of the H2A variants [26].

Histone variants, being the primary source of new histones in postmitotic terminally differentiated tissues, may play a significant role in brain-related diseases such as neurodevelopmental and neurodegenerative disorders. Studies indicate that variants like H3.3 and H2A.Z are crucial for brain function [26]. Considering the testis-specific nature of many histone variants, alterations or modifications of these variants may lead to male infertility [30]. Furthermore, recurrent mutations in genes responsible for chromatin-associated proteins, including histone variants and their deposition mechanisms, have been observed in various cancers and developmental syndromes [31].

2.3 Histone PTMs

The conserved tail domains, comprising approximately 25-30 % of the mass of core histones, extend beyond the structured regions that form the DNA spool. These domains, located at the N-terminal part of all four core histone proteins and the C-terminus of histone H2A, are commonly referred to as unstructured regions, as they adopt random coil conformations when the proteins are free in solution or released from their binding sites on nucleosomes in high salt conditions. Tail domains are rich in lysine and arginine residues, with significant amounts of glycine, alanine, and threonine. They serve as targets for enzymes involved in PTMs that play a crucial role in epigenetic signalling [23].

A wide range of PTMs, including acetylation, phosphorylation, methylation, ubiquitination, ADP-ribosylation and others occur on the tail domains of histones, indicating the existence of a histone code that may be interpreted by other proteins or protein complexes [32,33].

2.3.1 Acetylation

Histone acetylation is a biochemical process that involves the addition of acetyl groups to histone proteins (Figure 2). The discovery of enzymes responsible for catalysing acetylation, such as lysine acetyltransferases (KATs) and lysine deacetylases (KDACs), has provided compelling evidence of the crucial role played by acetylation. In the H3 histone of various species, lysines 9, 14, 18, and 23 are the prominent sites of acetylation. Similar to the functionally redundant tail of H3, specific lysines in H4 undergo acetylation during distinct cellular processes [32].

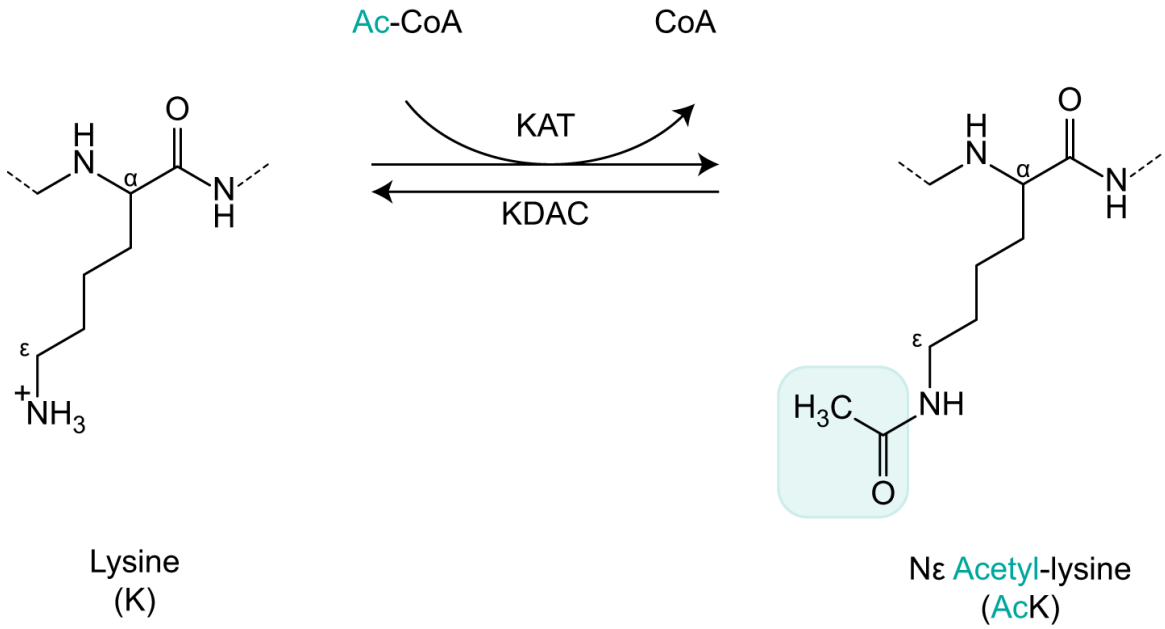


Figure 2 – Schematic overview of lysine acetylation and deacetylation, where lysine acetyltransferases (KATs) catalyse the transfer of an acetyl group (-$COCH_3$) from acetyl-CoA (Ac-CoA) to the ε-amino group of lysine (K) side chains with the formation of Nε Acetyl-lysine (AcK). The reverse reaction is carried out by lysine deacetylase (KDAC) [34]

The addition of acetyl groups neutralizes the positive charge of the lysine residues, enhancing hydrophobicity and weakening the electrostatic interaction between the histones and DNA [35]. The complex process initiated by this alteration in the histone structure allows for a more open chromatin structure, making the DNA more accessible to transcription factors and other regulatory proteins. Histone acetylation is generally associated with transcriptional activation, as the more open chromatin structure facilitates gene expression by allowing the transcription machinery to access the DNA and initiate transcription [36]. Conversely, deacetylation, the removal of acetyl groups, is associated with transcriptional repression, promoting a more compact chromatin structure and inhibiting gene expression.

In addition to transcription, specific lysine residues in H3 undergo acetylation during other biological processes. For instance, during DNA replication, newly synthesized histones are rapidly assembled onto the replicated DNA in a pre-acetylated state, which is subsequently removed as chromatin matures after replication finishes [32]. The initiation of DNA replication occurs at specific genomic sites known as the origins of replication. The activation and positioning of these origins are closely linked to the

acetylation of H3 and H4 histones and the increased accessibility of chromatin. This indicates that DNA replication not only involves the faithful duplication of genetic material but also the transmission of epigenetic information [37].

2.3.2 Methylation

Unlike acetylation, histone methylation presents a significantly more complex picture. Distinct histone lysine residues exhibit varying and sometimes opposing functions when methylated, and different degrees of methylation (mono-, di-, or trimethylation) on the same residue may lead to vastly different outcomes [38]. Additionally, histones can undergo methylation on arginine residues. Arginine modifications have demonstrated both positive and negative roles in transcription regulation. Currently, more than 20 methyl marks on lysine and arginine residues have been identified, with the five lysine residues on histone tails receiving the most attention [39].

Histone methylation can be divided into two main categories: lysine methylation and arginine methylation, which are catalysed by histone lysine methyltransferases (KMTs) and protein arginine methyltransferases (PRMTs), respectively (Figure 3). To date, over ten KMTs have been identified in humans, and many of them exhibit unique substrate specificity. Except DOT1L, all HKMTs possess an evolutionarily conserved SET domain responsible for their lysine methylation activity [40].

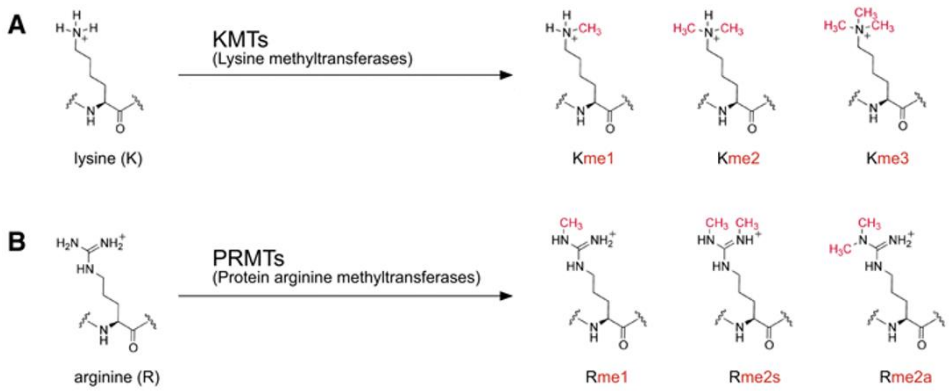


Figure 3 – Schematic overview of A) lysine and B) arginine methylation. The catalysis of mono-, di-, or tri-methylation (Kme1, Kme2, and Kme3 respectively) of the lysine ε-NH_2 group is carried out by lysine methyltransferases (KMTs). Protein arginine methyltransferases (PRMTs) are able to modify the guanidinium group of arginine residues, leading to the formation of either mono-methylated (Rme1) or symmetric (Rme2s) or asymmetric (Rme2a) di-methylated arginine. Adapted from Wesche *et al.*, 2017 [41]

Until recently, histone methylations were widely believed to be permanent marks, similar to DNA methylation. However, since 2004, a significant number of enzymes have been identified that possess the capability to remove methyl groups from histone

lysine residues through processes like amine oxidation, hydroxylation, or deimination. These enzymes responsible for demethylating histones are categorized into two families: lysine-specific demethylase (LSD) proteins and Jumonji domain (JMJD) containing (JMJC) proteins [40].

Methylation of certain histone lysine residues is associated with active gene transcription (e.g., H3K4me3, H3K36me3,). These modifications lead to an open chromatin structure, allowing transcription factors and the transcriptional machinery to access the DNA and initiate gene expression. Methylation of other lysine residues can participate and DNA replication (H3K79) or can lead to gene repression (e.g., H3K9me3, H3K27me3, H4K20me3) by promoting a compact chromatin structure that inhibits gene transcription. This can occur through blocking the access of transcriptional activators or recruiting repressive protein complexes [39]. Methylation of histones, especially H3K27me3, is crucial in X chromosome inactivation, a process that balances gene expression between males and females [42]. Histone methylation patterns can be heritable and play a role in cellular memory [43]. They are involved in maintaining cell-specific gene expression profiles, contributing to cell identity and differentiation [39]. Certain methylated histones are involved in the repair of DNA damage and the maintenance of genome stability [39]. They aid in the recruitment of repair proteins to damaged DNA sites. Aberrant histone methylation is implicated in various diseases, including cancer, neurological disorders, and developmental disorders. Dysregulation of histone methylation can lead to uncontrolled gene expression and disease progression.

2.4 MS-based approaches of histone analysis

A key challenge for scientists studying histones and their variants lies in deciphering the complexity of histone variants, identifying PTM combinations, and understanding their role in a pathophysiological context. Historically, this challenge has been addressed using antibody-based methods, such as western blot, immunofluorescence analysis, and chromatin immunoprecipitation (ChIP) [44]. ChIP techniques are essential for the study of epigenetic modifications in the genomic context but are limited in the analysis of the co-occurrence of distinct histone variant/modification combinations. For the less-studied modifications, antibodies are not commercially available. The sequence homology between histone variants and the large array of existing histone PTMs make the generation of highly specific antibodies challenging. Also, antibody-based assays require knowledge of the type and position of the modification of interest. As a result, histone antibodies are not always specific for their targets and often generate cross-reactivity [45]. Additionally, two neighbouring PTMs within the same histone may cause epitope occlusion, and thus impede recognition of the target PTM [46]. So, these approaches lack high-throughput capabilities, and each PTM must be analysed

separately, which makes the analysis of multiple occurring PTMs within the same histone expensive as well as time and effort-consuming.

As an alternative, mass spectrometry (MS) can provide a vast amount of information in a high-throughput way and without prior knowledge. With different techniques, it can be used for distinguishing histone variants, discovery of histone-binding proteins, analysis of combinatorial histone PTMs, *etc*. Historically, three main MS-based approaches, namely bottom-up, middle-down, and top-down are used, that was developed and adapted for the histone study (Figure 4). They differ in sample preparation, analysis procedures, and processing of the obtained results, but all of these approaches have been extensively applied to histones and successfully mapped over 500 different PTMs [45].

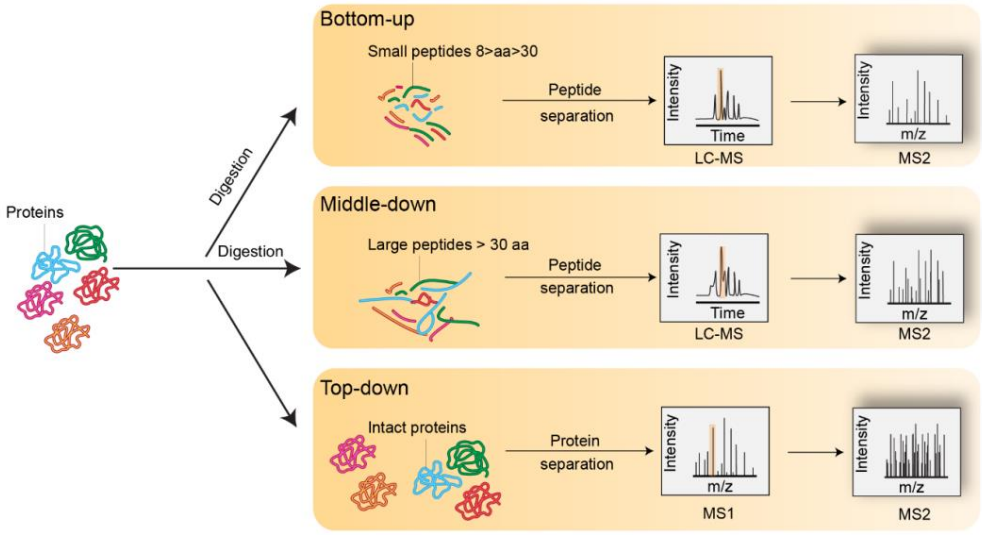


Figure 4 – Short description of bottom-up, middle-down and top-down MS approaches [47]

In the bottom-up approach, proteins are digested into short peptides which are then analysed. For protein digestion, the serine protease trypsin is used almost exclusively. Cleaving after lysine or arginine residues, trypsin generates small peptides with a charged residue at the C-terminal position. Peptides are ideal analytes for MS analysis because they are easy to solubilize, separate and ionize. Since peptides are the observed unit in bottom-up proteomics all information about the presence and abundance of proteins in the sample are inferred from the identified peptides. After peptides digestion, they lose their connectivity to the corresponding protein of origin, complicating the process of protein identification and determination [47].

In top-down proteomics, intact proteins are usually analysed. Here proteins are observed directly, and the relationship between the base amino acid sequence and the PTMs is preserved. However, top-down analysis is very complicated and there are

many challenges that must be overcome including but not limited to the low abundance of analyte, the low signal-to-noise ratio of proteins with large molecular weight, and low solubility of intact proteins [48].

As suggested by its name, the middle-down method falls between the other two approaches. The methodology differs from bottom-up classical sample preparation in the use of specific enzymes to yield considerably longer peptides. For instance, Glu-C can be used to digest histone H3 at the C-terminal of glutamic acid, producing longer peptides [47]. However, due to their greater length compared to bottom-up peptides, middle-down encounters analogous challenges to top-down and exhibits lower sensitivity than bottom-up.

