New tools to enhance the detection of post-translational modifications: Development and application of a mass spectrometry-based proteomic software and of specific antibodies detecting propionylated lysines

By
Magnus Øverlie Arntzen

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science

Department of Biomedicine
University of Bergen, Norway
August 2009

Preface

This work was done during my time as a master student at the Institute of Biomedicine, Faculty of Medicine and Odontology. It has been carried out in the group of Dr. Clive S. D'Santos, and all laboratory and computer work was performed at the national FUGE platform in proteomics, PROBE, located at the University of Bergen.

First I would like to thank my main supervisor Clive S. D'Santos who had the idea behind this project and has been involved in all aspects during its development. He has a genuine interest and expertise in this particular field of research and I am very grateful that I could work under his supervision. Thank you for being a great supervisor and friend; good luck on setting up the new lab in Cambridge!

I would also like thank my secondary supervisor Aurelia E. Lewis whom I could not have done this without! Aurelia knows everything worth knowing in the lab and is an enormous resource. Aurelia also seem to always "know someone" in cases when we are out of something. Thanks for teaching me all the neat tricks in practical biochemistry!

A bunch of "thank you"s goes to all the people at PROBE, past and present: Olav Mjaavatten, Hilde K. Garberg, Anne P. Døskeland, Jill A. Opsahl, Otto Barholm, Ann Kristin Frøyset, Ann Cathrine Kroksveen, Eystein Oveland, Frode Berven, Kristian Flikka, Kari E. Fladmark, Maja Zahl, Frode Selheim and Yogita Sharma. Thanks for creating such a good working environment and for all the useful and useless discussions about science and whatever. It will be hard to leave you all; I hope to see you at conferences from time to time.

Thanks to Christoffer L. Osland and Christopher Rasch-Olsen Raa in Stix AS for great collaboration on the software development.

Thanks to the people in Bjørn Tore Giertsens group; especially Sigrun Hjelle for the useful discussions on p53 and all the reagents and antibodies she has provided us with.

Thanks to the people in Harald Wikers group; especially Hiwa Målen, Gustavo de Souza and Tina Søfteland for good discussions about mass spectrometry and HPLC in general.

And thank you again Ann Cathrine, for using your summer to read through my thesis. I appreciated that a lot!

And last but not least: Thank you Jeanette for supporting and believing in me through this intensive year, and to my two kids, Filip Ambrosius and Sara Linnea, for being my inspiration and diversion from science. I love you!

Bergen 2009,

Magnus Øverlie Arntzen

Table of contents

Summary

Mass spectrometry based proteomics has evolved into a powerful tool in characterising proteins and their respective function in the cell. A proteins cellular localisation, concentration and activity is regulated by several cellular processes including the regulation by protein post-translational modifications (PTMs). PTMs can be viewed as cellular switches that enable or disable specific functions for a protein and they represent a huge potential in understanding protein pathways in the context of protein and protein complex regulation. To date there are over 300 known PTMs which regulate protein structure and function, and their characterisation, as well as the discovery of new PTMs, have immense implications for cell biological and clinical research. However, the detection of PTMs using mass spectrometry is not straightforward. An automatic setup of a mass spectrometer traditionally selects the 2-5 most intense peptides for fragmentation and identification, leaving the rest, less abundant peptides, to pass. Since the majority of post-translationally modified peptides are low abundant and therefore probably not selected for fragmentation and identification, this is a sub-optimal procedure for detecting PTMs. Based on this we wanted to develop bioinformatics software able to increase fragmentation events and identifications of post-translationally modified peptides, using a targeted mass spectrometry approach.

This resulted in the development of POSTMan (POST-translational Modification analysis). The underlying principle of this software is to compare two LC-MS runs and search for post-translationally modified peptides as pairs with their unmodified counterparts, either between the two individual LC-MS runs or within one run. POSTMan was designed to be as generic as possible, meaning that in principle any PTM irrespective of mass could be identified with this software. We applied the software to an acetylated model protein of Cytochrome c to verify its capability in identifying the acetylated peptides, as well as the tumour suppressor protein p53 to assess the PTM patterns of acetylation and propionylation.

Propionylation was recently discovered as a novel *in vivo* PTM both on histones and non-histone proteins. The biological role of propionylation however, is currently unknown. Propionylation as an *in vivo* PTM, as well as its possible involvement in p53 regulation, was addressed using POSTMan and immunoblotting. Novel pan-specific anti-propionyllysine antibodies were raised, characterised and utilized in the study of p53 propionylation. These

antibodies showed specificity for propionyllysine with no cross reaction to acetyllysine, and will be important reagents that can be used for global proteomic investigation of propionylation as a regulatory PTM.