In the present day, bottom-up proteomics has reached a mature stage and serves as the primary approach for MS analysis. Conversely, top-down and middle-down methodologies, while promising, demand specialized expertise to be effectively employed. The following paragraphs will focus on the bottom-up approach, as it was used in this work.

2.4.1 LC-MS and LC-MS/MS in histone analysis

Liquid Chromatography-Mass Spectrometry (LC-MS) stands as a powerful analytical method widely applied to investigate proteins, peptides, and their PTMs. This technique combines the separation proficiency of liquid chromatography with the analysis capabilities of MS, enabling the precise and sensitive determination and quantification of proteins in intricate biological specimens.

As evident from the name, MS analysis in this case is preceded by the separation of substances using various chromatographic methods. Then the isolated substances are introduced into the MS where they undergo ionization. In both LC-MS and LC-MS/MS analysis electrospray (ESI) source is usually used, which generates $M+2H^+$ and $M+3H^+$ ions from a solution [49]. Electrospray ionization operates through the application of a high voltage (2–6 kV) between the emitter at the end of the separation conduit and the mass spectrometer's inlet. Physicochemical processes within ESI encompass the generation of an electrically charged spray (Taylor cone), which involves the creation and desolvation of analyte-solvent droplets, assisted by a heated capillary. These ions are then fed into the mass analyser. In the first mass analyser, they are separated based on their mass-to-charge ratio (m/z), forming a spectrum that displays peptide masses and their respective abundances (MS1) [50]. Certain ions selected in this analysis are directed to the collision cell for fragmentation, producing product ions in the MS/MS (MS2) spectra. These spectra provide valuable details about the peptide sequence and structure. Utilizing these MS and MS/MS spectra, specific search engines, such as Mascot [51] together with protein sequence databases like UniProt [52] are used to determine the proteins and peptides present in the sample.

In bottom-up proteomics, for LC-MS and LC-MS/MS methods, analysers based on Ion trap instruments and their combinations are most commonly used. Standalone ion trap instruments are mostly used in LC-MS for protein identification studies involving complex samples and whole cell lysates, benefiting from their fast-scanning rates and high sensitivity that provide extensive coverage [50]. The Orbitrap, used in this study, is based on the principle of static electrostatic fields that are used to achieve orbital trapping of ions, enabling ions to orbit around a central electrode and oscillate in the axial direction (Figure 5). The orbitrap instrument employs a fast Fourier transform (FFT) algorithm to convert time-domain signals into mass-to-charge spectra. This analyser boasts a high resolution of up to 150,000, exceptional mass accuracy (2–5 ppm) and a broad mass-to-charge range of 6000. Regarding data analysis, high mass accuracy enables the integration of various approaches like database search, *de novo* search, Peptide Mass Fingerprint (PMF) search, and library lookup leading to improved coverage and accuracy in results.

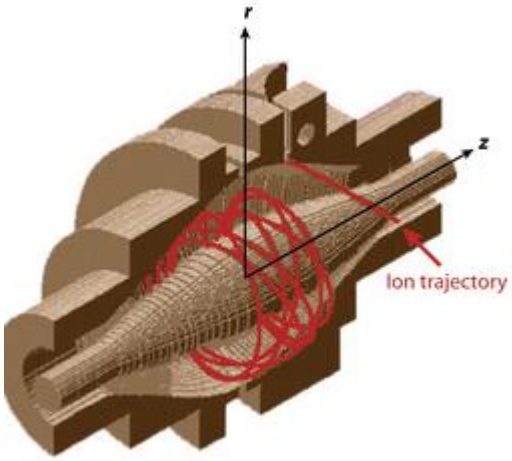


Figure 5 – Orbitrap analyser operating principle depicting the trajectory of ion movement [50]

As one of the bottom-up proteomics methods, LC-MS and LC-MS/MS also require prior enzymatic protein digestion, which means that the chromatographic separation involves the obtained peptides, not the entire protein. The prevalent separation method usually used in histone study is reverse phase chromatography (RPLC or RP-HPLC), which is based on the hydrophobic interactions of the analyte with the stationary phase of the column. A key benefit of RP-HPLC is its compatibility with ESI, utilizing buffers that align with this ionization method.

At present LC-MS/MS is actively used for the analysis of histone variants and the localization and identification of PTMs. As an instance, it can detect acetyl modifications across four lysine sites (K5, K12, K8, and K16) in the N-terminal fragment of

histone H4, distinguishing between isobaric amino acid variants and combinations of PTMs, quantify positional isomers and others [25].

2.4.2 MALDI-TOF

The majority of methods used in modern bottom-up proteomics are utilizing LC–MS or LC-MS/MS. Despite their popularity, other methods, such as matrix-assisted laser desorption/ ionization (MALDI) MS with a Time-of-Flight (TOF) detector, are often used in practice. MALDI-TOF MS finds extensive applications in various biochemical analyses including bacterial typing, proteomics, and MS imaging, as well as in polymer and inorganic chemistry for nanomaterial characterization, primarily due to its rapidity and sensitivity.

The experimental method of MALDI-TOF MS can be summarized as follows (Figure 6). First, the sample is mixed with a suitable matrix material and dropped onto a clean MALDI sample plate. In MALDI, an organic matrix is typically employed, which possesses the ability to absorb laser energy and disperse sample molecules [53]. Frequently used matrices encompass compounds like 2,5-dihydroxybenzoic acid (DHB), α-cyano-4-hydroxytrans-cinnamic acid (HCCA), sinapinic acid (SA), and others [54]. After drying, samples undergo ionization. Under laser irradiation from a pulsed laser of the nanosecond class (typically a nitrogen laser at 337 nm or a Nd:YAG laser at 355 nm wavelength [55]), the matrix molecules absorb the laser energy and convert it into electronically excited energy. This rapid process causes the solid mixture of the matrix and analyte to transition instantly into a gaseous state. Charge transfer occurs, inducing ionization of the analyte through collisions involving uncharged neutral molecules, matrix ions, protons, electrons, and metal cations. Subsequently, ions generated via photoablation, and photoionization are propelled by an electric field into the mass analyser for analysis of their mass-to-charge ratio (m/z).

Typically, TOF is chosen as the preferred mass analyser to be coupled with the MALDI ion source for separation and detection. Sample ions traverse through the TOF tube at velocities inversely linked to the square root of their m/z values. Measuring the flight time allows the calculation of the ions' m/z. Because of this, the TOF mass analyser has no strict limitations of mass range, which is important for coupling with MALDI, which generates singly charged ions with large m/z [55].

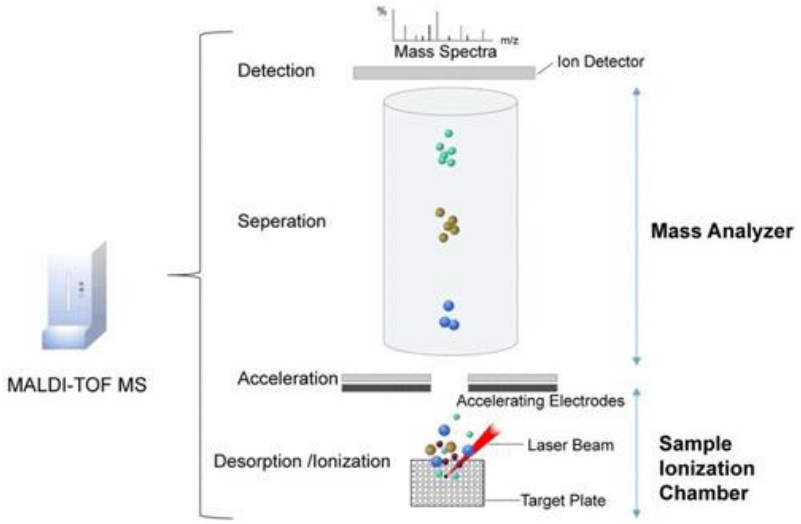


Figure 5 – Schematic description of MALDI-TOF MS [55]

Regarding histone analysis, MALDI-TOF is excellently suited for a rapid screening of new KDAC inhibitory molecules as well as a preliminary description of basic molecular responses of KDAC inhibitors under a wide range of cultivation conditions [56].

2.4.3 Problems of histone analysis

Despite recent advancements in MS techniques, significant challenges still exist in areas such as histone preparation, the MS process itself, and subsequent data analysis.

A characteristic feature of bottom-up proteomics is the necessity of enzymatic digestion before analysis. This simplifies the subsequent analysis of spectra but introduces challenges related to the physicochemical properties of histones. To digest proteins into peptides, different enzymes can be used. The commonly used enzyme for protein digestion is trypsin, which cleaves at the C-terminal of lysine and arginine residues unless they are followed by a proline. However, the direct use of trypsin on histones is problematic. The N-terminal tails of histones are lysine- and arginine-rich so trypsin digestion results in short peptides that are poorly retained on the chromatography column and can be lost. There are non-reproducible mis-cleavages, leading to a critical error in subsequent data processing and quantification hampering [57]. Other enzymes can be used but they are much less specific than trypsin and also result in non-reproducible digests [58].

The efficiency of RP-HPLC separation is associated with the polarity of the analyte as the time it spends in the column before being eluted (retention time) depends on the strength of attraction between the hydrophobic analyte and the nonpolar stationary phase of the column. Many peptides obtained after trypsin digestion are too small and hydrophilic or can be co-eluted in the same retention time, which affects the quality of the spectrum [58].

Histone variants distinguishing also poses a significant hurdle for MS-based techniques. In bottom-up approaches, the ability to differentiate variants relies on the identification of distinctive peptides. However, various factors like ionization efficiency, precursor selection, fragmentation and identification algorithms can hinder the detection of unique peptides [45].

To overcome above mentioned issues, chemical derivatization of amine groups has been introduced for histone preparation prior to LC-MS/MS [58].

2.5 Histone chemical derivatization

To address issues related to sample preparation for MS analysis modifying lysine side chains through derivatization has been used to prevent trypsin cleavage at lysine sites. This process typically includes modifying the free amine groups on N-termini and ε-NH_2 groups of unmodified or endogenously monomethylated lysines by reacting them with organic anhydride to create amides before employing trypsin cleavage (Figure 6) [58]. Anhydrides are preferred due to their heightened reactivity compared to their corresponding acids, leading to more efficient chemical reactions that require milder conditions and less time for the intended modification. Moreover, anhydrides often exhibit better solubility in organic solvents, facilitating easier handling and dissolution in the reaction mixture. This characteristic proves advantageous in specific reactions and purification processes. The selection of anhydrides is usually based on their compatibility with specific reaction conditions, temperature, and solvent systems, allowing optimization of the reaction for the desired modification.

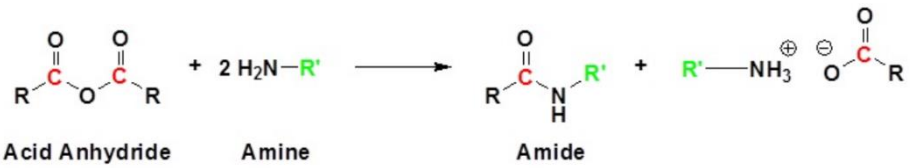


Figure 6 – General scheme of the reaction between acid anhydride and amine with the formation of amide bond [59]

Typically, sample preparation in bottom-up proteomics involves two stages of derivatization: one at the beginning, prior to enzymatic digestion, and the second one afterwards to label the newly formed N-termini of peptides after cleavage [58]. The first stage is crucial to block all lysine residues, ensuring that trypsin digestion occurs only on the C-terminal of arginine residues. This process maintains the efficiency and reproducibility of trypsin digestion while achieving the desired specificity. Subsequent derivatization at the peptide level is required for effective neutralization of the charge at the N-termini. Both additional derivatization of unmodified and monomethylated lysine residues make histone peptides less hydrophilic.

This modification enables the resolution of the peptides through standard RP-HPLC. Moreover, the peptides generate doubly and triply charged ions, resembling pseudo-tryptic-like fragments, simplifying the interpretation of their MS/MS spectra. This also provides an advantage over using the Arg-C protease, which cleaves at arginine residues but does not remove charge from lysine residues [58]. All these enhancements significantly improve LC retention, separation, and resolution of the peptides analysis.

Nevertheless, even after derivatization, there are still issues related to insufficient impart on the hydrophobicity to short peptides (*e.g.*, H3T3KQTAR8), where signal loss is common due to poor retention [57]. Furthermore, there are issues related to missing arginine residues in the N-terminal part of the sequence, which led to the formation of long peptides with multiple lysins. [60]. Clearly, such long peptides are difficult to characterize using bottom-up approaches due to incomplete ion series and the high combinatorial patterns of PTM sites.

2.5.1 Propionylation

Derivatization using propionic anhydride ($C_6H_{10}O_3$; Prop) stands out as one of the most widely used derivatization strategies in practice. Propionylation has demonstrated excellent results in terms of peptide MS signal when studying the digestion of unmodified histone tails, indicating highly efficient peptide digestion and ionization. It has been used in a huge list of histone PTMs research, such as *Trypanosoma brucei* histone [61] and *Neurospora crassa* H2B histone modifications study [62], H3 histone modifications profiling using LC-MS [63] and others.

Nonetheless, this derivatization approach has limitations. It doesn't significantly increase the hydrophobicity of short peptides like those spanning amino acids 3 to 8 of histone H3. This can lead to signal loss due to poor retention. Additionally, various peptides exhibit different ionization efficiencies, primarily due to differences in their sequences and endogenous modifications. Also, since propionylation is a natural PTM, it may interfere with the investigation of this specific modification [64]. Lastly, incomplete derivatization (underpropionylation) or side reactions of propionylation often occur: a non-specific side reactions on serine, threonine and tyrosine residues hydroxyl groups (overpropionylation) [65]. These pitfalls hinder identification and impair the direct comparison of precursor intensities of biologically modified peptides.

2.5.2 Alternative derivatization agents

The lack of any standard protocol for the derivatization of various histone variants and their modifications leads to an active search for the most successful and efficient candidates. For instance, in the study conducted by Sidoli *et al.* [64], several commercially available anhydrides (including Prop) were assessed, and chosen for their safety and hydrophobic properties (Table II).

The standard derivatization protocol was followed, with a slight modification for benzoic anhydride, involving a single cycle of the derivatization process. To evaluate efficiency at both the protein and peptide levels, different samples were employed:

- synthetic unmodified histone H3 tail,
- synthetic modified histone peptides,
- a histone extract from cell lysate [64].

Of all the tested anhydrides, in the end, benzoic and valeric anhydrides were chosen as the most promising. However, they have a set of advantages and disadvantages that do not allow their full practical use as an alternative to propionylation.

To address the challenge in analysing the H3K4me3 mark, which has a low natural abundance but a highly restricted genomic localization strongly associated with active gene promoters and enhancers, a hybrid method was proposed [66]. The method was tested on synthetic peptides corresponding to amino acid residues 1-17 of histone H3 (unmodified and modified) mixed with endogenous histones. In this protocol, two anhydrides were used at different stages of sample preparation:

1) propionic anhydride, at the protein level, before trypsin digestion;
2) phenyl isocyanate (PIC, Table II), at the peptide level, after digestion.

It was observed that all identified peptide forms exhibited later retention times compared to their Prop counterparts. Additionally, employing the Prop-PIC label led to substantial enhancements in peak areas for all peptide forms, with the most notable improvements observed for H3K4me2 and -me3, demonstrating an approximate two orders of magnitude increase in peak areas. However, this approach is impeded by side reactions of propionic anhydride with the hydroxy groups of serine, threonine, and tyrosine, causing excessive propionylation. Moreover, the reaction efficiency can be inadequate, resulting in insufficient propionylation, which is particularly challenging in pre-digestion labelling techniques. Additionally, certain propionylated isobaric peptides, such as positional isomers of monoacetylated peptide H3K18-K23 (K18ac and K23ac), were co-eluted [66]. Overall, this method can be applied to solve specific narrow tasks. However, after hybrid derivatizations result interpretation becomes significantly more complicated, which can lead to errors and incorrect conclusions.

As an alternative, the following method was proposed by Kuchaříková H. *et al* [60]. In their study, a microwave-assisted labelling procedure for histones before bottom–up analysis was established, employing trimethylacetic anhydride ($C_6H_{10}O_3$; TMA; Table II) for lysine derivatization (Figure 8). Microwave irradiation facilitated a stepwise labelling process of histone proteins and peptides, achieving a high derivatization efficiency. The efficacy of TMA labelling was demonstrated, encompassing chromatographic characteristics of derivatized isobaric peptide forms and variations in mass spectrometric data. They devised a derivatization protocol and effectively utilized it to differentiate various histone PTM states in human cell cultures treated with a KDAC inhibitor and controls.

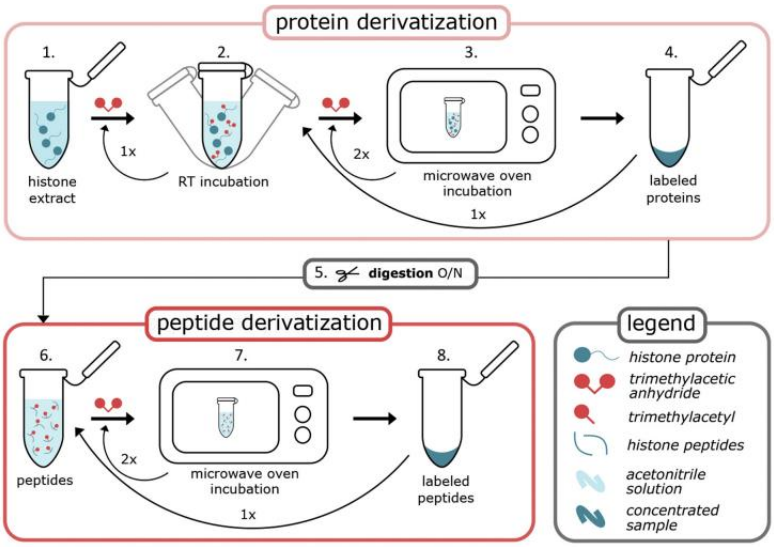
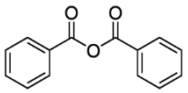
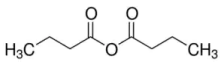
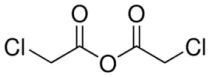
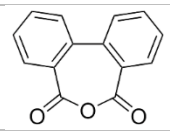
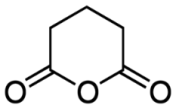
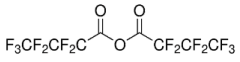
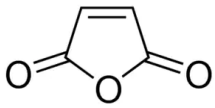
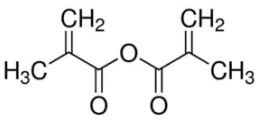


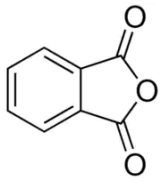
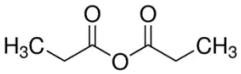
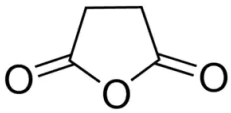
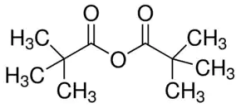
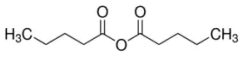
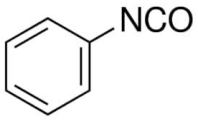
Figure 7 – Illustrative scheme of histone derivatization protocol with TMA [60]

The experimental comparison of the derivatization process using Prop and TMA on H3 and H4 histones showed that overall, the performance of labelling with both anhydrides is comparable (98% desired histone H3 peptides after TMA compared to 92% desired histone H3 peptides after propionylation)[60]. Regarding the efficiency of chromatographic separation, it was demonstrated that peptides labelled with TMA exhibited increased hydrophobicity, and their chromatographic peaks were more uniformly distributed across the LC. The observed enhancement in hydrophobicity significantly impacted the identification and quantification of various post-translationally modified peptide forms. Finally, enhanced separation facilitated the identification of all 16 modified forms of H4G4-R17, a feat not achievable with Prop.

However, the use of TMA does not completely exclude the formation of side products resulting from the reaction of acid anhydrides with hydroxyl groups of serine, threonine, or tyrosine, but is accompanied by a minimal amount of them. TMA labelling also does not completely solve the issue of the frequently disappearing H3T3-R8 peptide. In conclusion, there is no universal method for characterizing PTMs across all core histones and their variants, there why the analytical approach must be customized based on the primary research goals [60].

Table II: Overview of previously tested alternative derivatization agents

Anhydride	Structure	Acid pKa	Comments
Benzoic [64]		4.20	A separate protocol is needed Better signal intensity Equal retention time drift Significantly lower derivatization yield Significantly more complex chromatogram Significantly deviating signal intensity
Butyric [64]		4.82	Low retention time drift Natural PTM Significantly lower derivatization yield
Chloroacetic[64]		2.87	Very low signal intensity Lowest score after database searching
Diphenic [64]		3.20/ 5.06	Not detectable in the LC-MS analysis
Glutaric [64]		4.34	Low retention time drift Does not generate carboxyl groups
Heptafluoro-butyric [64]		0.4	Not detectable in the LC-MS analysis
Maleic [64]		1.83	Does not generate carboxyl groups High signal intensities Increased the differences in ionization efficiency between the peptides Lowest score after database searching
Methacrylic [64]		4.65	Large bias in ionization efficiency High signal intensities

			Increased the differences in ionization efficiency between the peptides
Phthalic [64]		2.89 and 5.51	Does not generate carboxyl groups Lowest score after database searching
Propionic [64,65]		4.88	Standard
Succinic [64]		4.21 and 5.64	Low retention time drift (decrease RT) Does not generate carboxyl groups
Trimethylacetic [60]		5.03	A separate protocol is needed Equal derivatization yield Better signal intensity High retention time drift Improved peptide identification
Valeric [64]		4.84	Better signal intensity Equal retention time drift Significantly lower derivatization yield Significantly deviating signal intensity
Phenyl isocyanate [66]		-	High retention time drift (increase RT) High reaction specificity Derivatization only on peptide label

3 The goal of the work

The aim of this thesis was to develop and improve methods for sample preparation and analysis of histones and their PTMs using MS. The majority of the work focused on selecting the most effective alternative derivatization agent and optimizing labelling conditions with the chosen anhydride for the most efficient labelling of lysine amino groups and N-termini of peptides.

The optimization of methods was conducted on various model systems, with the goal of applying this methodology to characterize histones and their PTMs in different biological systems.

4 Materials and Methods

4.1 Materials

To assess the performance of various organic anhydrides for histone derivatization synthetic peptide, recombinant protein, as well as histones extracted from cell culture, were used. To determine the effectiveness of derivatization at the peptide level, a synthetic peptide with a sequence ARTKme2QTARKS (Clonestar Peptide Services, s.r.o.; concentration of the stock solution unknown) identical to the dimethylated N-terminal region of histones H3.1 was used. To determine the effectiveness of derivatization at the protein level, recombinant histone H3.2 (New England Biolabs Inc., 1.0 mg/mL) and histones obtained from MEC-1 B-CLL cells (B chronic lymphocytic leukaemia; Leibniz Institute DSMZ (German Collection of Microorganisms and Cell Cultures GmbH)) were used. The MEC1 cell line was cultured in IMDM (Iscove's Modified Dulbecco's Medium) with phenol red, supplemented with 10 % FBS (fetal bovine serum), 2 mM L-glutamine, and 100 IU of penicillin/streptomycin. All samples were prepared as triplicates.

The following anhydrides were used for labelling: propionic (Sigma-Aldrich, Co; CAS Number:123-62-6; ≥99 %), trimethylacetic (Sigma-Aldrich, Co; CAS Number: 1538-75-6; 99 %), trichloroacetic (Sigma-Aldrich, Co; CAS Number: 4124-31-6; 95 %), trifluoroacetic (Sigma-Aldrich, Co; CAS Number: 407-25-0; ≥99 %), 2-bromoisobutyric (Sigma-Aldrich, Co; CAS Number: 42069-15-8; 95 %), 2-sulfobenzoic acid cyclic anhydride (Sigma-Aldrich, Co; CAS Number: 81-08-03; 90 %), benzoic (Sigma-Aldrich, Co; CAS Number: 93-97-0; ≥95 %), 3,3,3-trifluoropropionic (Fluorochem Ltd; CAS Number: 58668-07-8; 90 %) and phenyl isocyanate (Sigma-Aldrich, Co; CAS Number: 103-71-9; ≥98 %). To prepare the derivatization agent, the anhydrides were diluted in acetonitrile (ACN; Honeywell International Inc; CAS Number: 75-05-8; ≥99.9 %).

4.2 Methods

4.2.1 Sample preparation

For each sample, the MEC-1 histone extract (1 µg) or H3.2 recombinant histone stock solution (5 µL) was diluted with 50% ACN (v/v) to the final sample volume of 10 µL. For peptide labelling, 1 µL of ARTKme2QTARKS peptide stock solution was used and then diluted with 50% ACN (v/v) to achieve a final sample volume of 10 µL. For experiments where analysis was conducted after each derivatization cycle, the amount of peptide and final solution volume for each sample were doubled. Specifically, 2 µL of synthetic peptide stock solution was used to obtain 20 µL of solution.

4.2.2 Chemical derivatization at protein and peptide levels

0.5 µL of concentrated NH_4OH was added to maintain a pH of 8. Furthermore, the derivatization agent was prepared by mixing an anhydride with concentrated ACN in a ratio corresponding to the goals of the experiment (Chapter 5), added to the sample without interruptions, vortexed, and spun down for several seconds. The mixture was prepared fresh for each 3 samples. The derivatization agent and 1 µL of NH_4OH were added to the sample as quickly as possible, and the mixture was mixed using a vortex. After each step, the pH of the sample was checked. If it was different from pH 8, the required amount of NH_4OH was added. Next, the mixture was incubated (incubation conditions were optimized and specified in Chapter 5) and concentrated in SpeedVac (Savant SPD121P SpeedVac; Thermo Fisher Scientific Inc.) to 3-5ul, 35°C. Afterwards, the preparation of the derivatization agent, its addition to the sample, and incubation were repeated. Overall, the derivatization cycle was performed two to three times (corresponding with the goals of the experiment (Chapter 5)). After each of these cycles, the concentrated samples underwent a dilution with 10 µL of ACN. After derivatization at the protein level, the proteins were digested using trypsin. Derivatization at the peptide level followed the same principle and occurred under the same conditions.

4.2.3 Incubation conditions

In optimization experiments, two procedures were used for sample incubation with the derivatization agent: a procedure involving microwave irradiation and a procedure involving sample incubation in a thermomixer.

In the majority of experiments, incubation using a thermomixer (Thermomixer Comfort (Eppendorf SE)) was predominantly used. The temperature in this work ranged from room temperature (RT) to 40 ^{O}C. The shaking speed was set at 1000 rpm. The incubation time varied depending on the experiment's goals and ranged from 20 minutes to 16 hours per derivatization cycle. A detailed description of the conditions used in the experiment is provided in Chapter 5.

The derivatization of samples using microwave irradiation proceeded as follows. Samples were incubated in a microwave oven at the power of 350 W. The incubation time was 1 minute, and the cycle was repeated three times. After each cycle, the samples were cooled and spun down for several seconds.

4.2.4 Enzymatic digestion

To the concentrated labelled protein diluted in 100 mM NH_4HCO_3 up to 10 µL SOLu-Trypsin Dimethylated (Merck KGaA) was added in a ratio of 1:40 (E:S) for both recombinant H3.2 histone and MEC-1 histone extract. The incubation was carried out in a thermomixer at 37 ^{O}C overnight (12 hours).

4.2.5 Sample purification before MALDI-TOF analysis

After performing chemical derivatization, the samples were concentrated and subjected to purification. Three solutions were prepared in advance: binding solution (BS; 0.1 % trifluoroacetic acid (TFA) in water), elution solution 1 (ES1; 0.1 % TFA in 50 % ACN) and elution solution 2 (ES2; 0.1 % TFA in 75 % ACN). For purification, ZipTips with 0.2 µL C18 resin (ZTC18M096, Millipore; Merck KGaA) were used. 1 µL of the sample was dissolved in 9 µL of BS, after which the pH of the solution was checked (pH < 4).

Before starting the purification of each sample, ZipTips were first wetted once with 10 µL of ES2. Then, equilibration was carried out with three washes of 10 µL of BS solution. Afterwards, the sample was loaded using pipetting, with one sample requiring 10-20 pipetting cycles. After applying the sample, ZipTip was washed twice with 10 µL of BS to remove impurities. Elution was performed using 0.6 µL of ES2.

4.2.6 MALDI-TOF analysis

For MALDI-TOF analysis standard α-Cyano-4-hydroxycinnamic acid (HCCA) matrix was used. It was prepared by mixing 100 µL of ACN, 50 µL of water and 50 µL of 10 % TFA with HCCA powder. A sample in a volume of 0.6 µL was mixed with 2.4 µL of the matrix solution. The mixture was deposited onto the position (well) of a stainless steel MALDI target and allowed to dry at room temperature.

MALDI-TOF MS analyses of peptides were performed on an Ultraflex III mass spectrometer (Bruker Daltonik s.r.o.). MS were acquired in linear positive mode (20 kV ion source voltage) with 800 laser shots. Calibration of the mass spectra was performed using the Peptide Calibration Standard II (Bruker Daltonik s.r.o.) standard peaks. Ten spectra accumulations per well were acquired, giving a total of 30 spectra accumulations per sample.

4.2.7 Sample purification before LC-MS/MS analysis

After performing chemical derivatization at the protein and peptide levels, the samples were concentrated and subjected to purification. Three solutions were prepared in advance: BS (0.1 % TFA in water), ES1 (0.1 % TFA in 50 % ACN) and ES2 (0.1 % TFA in 75 % ACN). For purification, Pierce™ C18 Spin Tips (Thermo Fisher Scientific Inc.) were used. The concentrated sample was dissolved in BS to the final volume of 50 µL, after which the pH of the solution was checked (pH < 4).