Abbreviations

1D	One dimensional
ACN	Acetonitrile
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
CID	Collision induced dissociation
Da	Dalton
DDA	Data dependent acquisition
DNA	Deoxyribonucleic acid
DOC	Deoxycholate
DTT	Dithiotreitol
EDTA	Ethylene diamine tetra acetic acid
ESI	Electrospray Ionization
FA	Formic acid
GPL	General public license
GUI	Graphical user interface
HEK	Human embryonic kidney
HPLC	High performance liquid chromatography
kDa	Kilo Dalton
LC	Liquid chromatography
m/z	Mass-to-charge
MALDI	Matrix assisted laser desorption ionization
MDF	Mass distance fingerprint
MDM2	Murine double minute 2
MeOH	Methanol
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry scan
MS_1	Mass spectrometry scan, level 1
MS_2	Mass spectrometry scan, level 2. Same as MS/MS
Mw	Molecular weight
NAD	Nicotinamide adenine dinucleotide
NAM	Nicotinamide
NHS	*N*-hydroxysuccinimide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
POSTMan	POST-translational Modification Analysis
PTM	Post-translational modification
Q	Quadrupole
RT	Retention Time
SDS	Sodium dodecyl sulphate
Sir2	Silent Information Regulator 2
SIRT1	Silent mating type information regulation 2 homolog
SLD	Soft laser desorption
TBST	Tris buffered saline Tween
TFA	Trifluoroacetic acid

ToF	Time of flight
Tris	Tris(hydroxymethyl)-aminmethan
TSA	Tricostatin A
Tsv	Tab-separated values
UV	Ultra violet
WT	Wild type
XML	Extensible markup language

1 Introduction

1.1 Mass spectrometry based proteomics

Over a decade ago, in 1996, Marc Wilkins introduced the term “proteome” as the “total protein complement of a genome” (Wilkins et al, 1996b). The proteome is dynamic, highly regulated and in constant change in respect to cell type, cell cycle, stimuli, and other tissue or organ specific factors. There is a multitude of regulatory processes taking place simultaneously in the cell: Synthesis and degradation of proteins, protein transport, gene transcription and translation, exocytosis and endocytotic mechanisms to name a few. The cell therefore requires exquisite regulation of these proteins to successfully orchestrate these processes. Prominent mechanisms of protein regulation include post-translational modifications (PTMs) by other proteins, protein cleavages and thus activation of new functions, localisation of proteins in different cell compartments and translocation between them, all which contribute to the final concentration, location and activity of each protein. Altered behaviour of proteins due to regulation, or lack of such, results in a myriad of pathological conditions; proteins are for this reason the primary targets for drug design. Understanding protein structure, function, regulation and interaction with other proteins is essential in biomedical research with the aim of understanding and curing disease. This huge regulatory dynamic presents technological challenges for the proteomic researcher, and proteomic experiments need to be interpreted in the context of the right cell conditions.

Up until 15 years ago, Edman degradation was the only established method to reliably determine the amino acid sequence of a protein. Peptides with a maximum length of 20-40 amino acids were sequenced in an iterative manner by removing N-terminal amino acids one by one and identified by ultra-violet high performance liquid chromatography (UV-HPLC). This required the protein to be pure, for example from 2-dimensional poly-acrylamide gel electrophoresis (2D-PAGE), and contain a free N-terminus. However, the majority of all proteins have their N-terminus blocked by for example a formyl, acetyl or acyl group. These proteins had to be first cleaved to produce peptides with free N-terminus and then one or more could be purified and now sequenced. The introduction of mass spectrometry based proteomics however has enabled moving from the analysis of single proteins to analysing the entire population of expressed proteins in a cell, the proteome (Wilkins et al, 1996a).