Before use, all SpinTips were wetted once with 20 µL of ES2 and then centrifuged at 1000 *g* for 1 minute. After the wash, they were equilibrated twice with 20 µL of BS and also centrifuged at 1000 *g* for 1 minute. Each SpinTip was loaded with 50 µL of the sample, which was centrifuged at 600 *g* for 1 minute and washed with 20 µL of BS.

After that, the sample was loaded and centrifuged once again. Afterwards, the SpinTips were washed twice with 20 µL of BS, and after each washing, they were centrifuged at 1000 *g* for 1 minute. The peptides were eluted using two washes of 10 µL ES1, followed by two washes of 10 µL ES2, and after each elution, they were centrifuged at 600 *g* for 1 minute. The eluted samples were transferred to an LC vial. The tube used for transfer was rinsed with 30 µL ACN, and the contents were also transferred to the same LC vial. The eluate was concentrated in a SpeedVac to a final volume of 12 µL. After that, 1 µL was taken from each sample for quality control (QC), and it was diluted 10 times in 1% formic acid (FA). To all eluted samples, 1.2 µL of 10% FA was added to achieve a final FA concentration of 1%.

4.2.8 LC-MS/MS analysis

LC-MS/MS analysis was used to analyze recombinant protein samples as well as histone extract from cell culture. The samples were submitted for MS analysis to the CL Proteomics laboratory. For analysis, an Ultimate 3000 RSLCnano (Dionex) liquid chromatograph connected online to the Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific Inc.) was used. To prepare and condition samples for introduction into the analytical column, a cartridge-based trap column (Thermo Fisher Scientific Inc.) was used. Equilibration of the pre-column and the column was performed before sample injection into the loop. Subsequently, peptides were separated on an analytical column Ion Opticks, Aurora C18 (25 cm long, 75 µm inner diameter, 1.6 µm particles, Ion Opticks Pty Ltd). The analytical column was tempered at 50 ^{O}C. Two mobile phases were used: A (0.1 % FA in Milli-Q water (Millipore)) and B (0.1 % FA in 80 % acetonitrile), with a flow rate of 0.3 µL/min. The LC gradient (Table III) for separating histone peptides was designed to account for their increased hydrophobicity resulting from modification by anhydrides and was optimized to achieve the best possible separation of isobaric peptides (Figure 8). The output from the analytical column was connected to the ion source, Digital PicoView 550 (New Objective). Mass spectra were acquired using an Orbitrap detector and DDA (Data-Dependent Acquisition) method. The resolution was set at 60,000 (at 400 m/z) with a target scan range of 350 – 2000 m/z in positive mode. Dynamic exclusion was enabled for 20 seconds after obtaining one MS spectrum. Precursors were isolated using quadrupole mass filtration with a 1.6 m/z isolation window. MS/MS spectra were acquired with a resolution of 15,000 (at 400 m/z). The maximum injection time was set at 22 ms.

Table III: The change in the mobile phase B gradient over time

Time (min)	Flow (µL/min)	Mobil phase B (%)
0	0.3	5.0
3	0.3	5.0
5	0.3	5.0
25	0.3	25.0
35	0.3	29.0
45	0.3	32.0
60	0.3	38.0
80	0.3	50.0
90	0.3	85.0
104.9	0.3	85.0

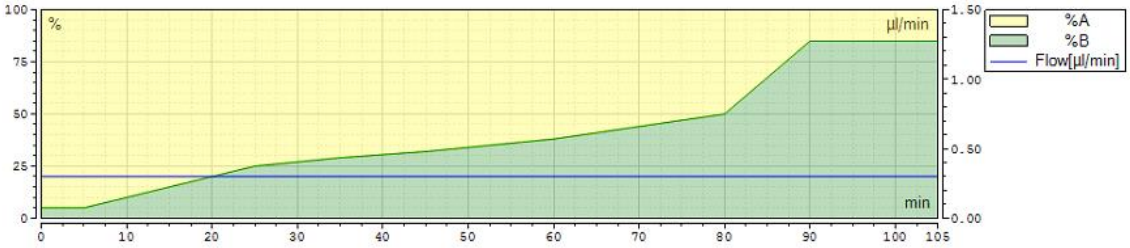


Figure 8 – The LC gradient utilized in the LC-MS/MS analysis of derivatized histones

4.2.9 Data processing

For MALDI-TOF MS data processing Flex Analysis 3.4 (Bruker Daltonik, Germany) software was used. Grubbs' test was employed to exclude extreme values, and calculations were performed for average ratios along with standard error, and 95% confidence intervals.

The acquired LC-MS/MS data were analysed using the Proteome Discoverer software (Thermo Fisher Scientific, version 2.42), which is linked to the search engine Mascot (Matrixscience, version 2.6.2). The Mascot database search engine allows for comparing the obtained MS spectra with a selected protein database. Specifically, the databases selected for the search included the in-house human histone protein database (histone human; v191011; 52 protein sequences, generated from UniProt database), the database of all human proteins (UniprotKB human; v230301; 20591 protein sequences), and a database containing sequences of commonly occurring protein contaminants (cRAP; v181122; 112 protein sequences; downloaded from http://www.thegpm.org/crap/). For all database searches, the parameters for trypsin semiArg-C cleavage with two missed cleavage sites were set. Dynamic modifications

were applied to the human histone database, including TFPA (K, S, T, Y), TFPA (N-terminus), acetylation (K), methylation (K, R), dimethylation (K), trimethylation (K), and phosphorylation (S, T). The fragment mass tolerance was set to 0.03 Da, with a precursor mass tolerance of 10 ppm. For the UniProtKB human database, dynamic modifications included TFPA (K, S, T, Y) and TFPA (N-termini). The fragment mass tolerance was set to 0.5 Da, with a precursor mass tolerance of 10 ppm. For the cRAP database search, dynamic modifications were set for oxidation (M), deamidation (N, Q), and TFPA (K, S, T, Y) as well as TFPA (N-terminus) with the fragment mass tolerance 0.5 Da and the precursor mass tolerance 10 ppm. For quantitative evaluation, the Skyline (version 22.2.1.425, Agilent Technologies) software was utilized. Manual validation of selected peptide identifications was performed, and the peptide area obtained from the extracted ion chromatogram (EIC) was quantified, including identifications across data files based on retention time and m/z.

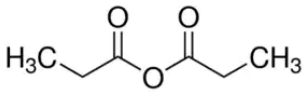
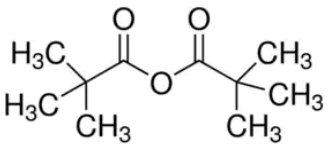
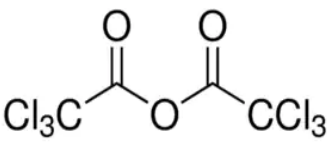
5 Results

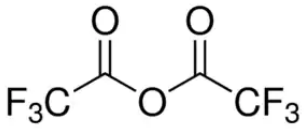
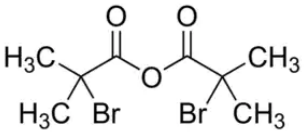
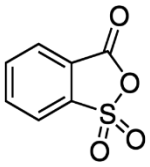
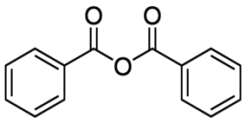
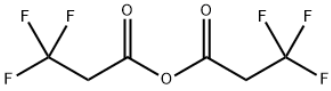
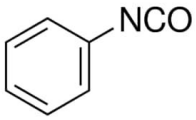
The aim of the work was to test various anhydrides that could be used as an alternative to Prop and TMA. To assess the performance of a particular derivatization agent, the efficiency of derivatization at the peptide level was investigated. Additionally, for the best candidate, an evaluation of its effectiveness was conducted on both recombinant H3.2 protein and histone extract isolated from cell culture. The evaluation considered general labelling efficiency, including assessing the quality of cleavage.

5.1 Primary selection of derivatization agents using a synthetic peptide

As the first selection step, derivatization at the peptide level was performed. A synthetic peptide ARTKme2QTARKS with the m/z of 1174,701 was used as a representative histone peptide. The aim was to achieve chemical modification of all free NH_2 groups in the peptide, specifically one NH_2 group at lysine 9 and one NH_2 group at the N-termini. Seven anhydrides were selected as potential alternative derivatization agents (Table IV). The efficiency of labelling was compared to the efficiency of derivatization using previously employed methods, such as labelling with TMA[60], Prop[67], PIC[66] and BA[64].

Table IV: List of the anhydrides, tested as derivatization agents. The mass of the anhydride, the structural formula, and the pKa of the respective carboxylic acid are indicated

Anhydride	**Mass (Da)**	**Formula**	**pKa (acid)**
Propionic (Prop)	130,14		4,86
Trimethylacetic (TMA)	186,25		5,03
Trichloroacetic (TCAA)	308,76		0,17

Trifluoroacetic (TFAA)	210,03		-0,3
Bromoisobutyric (BiBA)	315,99		2,91±0,10 (Predicted)
2-Sulfobenzoic acid cyclic anhydride (SBA)	184,17		-1,12±0,15 (Predicted)
Benzoic anhydride (BA)	226,23		4,19 (25°C)
3,3,3-Trifluoropropionic anhydride (TFPA)	238,085		3,06 (25°C)
Phenyl isocyanate (PIC)	119,12		–

All derivatization agents were prepared in a ratio of 1:11 (v/v) in ACN, as described in Chapter 4.2.2. To 10 µL of the sample with synthetic peptide, 1 µL of the mixture was added, and derivatization proceeded according to Chapter 4.2.2. The incubation took place in a thermomixer at RT with 1000 rpm for 20 minutes. The pH 8.0 was adjusted with NH_4OH individually based on the properties of the anhydride and the pKa of the acids formed as an undesirable reaction product. Each sample underwent two cycles of derivatization.

MALDI-TOF MS was performed to check the performance of all anhydrides (Figure 9). The relative representation (R_f) of individual modified forms of histone peptides from MS spectra was determined as the ratio of the peak intensity of particular form to the sum of intensities of unlabelled (I_n), mono-modified (I_m; either modified NH_2 group at the N-terminus or at lysine 9) and di-modified form (I_d; modified NH_2 groups at both the N-terminus and lysine 9) of the synthetic peptide (Equation 1).

$$R_d(\%) = \frac{I_d}{I_n + I_m + I_d} \times 100 \quad (1)$$

As the synthetic peptide contains two free NH_2 groups (at N-terminal alanine and lysine 9), the most significant factor for assessing derivatization with a particular anhydride was the presence of its di-modified form. Nevertheless, to assess the labelling efficiency, all peptide forms—unmodified, mono-modified, di-modified and their ratios — were considered (Table V).

The highest percentage of di-modified forms was obtained with previously introduced anhydrides: 97.66 %, 100 %, and 99.88 % using PIC, BA, and Prop, respectively. TMA and TFPA showed significantly lower levels of di-modified forms, 20.34 % and 12.11 %, respectively, but overall had a high percentage of modified forms relative to unmodified ones. Among the tested anhydrides, TCAA, TFAA, BiBA, and SBA were the derivatization agents for which trace amounts (BiBA, 0.39 %) or no di-modified peptide forms were observed. TFAA and TCAA were excluded from further analyses, as in their case, the conversion did not occur at all. Then BiBA was tested at two concentrations: 1 μg/μL and 0.1 μg/μL. However, changing the concentration did not improve the conversion rate and BiBA was also excluded from further analyses. Therefore, among this group of candidates, the highest percentage of labelled peptide forms was observed after two derivatization cycles with SBA, 36.67 %. Based on the results of the first experiment, two of the most promising candidates, previously not tested as potential derivatization agents, were selected for further optimization of the protocol: TFPA and SBA (Figure 10).

Table V: List of the used anhydrides with the mass of the respective adduct and masses of corresponding unlabeled, mono- and di-modified forms of the synthetic peptide

Anhydride	Adduct mass, Da	Unlabeled form, Da	Mono-modified form, Da	Di-modified form, Da
Prop	56,026	1174,701	1230,727	1286,753
TMA	84,058		1258,759	1342,817
TCAA	143,893		1318,594	1462,487
TFAA	95,982		1270,683	1366,665
BiBA	147,952		1322,653	1470,605
SBA	182,975		1357,676	1540,651
BA	104,026		1278,727	1382,753
TFPA	109,998		1284,699	1394,697
PIC	119,037		1293,738	1412,775

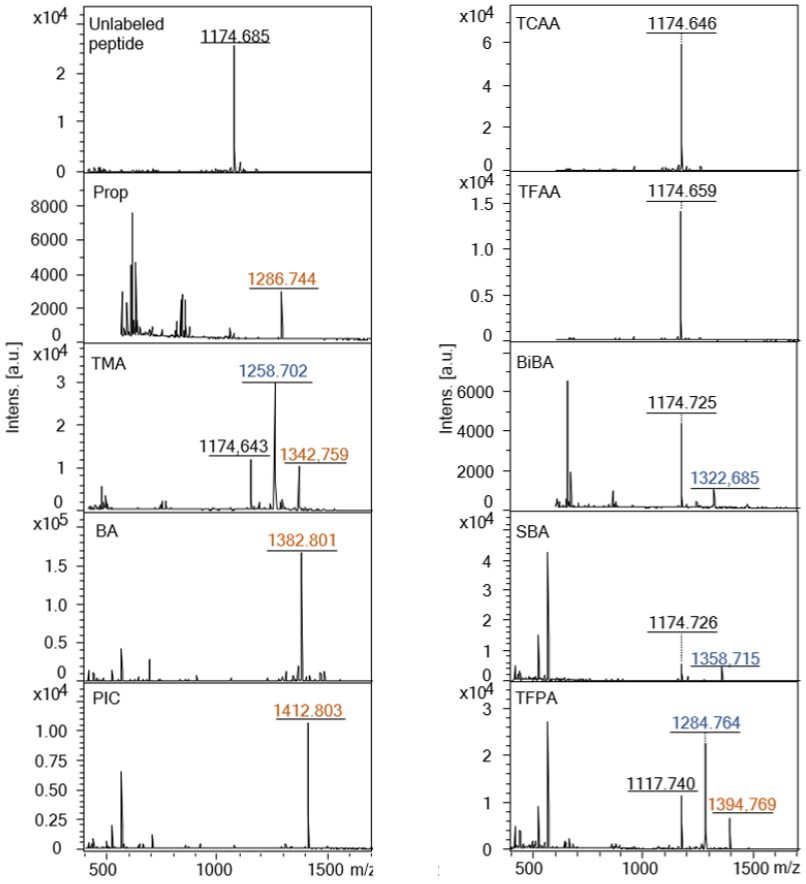


Figure 9 – Mass spectra of modified peptide obtained using MALDI-TOF MS. The mass of the peak corresponding to the unmodified form of the peptide is indicated in grey, the mono-modified form in blue, and the di-modified form in orange

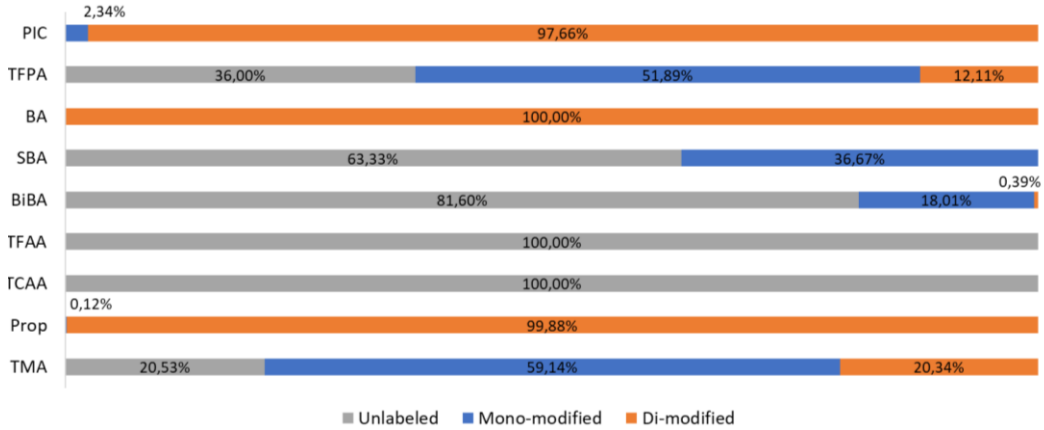


Figure 10 – The efficiency of derivatization of the synthetic peptide by various anhydrides. The percentage of unlabelled, mono-modified, and di-modified peptides was calculated as the average of triplicates

5.2 Comparison of the best candidates' performances

The performance of TFPA and SBA was compared, considering the influence of the concentrations of the used anhydrides. The concentration of TFPA used in the experiment corresponds to the commercial stock solution (90% purity, Chapter 4.1). Since SBA is commercially available as a powder, two stock solutions in ACN were prepared for its testing, corresponding to concentrations of 100 µg/mL and 500 µg/mL. Samples labelled with anhydrides according to the scheme of 1:11 (v/v) in ACN with the addition of 1 µL of the mixture to the sample, were used as a standard, as the same sample preparation conditions were used in the previous experiment. As an alternative, a 1:3 (v/v) ratio of anhydride to ACN was used, and 3 µL of the derivatization agent was added to the sample (the derivatization agent was prepared as described in Chapter 4.2.2). This resulted in a significantly higher concentration of anhydride during labelling (Table VI). In this experiment, samples underwent three cycles of derivatization. Incubation took place in a thermomixer, at room temperature, and 1000 rpm for 20 minutes.

Table VI – Key parameters of the experiment to determine the efficiency of derivatization of synthetic peptide using SBA and TFPA.

Anhydride	Concentration	Ratio (v/v)	Volume, µL
SBA	100 µg/µL	1:11	1
	500 µg/µL	1:11	1
	100 µg/µL	1:3	3
TFPA	Commercial solution	1:11	1
	Commercial solution	1:3	3

In the case of SBA, changes in substance concentration or reaction conditions did not lead to the formation of di-modified peptide forms (Figure 11). The use of a higher SBA:ACN ratio led to the overall lower number of labelled peptide forms. Increasing the concentration of the stock solution from 100 µg/µL to 500 µg/µL for 1:11 (v/v) SBA:ACN ratio decreases the percentage of modified peptide forms from 67.60 % to 39.10 % after two derivatization cycles. The introduction of the third derivatization cycle also decreased the overall content of SBA-labelled peptide forms compared to the previously applied two cycles. Thus, none of the tested conditions improved the efficiency of labelling using this anhydride.

For both TFPA:ACN ratios used, the di-modified form of the synthetic peptide was already detected in the mixture after the first derivatization cycle (2,00 % for 1:11 TFPA:ACN and 11.90 % for 1:3 TFPA:ACN ratio). However, the addition of a third derivatization cycle significantly increased the content of labelled forms. Changing the TFPA:ACN ratio from 1:11 to 1:3 (v/v) significantly increased the overall quantity of

modified forms up to 99.50 % but did not alter the final content of forms with both labelled NH_2 groups (64.80 % for 1:11 TFPA:ACN and 62.30 % for 1:3 TFPA:CAN ratio).

As the presence of di-modified forms is crucial, TFPA in ACN at ratios of 1:11 and 1:3 (v/v) were selected for further optimization. The labelling efficiency of SBA proved to be insufficient and less reproducible; thus, SBA was excluded from further experiments (Figure 11).

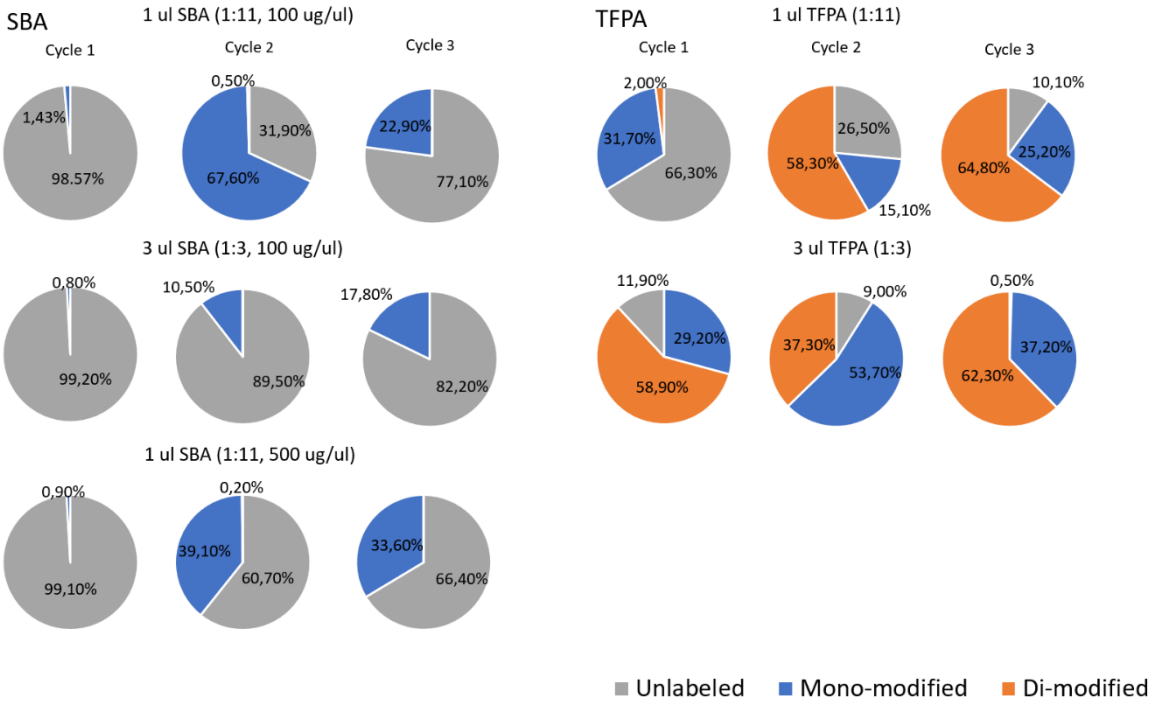


Figure 11 – The ratio of unlabeled, mono- and di-modified forms of the synthetic peptide after three cycles of derivatization using different amounts of SBA and TFPA

5.3 Optimization of TFPA derivatization protocol

5.3.1 Incubation temperature and time impact on labelling efficiency

The next experiment was conducted to determine the influence of the temperature and incubation time on the derivatization process. Samples incubated in a thermomixer at room temperature and 1000 rpm for 20 minutes after one cycle of derivatization with both 1:11 and 1:3 TFPA:ACN ratios were prepared according to the Chapter 4.2.2. The conditions of the experiment are described in Table VII.

Table VII: The incubation conditions that were tested in the experiment

TFPA:ACN ratio (v/v)	Incubation type	Temperature	Frequency, rpm	Time, h
1:11	Thermomixer	RT	1000	1
				4
				16
		40°C		1
				4
				16
1:3		RT		1
				4
				16
		40°C		1
				4
				16

The obtained results led to the conclusion that increasing the temperature does not affect the efficiency of derivatization. Furthermore, an increase in incubation time does not lead to an increase in the quantity of mono- and di-modified forms of the peptide too. The 1:3 (V/V) ratio of TFPA to ACN was more effective and led to the modification of both NH_2 groups after the first derivatization cycle (Figure 12).

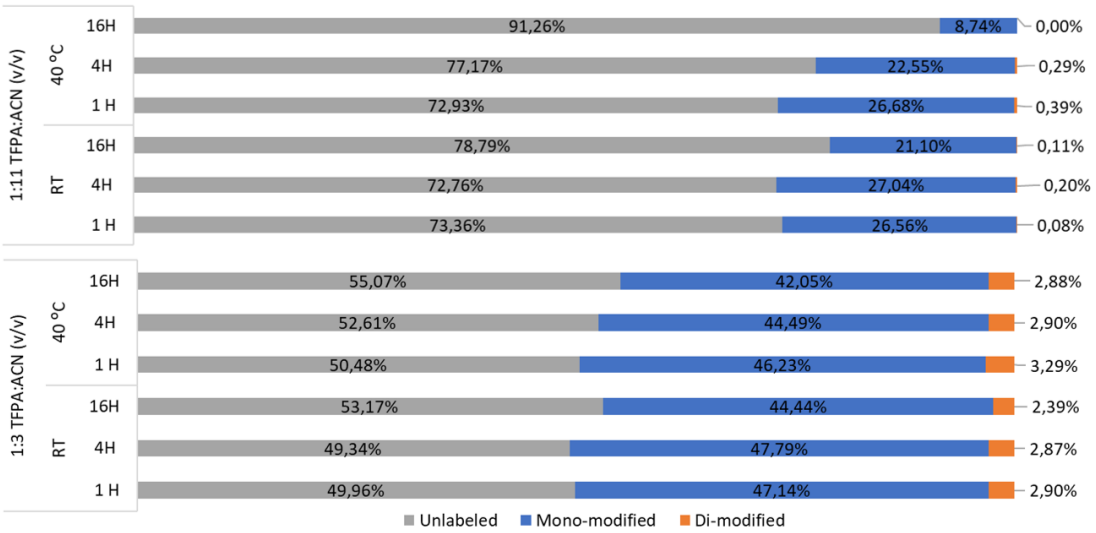


Figure 12 – The percentage of unmodified, mono- and di-modified forms of the peptide after incubation with a TFPA:ACN ratio of 1:11 (v/v) and with a TFPA:ACN ratio of 1:3 (v/v) after one derivatization cycle

Overall, increasing the incubation time neither caused changes in labelling efficiency nor led to an increase in the content of labelled forms with an increase in the incubation period. Compared to the control tested in previous experiments (RT, 1000 rpm, 20 min), the overall amount of modified peptide forms significantly decreased in both TFPA 1:11 (v/v) and TFPA 1:3 (v/v). Di-modified peptide forms were present in trace amounts in the case of TFPA 1:11 and in the range of 2.39-3.29 % in the case of TFPA 1:3 (v/v), which is significantly lower than the ratio of di-modified forms to the total peptide amount after 20 minutes of incubation (Figure 13). Also, since in all previous experiments, the use of TFPA in ACN prepared at a ratio of 1:3 (v/v) resulted in the formation of a larger number of modified forms, this ratio was chosen as the most effective for further work.

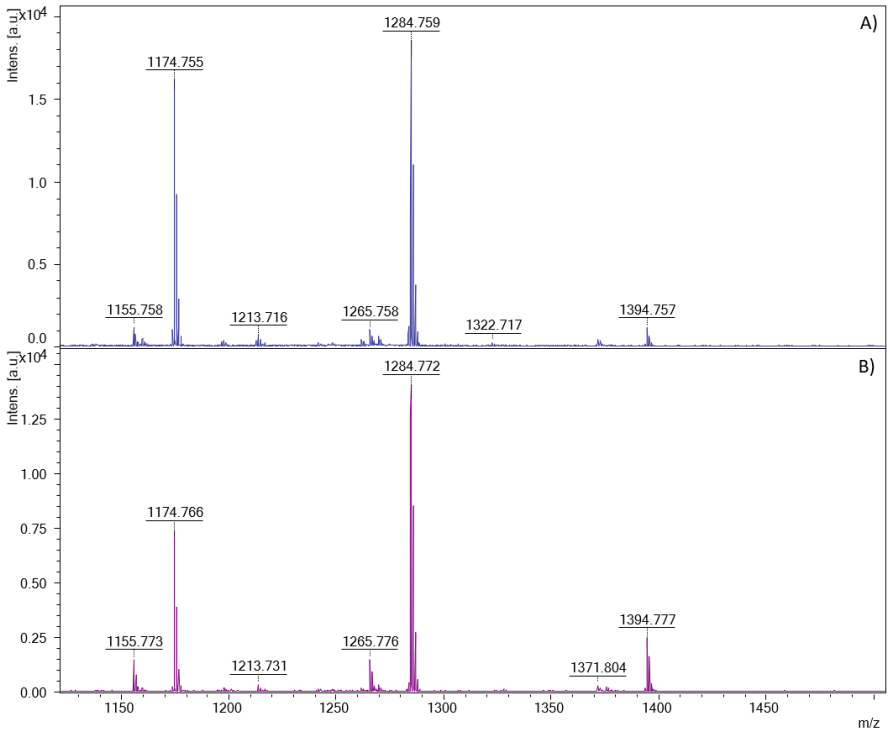


Figure 13 – Mass spectrum of the synthetic peptide obtained after one derivatization cycle with TFPA in ACN at a ratio of 1:3 after incubation using a thermomixer at room temperature and 1000 rpm for A) 1 hour, B) 20 min. Peaks correspond to the unmodified form of the peptide (m/z 1174.8), mono-modified form (m/z 1284.8), and di-modified form (m/z 1394.8)

5.3.2 Impact of the number of derivatization cycles on labelling efficiency

As the increase in incubation time did not affect the efficiency of derivatization and did not significantly increase the quantity of mono- and di-modified forms of the peptide, further incubation was conducted for 20 minutes at 1000 rpm and room temperature. To monitor the reaction dynamics, an experiment was conducted in which the synthetic peptide underwent three cycles of derivatization using TFPA at a ratio of 1:3 (v/v) in ACN according to the Chapter 4.2.2. Samples for analysis were applied to the MALDI target in 1 minute after the first addition of the derivatizing agent and after 10 minutes. Samples were also collected after the second and third cycles of derivatization (20 min, 40 min). The last sample was applied at the end of the third incubation cycle, 60 minutes after the start of the experiment.

Immediately after 1 minute of derivatization reaction, the conversion occurred, showing a substantial increase in the content of mono- and di-modified forms (51-53 % and 7-9 %, respectively; Figure 14). The total amount of unlabelled peptide during the first derivatization cycle incubation remained almost unchanged and was approximately 39-49 %. Overall, between 1 minute into the experiment and up to 20 minutes, the signal remained almost stable.

The conversion was supported by the subsequent derivatization cycles. The second addition of TFPA (40 minutes) led to an increase in the di-modified form to 24-26 %, and the amount of unlabelled forms dropped to 23-25 %. The overall content of mono-modified forms did not change significantly. By the end of the third cycle (60 minutes), the number of unlabelled forms dropped to 4-11 %, while the di-modified forms increased to 34-47 %. The number of mono-modified forms also remained unchanged, settling at 48-51 %.

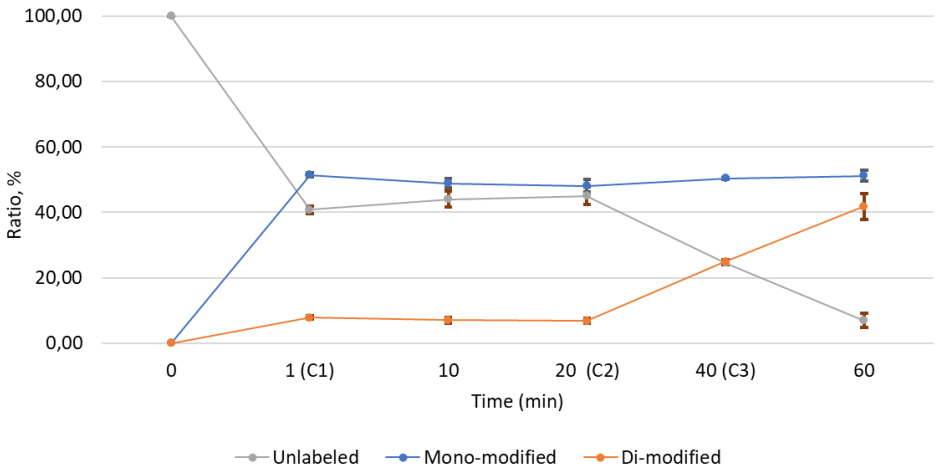


Figure 14 – The dynamics of the change in the ratio (%) of unmodified, mono- and di-modified forms of the synthetic peptide over time after the first (C1), second (C2), and third (C3) cycles of derivatization. Standard error is shown as error bars

5.3.3 Impact of microwave-assisted Incubation on labelling efficiency

In the next experiment, the standard incubation in a thermomixer was compared with microwave-assisted derivatization, which has been effectively applied in the TMA derivatization protocol.

The synthetic peptide sample underwent three cycles of derivatization using TFPA in ACN at a ratio of 1:3 (v/v) according to the Chapter 4.2.2. Incubation in the thermomixer was carried out at room temperature, 1000 rpm for 20 minutes. Microwave incubation was performed according to Chapter 4.2.3. The results obtained showed that the use of microwaves after three cycles of derivatization, in general, led to lower content of di-modified peptide forms (Figure 15).

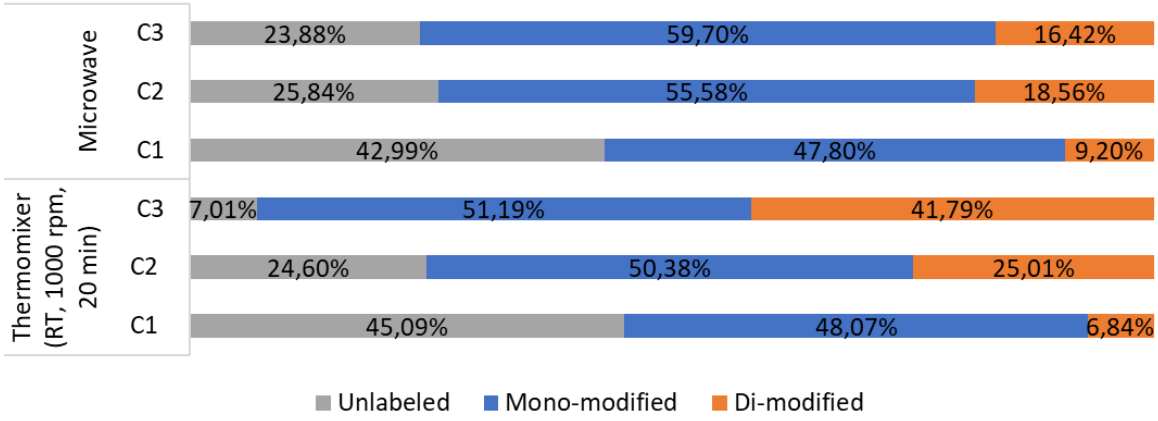


Figure 15 –The ratio of unmodified, mono- and di-modified forms of the peptide after incubation using a microwave oven and incubation in a thermomixer (RT, 1000 rpm, 20 min)

5.3.4 Desalting of TFPA-labeled peptide

For obtaining more accurate and precise results, as well as for removing excess reagents, such as reaction by-products from the mixture, desalting was performed after the third cycle of derivatization before spectrum acquisition. Although this step substantially reduced the variability in peak intensities, making the data more homogeneous, it also altered the ratio of modified forms of the peptide (Figure 16 A, B).

Purification was accompanied by a partial loss of di-modified forms of the peptide by 19.3% in the standard incubation method and without any significant loss in the case of microwave irradiation. Additionally, desalting resulted in an average increase of 11.12 % in mono-modified forms. Since desalting leads to the removal of MS- interfering contaminants, this step was included in the sample preparation protocol for MS and MS/MS analysis.

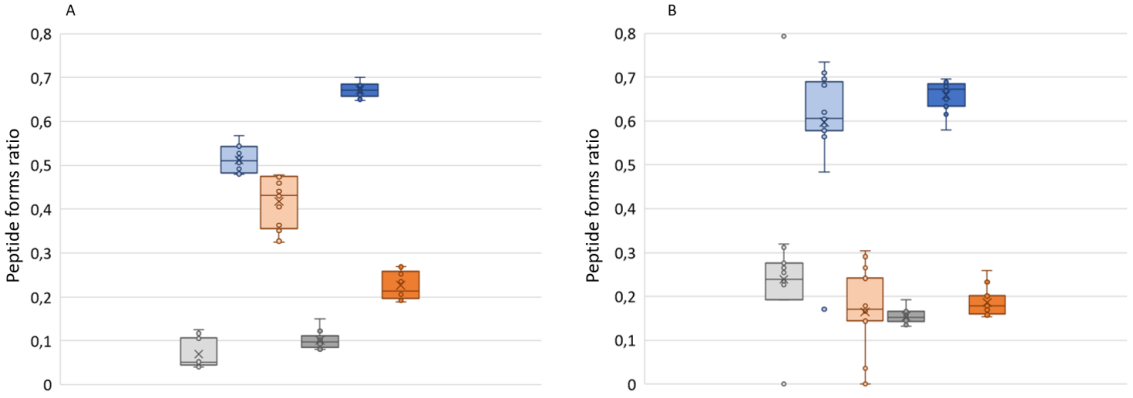


Figure 16 – The influence of the purification step on the ratio of unmodified, mono-, and di-modified forms of the synthetic peptide after three cycles of derivatization with TFPA and incubation in A) thermomixing at room temperature, 1000 rpm for 20 minutes and B) microwave-assisted incubation. The quantity of unmodified, mono-modified, and di-modified forms before desalting is indicated by light grey, light blue, and light orange, respectively. The quantity of unmodified, mono-, and di-modified forms after purification is represented by grey, blue, and orange colours, respectively. The box plot for each peptide form was based on fifteen values obtained from triplicates

5.3.5 Efficiency of TFPA-labelling on recombinant histone

After the main conditions for derivatization were determined, this protocol was applied to the recombinant protein H3.2. All samples underwent three cycles of derivatization with TFPA in ACN at a ratio of 1:3 (v/v) according to Chapter 4.2.2, including incubation in a thermomixer at room temperature, 1000 rpm, for 20 minutes. After the third cycle of derivatization, the samples were desalted and transferred to vials for LC-MS/MS analysis. The obtained spectra were processed using the Skyline software, which allows for manual editing of selected peaks and quantitative analysis of the data.

In the analysis of modified recombinant proteins, peptides were taken into account based on the following categories: desired - peptides correctly modified by TFPA on all lysines NH_2 groups, including those correctly digested by trypsin on arginines C-ends; AS (acceptable sequence) – acceptable, i.e. shorter or longer sequence peptides containing the same number of lysines as the corresponding desired peptides; WS (wrong sequence) - non-specifically digested peptides, i.e. peptides containing a different number of lysines in the sequence than the desired peptides; over - peptides cleaved at the C-terminus of arginine containing non-specific reactions with OH-groups on amino acids serine, threonine or tyrosine; under - peptides cleaved at the C-terminus of arginine with incomplete labelling of NH_2 groups (i.e. at least one NH2-group did not undergo TFPA labelling); (Figure 19).

The main indicator of labelling efficiency was the quantity of desired and acceptable peptides. Among them, those correctly digested by trypsin constitute an average of 17 %, and those differing in length an average of 10 %. According to the obtained data, labelling the protein with TFPA did not result in side reactions at hydroxyl groups of serine, threonine, or tyrosine. Among all peptides, approximately half are those in which the part of NH_2 groups of lysines remained unlabeled. However, all desired, acceptable, under- and over-modified peptides containing the required number of lysines formed a counting category. About 24% of all peptides were nonspecifically digested and were considered non-countable. The ratio of the number of counting to non-counting peptides directly reflected the ratio of the peak areas of the corresponding peaks and the efficiency of labelling. After labelling the recombinant protein with TFPA, the total amount of counting peptides was approximately 76%, with the majority of them being under-modified.

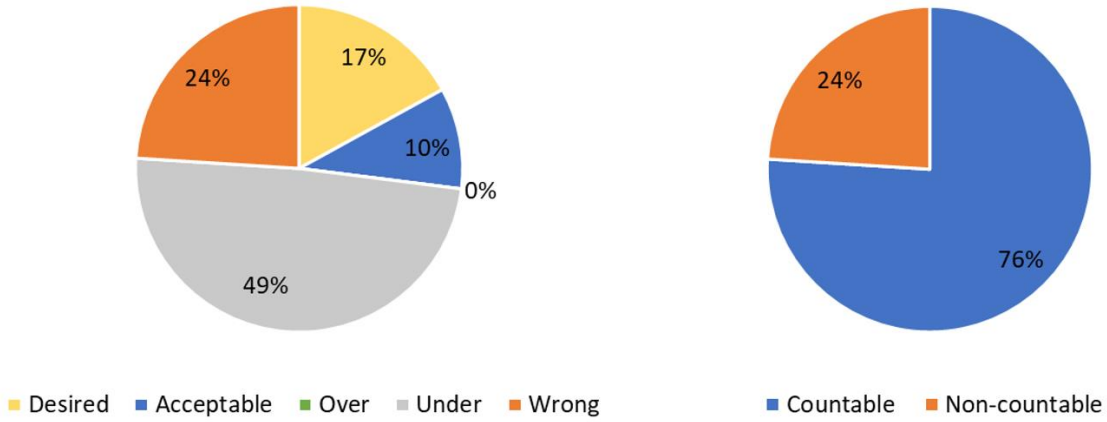


Figure 17 – The ratio of all peptides of recombinant histone H3.2 obtained after analysis using LC-MS/MS

5.3.6 Performance of TFPA-labeling on MEC-1 histone extract

Another important factor indicating the performance of labelling is the ability to increase peptides' hydrophobicity, thereby retaining and detecting short hydrophilic peptides. For the comparison of labelling performance using TFPA and Prop, the histone extract from MEC-1 B-CLL cells was used. All samples were derivatized using the same protocol as applied in Chapter 5.3.5, and data were measured using LC-MS/MS. Obtained spectra were processed using Proteome Discoverer 2.2, and retention time was determined from manually edited precursor ion peaks in Skyline. To assess TFPA and Prop performance, naturally unmodified peptide forms of histone H3.3 with varying lysine content in their sequences were selected, along with a peptide

corresponding to the initial sequence of histone H3.3 (Table VIII). All analysed peptide forms contained one modified NH_2 group at the peptide N-terminus. In cases where lysine was the N-terminal amino acid, modification could occur on both the N-terminal and the free ε-NH_2 group of the side chain. Peptides with multiple lysines in their sequences could be fully or partially modified. These peptides will have identical sequences but will differ in mass and retention times due to varying amounts of TFPA attached.

Labelling with TFPA substantially increased the RT of the majority of histone peptides. The difference in retention time between TFPA- and Prop-labelled peptides is showing a shift from 3 to 25 minutes. Moreover, the peak corresponding to the short hydrophilic N-terminal peptide of histone H3.3, A1RTKQTAR8, was absent in the spectrum obtained after Prop labelling.

Table VIII – Retention time (RT) comparison of selected TFPA- and Prop-labelled peptides of histones H3. The number of modified lysine residues is indicated by the number of labelled NH_2 groups, including those at the N-terminus and on lysine residues (N-term + Lys)

# (N-term + Lys)	Sequence	Peptide	TFPA		Prop		ΔRT [min]
			m/z	RT [min]	m/z	RT [min]	
1+1	**A**RT**K**QTAR	H3.3_K4	576.27	***23***	-	**-**	-
1+2	**K**STGG**K**APR	H3.3_K9K14	616.26	***35***	535.30	***25***	10
1+1	**K**STGG**K**APR	H3_K9K14	561.26	***21***	507.29	***18***	3
1+2	**K**QLAT**K**AAR	H3_K18K23	658.81	***59***	577.85	***36***	25
1+1	**K**QLAT**K**AAR	H3_K18K23	603.81	***32***	549.84	***25***	7
1+3	**K**SAPATGGV**K** **K**PHR	H3_K27K36 K37	937.42	***42***	829.47	***29***	13
1+2	**K**SAPATGGV**K** **K**PHR	H3_K27K36 K37	882.42	***29***	801,46	***24***	5

- number of labelled NH_2 groups

6 Discussion

Histones, their variants, and PTMs are directly related to chromatin remodelling, and therefore, they play a crucial role in many biological processes. Currently, MS is the predominant approach for characterising this group of proteins. [68]. However, the extensive variety of histone PTMs, the variability in the patterns of their modifications, and the subtle differences in the sequences of histone variants still pose challenges for analysis.

Investigations into histone PTMs have incorporated significant proteomics methodologies, including top-down, middle-down, and bottom-up approaches, along with various MS acquisition techniques [45, 69, 70]. In DDA, a subset of the most abundant ions reaching the mass spectrometer detector during an MS1 scan are individually isolated and fragmented in sequential MS2 scans. Alternative LC-MS/MS targeted acquisition strategies have emerged, including multiple-reaction monitoring (MRM) [71], parallel reaction monitoring (PRM)[72], and data-independent acquisition (DIA)[73]. DIA provides a synthesis of the advantages associated with MRM/PRM and DDA analysis modes, accomplishing this by repetitively selecting mixtures of peptide species within predetermined mass ranges for subsequent MS2 scans. The method has been shown to be effective in examining post-translationally modified histone peptides, including those with isobaric co-eluting characteristics. However, it relies on a limited set of specific fragment ions when analysing histone positional isomers [74, 75].

The involvement of histone epigenetic marks in various physiological and pathological processes has been increasingly acknowledged. This has led scientists to develop MS strategies for PTMs characterising [76, 77, 78].

While MS-based proteomics has become increasingly important in studying protein modifications, the investigation of histone PTMs and variants remains a challenge due to several factors. Distinguishing specific histone variants can be challenging due to small differences in histone sequences, represented by only a few amino acids. Differentiating diverse proteoforms is complicated by the complexity of PTM combinations. The abundance of basic amino acids can result in the formation of short peptides after tryptic digestion, which can impede RP-HPLC separation and subsequent detection by MS. Additionally, the presence of multiple lysines in histone sequences often leads to the co-elution of more than two isobaric peptide forms [79]. This underscores the need to develop and optimize not only MS analysis and data evaluation but also sample preparation.

The most commonly used method for protein identification and quantification is the bottom-up MS approach. This method offers enhanced sensitivity and mass accuracy while requiring fewer complex data analyses compared to middle- and top-down proteomics. Histones, which are rich in lysine and arginine residues, pose a challenge for the bottom-up approach (Chapter 2.4.3). To address this issue, chemical derivatization

of NH_2 groups was introduced. However, among actual derivatization agents, there is no single protocol that would be optimal for the analysis of all histone variants and their PTMs.

The goal of this work was to select an alternative derivatization agent and optimize the protocol to address the current shortcomings in histones labelling and characterization. To analyse the efficiency of used anhydrides, they were compared with Prop, which is currently the most commonly used derivatization agent. However, Prop has several drawbacks. These include incomplete and nonspecific derivatization, the inability to study natural propionylation, insufficient retention time increase, *etc* [64]. Many research groups have attempted to modify the Prop derivatization protocol to address different issues: overpropionylation, using boiling for 1 hour or adding hydroxylamine [80], hybrid methods for detecting specific modifications [66] or minimizing unwanted side reactions by changing the solvent [81], among other strategies. Each of them has its pros and cons, but none of them completely addresses the drawbacks of Prop labelling.

In the experiment conducted by Sidoli *et al.* [64], twelve commercially available anhydrides were tested. Synthetic unmodified histone H3 N-terminus, synthetically modified histone peptides, and histone extracts from cell lysates were used as samples. After analysing the quality of the obtained MS/MS spectra, as well as filtering candidates based on their prevalence as natural modifications, retention time shift indicators and reaction yields, valeric and benzoic anhydrides remained the most successful candidates for comparison with the classical derivatization using Prop. Despite benzoic anhydride demonstrating promise as a top candidate for histone derivatization, the analysis of the unmodified histone H3 tail revealed a significant presence of side products and incomplete reactions. Furthermore, due to benzoic anhydride being in a solid state naturally, it results in the formation of salt residue after the solution dries, prompting the need to minimize the number of derivatization steps. In general, while benzoic anhydride demonstrated extremely high labelling efficiency at the peptide level and had great potential as an alternative candidate, the chromatogram of the sample treated with it exhibited notably greater complexity compared to other anhydrides, such as propionic. After employing the Prop protocol, 57 unique combinations of PTMs were successfully identified and quantified using a standard LC-MS/MS bottom-up proteomics approach. In comparison, utilizing valeric and benzoic derivatization led to the identification of only 43 and 48 combinations of PTMs, respectively [64]. Thus, the most common approach to chemical derivatization of histones currently relies on the reaction of histones with Prop at both the peptide and protein levels.

Maile *et al.* [66] proposed a hybrid technique (using PIC and Prop for labelling at different levels, Chapter 2.5.2) to address the detection and quantification issues of H3K4me2/me3 transcriptional activators. The use of hybrid chemical derivatization allows the detection and quantitative determination of short peptide with K4 in all methylation states and results in a fourfold increase in the number of detected histone

peptides for the same input compared to the standard Prop derivatization [66]. The labelling efficiency of PIC at the peptide level was also assessed in our study and is among the best of all evaluated candidates. Moreover, the hybrid labelling method allowed for the detection of low-abundant acetylated lysines, such as K27ac and K36ac, which were barely or not detected at all using the standard Prop protocol. Nevertheless, the application of the hybrid method still leads to the formation of products from nonspecific reactions with serine and threonine. PIC faces challenges in labelling at the protein level, hindering its potential use as the sole derivatization agent.

Recently, the protocol using TMA as an alternative derivatization agent was published by Kuchaříková *et al* [60]. This protocol employed a microwave-assisted incubation method, significantly reducing sample preparation time without compromising the labelling efficiency. It was found that, overall, the labelling efficiency with TMA is inferior to Prop, and the amount of nonspecific reaction products is less than 0,4 %. In addition, peptides labelled with TMA exhibited greater hydrophobicity than peptides modified with Prop, increasing the retention time from 17 to 42 minutes, depending on the number of lysines and TMA groups in the peptide. Such an increase in retention time leads to a greater number of identifiable PTMs. Importantly, the chromatographic separation improved by TMA enabled the separation and quantification of isobaric peptides at the MS1 level. However, the identification of the H3A1-R8 peptide, especially its modified forms, still poses challenges [60].

It's also worth noting that, in addition to the choice of the derivatization agent, various derivatization protocols often involve different reaction conditions. For example, the solvent, methods, and duration of incubation, as well as the number of derivatization cycles, can vary. The detailed dependence of Prop and TMA labelling performance on different conditions is described in the master's thesis of Pavlína Dobrovolná [82].

In this work, the performance of five anhydrides was evaluated. The initial derivatization protocol follows the method published by Sidoli *et al* [67] for derivatization using Prop in ACN, except for the initial anhydride content. For the preliminary analysis of labelling efficiency, a different ratio, 1:11 in ACN instead of 1:3 (v/v), was chosen. A lower anhydride content was selected because the respective acids released during the derivatization reaction have a very low pKa, which could significantly affect the sample's pH. For Prop, it has been observed that underpropionylation occurs at lower pH levels. However, propionylation at pH greater than 10 not only increases labelling efficiency but also leads to an increase in overpropionylated forms of peptides, *i.e.*, nonspecific reactions. Two derivatization cycles, that were considered the most effective according to the original protocol, were maintained.

The labelling efficiency was assessed by determining the ratio of the total amount of labelled and unlabelled peptide forms, taking into account the content of di-modified forms. The obtained ratio was then compared with the results of anhydrides used in previous studies [60, 64]. TFPA was chosen as the most promising alternative

derivatization agent, surpassing other candidates in overall labelling efficiency, as well as the quantity of peptide forms with labelled N-terminal NH_2 groups and ε-NH_2 group of lysine. In addition, the adduct masses associated with natural methylation and acylation (such as mono-, di- and tri-methylation, acetylation, propionylation, butyrylation, *etc*.) are multiples of fourteen. This feature can lead to ambiguous results when searching databases for incomplete fragment ion series of peptides containing multiple lysine residues. The introduction of multiples of fourteen by chemical derivatisation using Prop and TMA can introduce bias into MS/MS data. In contrast, the mass of the TFPA adduct varies by multiples of fourteen, making it easier to identify and localise natural modifications.

To enhance the labelling efficiency, the reaction conditions were modified while optimizing the derivatization protocol using TFPA. Several conditions were tested, including the dependence on the concentration of the anhydride in the derivatization mixture, and the incubation method, including incubation time and temperature. Furthermore, the number of derivatization cycles was explored. Among all the tested conditions, a significant change in the overall content of modified forms was observed when the ratio of the added amount of anhydride was increased and an extra derivatization cycle was added. Including a purifying step following the third derivatisation gave more homogeneous data. However, it also decreased the number of peptide forms labelled with TFPA on both the N-terminal NH_2 group and the ε-NH_2 group of lysine. This suggests that the added groups made some peptides with both modified NH_2 groups hydrophobic enough not to be quantitatively eluted. Therefore, it can be assumed that RP-HPLC separation prior to MS may have similar effects. The pre-LC purification is a crucial step that must be included in the protocol to obtain precise, accurate, and reproducible data. Therefore, the subsequent experiment should focus on resolving this issue. For example, testing different types of sorbents and mobile phases for both desalting and LC-MS/MS.

Thus far, among all the tested anhydrides, TFPA appears to be the best potential alternative derivatization agent. Its optimal protocol does not require any unique techniques or long incubation steps, significantly saving time in sample preparation for LC-MS/MS analysis. However, the labelling efficiency with TFPA cannot be considered comparable to that achieved with Prop and TMA. The low labelling efficiency, manifested by the complete or partial absence of labelled free amino groups of lysines at the protein level, leads to a large number of incorrectly cleaved trypsin peptides that cannot be counted in the analysis. Despite these challenges, chemical derivatization with TFPA has its advantages. Labelling leads to a significant increase in the hydrophobicity of peptides compared to Prop derivatives, and it facilitates the identification and localization of natural acylation by avoiding mass adduct of multiples of fourteen. Additionally, the use of TFPA allows the detection of peptides on MS/MS spectra that are lost when using other derivatization agents. Therefore, the derivatization protocol using TFPA requires further adjustment.

7 Conclusion

- In this diploma thesis the performance of five organic anhydrides was analyzed as potential derivatization agents for histone preparation prior to LC-MS/MS.

- TFPA was selected as the most promising candidate, and its protocol was optimized to improve labelling efficiency.

- Significantly lower efficiency of labelling was obtained using TFPA compared to TMA and Prop. However, non-specific reactions were completely absent, and the hydrophobicity of peptides was significantly higher compared to Prop, enabling the detection of peptides that are often lost.

- Overall, TFPA protocol in its current form cannot be used as an alternative to labelling with TMA or Prop. However, further improvement in the overall labelling efficiency could significantly increase its potential.

Bibliography

1. Dawson MA, Kouzarides T. Cancer Epigenetics: From Mechanism to Therapy. *Cell*. 2012;150(1):12-27. doi:10.1016/j.cell.2012.06.013

2. Zhang L, Lu Q, Chang C. Epigenetics in Health and Disease. *Adv Exp Med Biol*. 2020;1253:3-55. doi:10.1007/978-981-15-3449-2_1

3. Liang G, Weisenberger DJ. DNA methylation aberrancies as a guide for surveillance and treatment of human cancers. *Epigenetics*. 2017;12(6):416-432. doi:10.1080/15592294.2017.1311434

4. Basheer F, Giotopoulos G, Meduri E, et al. Contrasting requirements during disease evolution identify EZH2 as a therapeutic target in AML. *J Exp Med*. 2019;216(4):966-981. doi:10.1084/jem.20181276

5. Coit P, Jeffries M, Altorok N, et al. Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naïve CD4+ T cells from lupus patients. *J Autoimmun*. 2013;43:78-84. doi:10.1016/j.jaut.2013.04.003

6. Klein K, Gay S. Epigenetics in rheumatoid arthritis. *Curr Opin Rheumatol*. 2015;27(1):76. doi:10.1097/BOR.0000000000000128

7. Ntontsi P, Photiades A, Zervas E, Xanthou G, Samitas K. Genetics and Epigenetics in Asthma. *Int J Mol Sci*. 2021;22(5):2412. doi:10.3390/ijms22052412

8. Hikichi M, Mizumura K, Maruoka S, Gon Y. Pathogenesis of chronic obstructive pulmonary disease (COPD) induced by cigarette smoke. *J Thorac Dis*. 2019;11(Suppl 17):S2129-S2140. doi:10.21037/jtd.2019.10.43

9. Schon SB, Luense LJ, Wang X, et al. Histone modification signatures in human sperm distinguish clinical abnormalities. *J Assist Reprod Genet*. 2019;36(2):267-275. doi:10.1007/s10815-018-1354-7

10. Bertogliat MJ, Morris-Blanco KC, Vemuganti R. Epigenetic mechanisms of neurodegenerative diseases and acute brain injury. *Neurochem Int*. 2020;133:104642. doi:10.1016/j.neuint.2019.104642

11. Cacabelos R, Torrellas C. Epigenetics of Aging and Alzheimer's Disease: Implications for Pharmacogenomics and Drug Response. *Int J Mol Sci*. 2015;16(12):30483-30543. doi:10.3390/ijms161226236

12. Landgrave-Gómez J, Mercado-Gómez O, Guevara-Guzmán R. Epigenetic mechanisms in neurological and neurodegenerative diseases. *Front Cell Neurosci*. 2015;9. Accessed October 12, 2023. https://www.frontiersin.org/articles/10.3389/fncel.2015.00058

13. Chang C. *Advances in Experimental Medicine and Biology*. Accessed October 12, 2023. https://www.springer.com/series/5584

14. Mottamal M, Zheng S, Huang TL, Wang G. Histone Deacetylase Inhibitors in Clinical Studies as Templates for New Anticancer Agents. *Molecules*. 2015;20(3):3898-3941. doi:10.3390/molecules20033898

15. Ohno M, Priest DG, Taniguchi Y. Nucleosome-level 3D organization of the genome. *Biochem Soc Trans*. 2018;46(3):491-501. doi:10.1042/BST20170388

16. White CL, Suto RK, Luger K. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J*. 2001;20(18):5207-5218. doi:10.1093/emboj/20.18.5207

17. Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A. Structure of the nucleosome core particle at 7 Å resolution. *Nature*. 1984;311(5986):532-537. doi:10.1038/311532a0

18. Lawrence M, Daujat S, Schneider R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet TIG*. 2016;32(1):42-56. doi:10.1016/j.tig.2015.10.007

19. Gangloff YG, Romier C, Thuault S, Werten S, Davidson I. The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem Sci*. 2001;26(4):250-257. doi:10.1016/s0968-0004(00)01741-2

20. Kumar A, Yadav G. Shared ancestry of core-histone subunits and non-histone plant proteins containing the Histone Fold Motif (HFM). *J Bioinform Comput Biol*. 2021;19(2):2140001. doi:10.1142/S0219720021400011

21. Bank RPD. RCSB PDB. Accessed October 18, 2023. https://www.rcsb.org/search?request=%7B%22query%22%3A%7B%22type%22%3A%22group%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22group%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22group%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22terminal%22%2C%22service%22%3A%22text%22%2C%22parameters%22%3A%7B%22attribute%22%3A%22struct.title%22%2C%22operator%22%3A%22contains_phrase%22%2C%22value%22%3A%22X-

Ray%20Structure%20of%20the%20Nucleosome%20Core%20Particle%2C%20NCP147%2C%20at%201.9%20A%20Resolution%22%7D%7D%5D%2C%22logical_operator%22%3A%22and%22%7D%5D%2C%22logical_operator%22%3A%22and%22%2C%22label%22%3A%22text%22%7D%5D%2C%22logical_operator%22%3A%22and%22%7D%2C%22return_type%22%3A%22entry%22%2C%22request_options%22%3A%7B%22paginate%22%3A%7B%22start%22%3A0%2C%22rows%22%3A25%7D%2C%22results_content_type%22%3A%5B%22experimental%22%5D%2C%22sort%22%3A%5B%7B%22sort_by%22%3A%22score%22%2C%22direction%22%3A%22desc%22%7D%5D%2C%22scoring_strategy%22%3A%22combined%22%7D%2C%22request_info%22%3A%7B%22query_id%22%3A%2231fd7cd161fe37b0d15e07a1d6768be6%22%7D%7D

22. Martinsen JH, Saar D, Fernandes CB, Schuler B, Bugge K, Kragelund BB. Structure, dynamics, and stability of the globular domain of human linker histone H1.0 and the role of positive charges. *Protein Sci Publ Protein Soc*. 2022;31(4):918-932. doi:10.1002/pro.4281

23. Cutter AR, Hayes JJ. A brief review of nucleosome structure. *FEBS Lett*. 2015;589(20 Pt A):2914-2922. doi:10.1016/j.febslet.2015.05.016

24. Biterge B, Schneider R. Histone variants: key players of chromatin. *Cell Tissue Res*. 2014;356(3):457-466. doi:10.1007/s00441-014-1862-4

25. El Kennani S, Crespo M, Govin J, Pflieger D. Proteomic Analysis of Histone Variants and Their PTMs: Strategies and Pitfalls. *Proteomes*. 2018;6(3):29. doi:10.3390/proteomes6030029

26. Martire S, Banaszynski LA. The roles of histone variants in fine-tuning chromatin organization and function. *Nat Rev Mol Cell Biol*. 2020;21(9):522-541. doi:10.1038/s41580-020-0262-8

27. Filipescu D, Müller S, Almouzni G. Histone H3 Variants and Their Chaperones During Development and Disease: Contributing to Epigenetic Control. *Annu Rev Cell Dev Biol*. 2014;30(1):615-646. doi:10.1146/annurev-cellbio-100913-013311

28. El Kennani S, Adrait A, Shaytan AK, et al. MS_HistoneDB, a manually curated resource for proteomic analysis of human and mouse histones. *Epigenetics Chromatin*. 2017;10:2. doi:10.1186/s13072-016-0109-x

29. Oberdoerffer P, Miller KM. Histone H2A variants: Diversifying chromatin to ensure genome integrity. *Semin Cell Dev Biol*. 2023;135:59-72. doi:10.1016/j.semcdb.2022.03.011

30. Ryu DY, Pang WK, Adegoke EO, Rahman MS, Park YJ, Pang MG. Abnormal histone replacement following BPA exposure affects spermatogenesis and fertility sequentially. *Environ Int*. 2022;170:107617. doi:10.1016/j.envint.2022.107617

31. Chen P, Zhao J, Li G. Histone variants in development and diseases. *J Genet Genomics Yi Chuan Xue Bao*. 2013;40(7):355-365. doi:10.1016/j.jgg.2013.05.001

32. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000;403(6765):41-45. doi:10.1038/47412

33. Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta BBA - Gene Regul Mech*. 2014;1839(8):627-643. doi:10.1016/j.bbagrm.2014.03.001

34. Ree R, Varland S, Arnesen T. Spotlight on protein N-terminal acetylation. *Exp Mol Med*. 2018;50(7):1-13. doi:10.1038/s12276-018-0116-z

35. Sabari BR, Zhang D, Allis CD, Zhao Y. Metabolic regulation of gene expression through histone acylations. *Nat Rev Mol Cell Biol*. 2017;18(2):90-101. doi:10.1038/nrm.2016.140

36. Martin BJE, Brind'Amour J, Kuzmin A, et al. Transcription shapes genome-wide histone acetylation patterns. *Nat Commun*. 2021;12(1):210. doi:10.1038/s41467-020-20543-z

37. Shvedunova M, Akhtar A. Modulation of cellular processes by histone and non-histone protein acetylation. *Nat Rev Mol Cell Biol*. 2022;23(5):329-349. doi:10.1038/s41580-021-00441-y

38. Li Y, Chen X, Lu C. The interplay between DNA and histone methylation: molecular mechanisms and disease implications. *EMBO Rep*. 2021;22(5):e51803. doi:10.15252/embr.202051803

39. Mosammaparast N, Shi Y. Reversal of Histone Methylation: Biochemical and Molecular Mechanisms of Histone Demethylases. *Annu Rev Biochem*. 2010;79(1):155-179. doi:10.1146/annurev.biochem.78.070907.103946

40. Li KK, Luo C, Wang D, Jiang H, Zheng YG. Chemical and biochemical approaches in the study of histone methylation and demethylation. *Med Res Rev*. 2012;32(4):815-867. doi:10.1002/mrr.20228

41. Wesche J, Kühn S, Kessler B, Salton M, Wolf A. Protein arginine methylation: a prominent modification and its demethylation. *Cell Mol Life Sci CMLS*. 2017;74. doi:10.1007/s00018-017-2515-z

42. Chen Z, Zhang Y. Maternal H3K27me3-dependent autosomal and X chromosome imprinting. *Nat Rev Genet*. 2020;21(9):555-571. doi:10.1038/s41576-020-0245-9

43. Fabrizio P, Garvis S, Palladino F. Histone Methylation and Memory of Environmental Stress. *Cells*. 2019;8(4):339. doi:10.3390/cells8040339

44. Kungulovski G, Mauser R, Jeltsch A. Affinity reagents for studying histone modifications & guidelines for their quality control. *Epigenomics*. 2015;7(7):1185-1196. doi:10.2217/epi.15.59

45. El Kennani S, Crespo M, Govin J, Pflieger D. Proteomic Analysis of Histone Variants and Their PTMs: Strategies and Pitfalls. *Proteomes*. 2018;6(3):29. doi:10.3390/proteomes6030029

46. Britton LMP, Gonzales-Cope M, Zee BM, Garcia BA. Breaking the histone code with quantitative mass spectrometry. *Expert Rev Proteomics*. 2011;8(5):631-643. doi:10.1586/epr.11.47

47. Miller RM, Smith LM. Overview and considerations in bottom-up proteomics. *Analyst*. 2023;148(3):475-486. doi:10.1039/D2AN01246D

48. Patrie SM. Top-Down Mass Spectrometry: Proteomics to Proteoforms. In: Mirzaei H, Carrasco M, eds. *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*. Advances in Experimental Medicine and Biology. Springer International Publishing; 2016:171-200. doi:10.1007/978-3-319-41448-5_8

49. Zhang S (Weihua), Jian W. Recent advances in absolute quantification of peptides and proteins using LC-MS. *Rev Anal Chem*. 2014;33(1):31-47. doi:10.1515/revac-2013-0019

50. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu Rev Biomed Eng*. 2009;11(1):49-79. doi:10.1146/annurev-bioeng-061008-124934

51. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS*. 1999;20(18):3551-3567. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

52. UniProt: the Universal Protein Knowledgebase in 2023 | Nucleic Acids Research | Oxford Academic. Accessed October 18, 2023. https://academic.oup.com/nar/article/51/D1/D523/6835362?login=false

53. Goodwin RJA, Pennington SR, Pitt AR. Protein and peptides in pictures: Imaging with MALDI mass spectrometry. *PROTEOMICS*. 2008;8(18):3785-3800. doi:10.1002/pmic.200800320

54. Calvano CD, Monopoli A, Cataldi TRI, Palmisano F. MALDI matrices for low molecular weight compounds: an endless story? *Anal Bioanal Chem*. 2018;410(17):4015-4038. doi:10.1007/s00216-018-1014-x

55. Li D, Yi J, Han G, Qiao L. MALDI-TOF Mass Spectrometry in Clinical Analysis and Research. *ACS Meas Sci Au*. 2022;2(5):385-404. doi:10.1021/acsmeasuresciau.2c00019

56. Činčárová L, Lochmanová G, Nováková K, et al. A combined approach for the study of histone deacetylase inhibitors. *Mol Biosyst*. 2012;8(11):2937-2945. doi:10.1039/C2MB25136A

57. Sidoli S, Yuan ZF, Lin S, et al. Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis. *Proteomics*. 2015;15(9):1459-1469. doi:10.1002/pmic.201400483

58. Garcia BA, Mollah S, Ueberheide BM, et al. Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protoc*. 2007;2(4):933-938. doi:10.1038/nprot.2007.106

59. 8.7: Chemistry of Acid Anhydrides. Chemistry LibreTexts. Published December 27, 2021. Accessed December 6, 2023. https://chem.libretexts.org/Courses/University_of_Connecticut/Chem_2444%3A_(Second_Semester_Organic_Chemistry)_UConn/08%3A_Reactions_of_Carboxylic_Acid_Derivatives/8.07%3A_Chemistry_of_Acid_Anhydrides

60. Kuchaříková H, Dobrovolná P, Lochmanová G, Zdráhal Z. Trimethylacetic Anhydride–Based Derivatization Facilitates Quantification of Histone Marks at the MS1 Level. *Mol Cell Proteomics*. 2021;20. doi:10.1016/j.mcpro.2021.100114

61. Mandava V, Fernandez JP, Deng H, Janzen CJ, Hake SB, Cross GAM. Histone modifications in Trypanosoma brucei. *Mol Biochem Parasitol*. 2007;156(1):41-50. doi:10.1016/j.molbiopara.2007.07.005

62. Anderson DC, Green GR, Smith K, Selker EU. Extensive and Varied Modifications in Histone H2B of Wild-Type and Histone Deacetylase 1 Mutant Neurospora crassa. *Biochemistry*. 2010;49(25):5244-5257. doi:10.1021/bi100391w

63. Drogaris P, Wurtele H, Masumoto H, Verreault A, Thibault P. Comprehensive Profiling of Histone Modifications Using a Label-Free Approach and Its Applications in Determining Structure−Function Relationships. *Anal Chem*. 2008;80(17):6698-6707. doi:10.1021/ac800739d

64. Sidoli S, Yuan ZF, Lin S, et al. Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis. *PROTEOMICS*. 2015;15(9):1459-1469. doi:10.1002/pmic.201400483

65. Meert P, Dierickx S, Govaert E, et al. Tackling aspecific side reactions during histone propionylation: The promise of reversing overpropionylation. *Proteomics*. 2016;16(14):1970-1974. doi:10.1002/pmic.201600045

66. Maile TM, Izrael-Tomasevic A, Cheung T, et al. Mass Spectrometric Quantification of Histone Post-translational Modifications by a Hybrid Chemical Labeling Method[S]. *Mol Cell Proteomics*. 2015;14(4):1148-1158. doi:10.1074/mcp.O114.046573

67. Sidoli S, Fujiwara R, Garcia BA. Multiplexed data independent acquisition (MSX-DIA) applied by high resolution mass spectrometry improves quantification quality for the analysis of histone peptides. *Proteomics*. 2016;16(15-16):2095-2105. doi:10.1002/pmic.201500527

68. Lu C, Coradin M, Porter EG, Garcia BA. Accelerating the Field of Epigenetic Histone Modification Through Mass Spectrometry–Based Approaches. *Mol Cell Proteomics*. 2021;20. doi:10.1074/mcp.R120.002257

69. Progress in Top-Down Proteomics and the Analysis of Proteoforms | Annual Review of Analytical Chemistry. Accessed December 25, 2023. https://www.annualreviews.org/doi/10.1146/annurev-anchem-071015-041550

70. Middle-down proteomics: a still unexploited resource for chromatin biology: Expert Review of Proteomics: Vol 14, No 7. Accessed December 25, 2023. https://www.tandfonline.com/doi/full/10.1080/14789450.2017.1345632

71. Selected reaction monitoring for quantitative proteomics: a tutorial | Molecular Systems Biology. Accessed December 25, 2023. https://www.embopress.org/doi/full/10.1038/msb.2008.61

72. Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics *. *Mol Cell Proteomics*. 2012;11(11):1475-1488. doi:10.1074/mcp.O112.020131

73. Gillet LC, Navarro P, Tate S, et al. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis *. *Mol Cell Proteomics*. 2012;11(6). doi:10.1074/mcp.O111.016717

74. Sidoli S, Lin S, Xiong L, et al. Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) Analysis for Characterization and Quantification of Histone Post-translational Modifications *[S]. *Mol Cell Proteomics*. 2015;14(9):2420-2428. doi:10.1074/mcp.O114.046102

75. Krautkramer KA, Reiter L, Denu JM, Dowell JA. Quantification of SAHA-Dependent Changes in Histone Modifications Using Data-Independent Acquisition Mass Spectrometry. *J Proteome Res*. 2015;14(8):3252-3262. doi:10.1021/acs.jproteome.5b00245

76. Janssen KA, Coradin M, Lu C, Sidoli S, Garcia BA. Quantitation of Single and Combinatorial Histone Modifications by Integrated Chromatography of Bottom-up Peptides and Middle-down Polypeptide Tails. *J Am Soc Mass Spectrom*. 2019;30(12):2449-2459. doi:10.1007/s13361-019-02303-6

77. Holt MV, Wang T, Young NL. High-Throughput Quantitative Top-Down Proteomics: Histone H4. *J Am Soc Mass Spectrom*. 2019;30(12):2548-2560. doi:10.1007/s13361-019-02350-z

78. Gargano AFG, Shaw JB, Zhou M, et al. Increasing the Separation Capacity of Intact Histone Proteoforms Chromatography Coupling Online Weak Cation Exchange-HILIC to Reversed Phase LC UVPD-HRMS. *J Proteome Res*. 2018;17(11):3791-3800. doi:10.1021/acs.jproteome.8b00458

79. Yuan ZF, Arnaudo AM, Garcia BA. Mass Spectrometric Analysis of Histone Proteoforms. *Annu Rev Anal Chem*. 2014;7(1):113-128. doi:10.1146/annurev-anchem-071213-015959

80. Meert P, Dierickx S, Govaert E, et al. Tackling aspecific side reactions during histone propionylation: The promise of reversing overpropionylation. *Proteomics*. 2016;16(14):1970-1974. doi:10.1002/pmic.201600045

81. Paternoster V, Edhager AV, Sibbersen C, et al. Quantitative assessment of methylesterification and other side reactions in a standard propionylation protocol for detection of histone modifications. *Proteomics*. 2016;16(14):2059-2063. doi:10.1002/pmic.201500425

82. Dobrovolná P. *Kvantitativní proteomická analýza histonových kódů*. Masarykova univerzita, Přírodovědecká fakulta; 2020. Accessed December 5, 2023. https://is.muni.cz/auth/th/styy4/