Mass spectrometry based proteomics includes the use of annotated genomes for peptide and protein identification. Annotated genomes are a prerequisite for high-throughput proteomics, while in the cases of un-annotated genomes it is limited by time-consuming peptide *de novo* sequencing. The high-throughput sequencing of genomes together with recent improvements in instrumentation, both mass spectrometers and liquid chromatography (presented later), are the major technological advances for proteomics-based research. Mass spectrometry based proteomics is usually divided into "top-down" and "bottom-up" proteomics. There exists some ambiguity about the usage of these two terms, but the most common usage describes top-down as a protein centric method where proteins are identified by mass spectrometry without the use of proteolytic enzymes. Bottom-up proteomics on the other hand describes a peptide centric method where proteins are cleaved by proteolytic enzymes, for example trypsin, and identified by the presence of their respective peptides. This latter proteomic approach is mostly used in large scale proteomics experiments due to its robustness to complex protein mixtures and is the method chosen for this thesis. To a large degree the complexity of the peptide mixture dictates which method of generating peptide ions (ionisation) would be most suitable. Matrix-assisted laser desorption ionization (MALDI) instruments are suitable for low complexity peptide mixtures whereas HPLC coupled electrospray ionization (ESI) instruments are ideal for highly complex peptide mixtures. In addition, a number of different mass analysers exist: quadrupole (Q), time-of-flight (ToF) tube, ion-trap and orbitrap to mention a few. Mass analysers can be used in different combinations inside the instrument, creating a tandem mass spectrometer. When peptides are analysed using a single mass analyser the data is termed MS_1 and the spectrum an MS spectrum. When peptides are analysed by two mass analysers (and usually but not necessarily fragmented in between) the data is termed MS_2 and the obtained spectrum an MS/MS spectrum. In order to fragment the peptides and generate MS/MS spectra a tandem mass spectrometer is needed.

1.1.1 Matrix-assisted laser desorption ionization time-of-flight instruments in proteomics

MALDI-ToF instruments have for more than a decade been the cornerstone of peptide and protein mass spectrometry, and represents an improvement on the soft laser desorption (SLD) work by Koichi Tanaka in 1987. Using SLD, Tanaka was able to ionise biomolecules up

to 34 kDa whereas further experiments with different matrices, laser wavelengths and laser types enabled MALDI to ionise even larger biomolecules. The principles of MALDI analysis rely on the transfer of energy from a laser source to the peptides embedded within a defined crystal structure or matrix. Peptides are embedded in the matrix and placed on a steel target plate to air dry. Inside the source of the mass spectrometer a laser beam fires high energy photons onto the matrix-peptide structure which absorbs this energy, evaporates and thus desorbs the peptides embedded in the matrix. The matrix catalyses the transfer of energy from the laser beam to the peptides, generating ionised peptides. Predominantly singly charged ion species are generated using MALDI and this allows relatively simple determination of peptide mass. A mass spectrum of all ionised peptides from a protein is termed a peptide mass fingerprint (PMF) and the masses for all the peptides therein can be searched against publically available protein databases for identification of the protein. However, if the protein is not in the database, for example due to non-existing annotated genomes, the protein can not be identified this way. If the instrument used is a tandem mass spectrometer, peptides can be selected for fragmentation either with collision induced dissociation (CID) or post-source decay resulting in a MS/MS fragment spectrum where the amino acid sequence can be determined. A detailed description of fragmentation theory and nomenclature is beyond the scope of this thesis; the reader is referred to excellent literature covering this topic (Roepstorff & Fohlman, 1984). However it is worth noting that a significant advantage of MALDI is its capacity to generate and detect low-mass ions during the fragmentation of peptides. In the low-mass region amino acids, dipeptides and immonium-ions can be observed. Immonium-ions are ions resulting from an internal peptide cleavage during fragmentation and specific immonium-ions and further fragments of these have previously been used to determine the presence of PTMs on specific amino acids within the peptide sequence (Kim et al, 2002). A significant disadvantage of MALDI is its capacity for high throughput analysis of peptides. MALDI is limited to low complexity peptide mixtures for the reason that all different peptides are detected in the same spectrum making ion suppression and reliable peak detection a problem. For further reading, a nice introduction to MALDI mass spectrometry has been published by Bonk and Humeny (Bonk & Humeny, 2001).

1.1.2 Electrospray hybrid quadrupole time-of-flight instruments in proteomics

A complementary method to MALDI for generating peptides as ions in the gas phase is ESI. Electrospray as ionization technique of peptides was developed by John Bennett Fenn in 1988; one year after SLD had its breakthrough. The big advantage over MALDI instruments is its readily coupling with HPLC and ionisation of peptides in solution. Peptides or complex peptide mixtures can be separated on a HPLC and eluted from the HPLC column straight into a spray needle. This separation of peptides is usually performed as reverse phase chromatography where the solution used for peptide elution is organic, for example methanol or acetonitrile, both being ideal mass spectrometry friendly eluents. In the presence of an applied voltage, positive charge is concentrated at the needle tip causing peptide ionisation in solution due to the low pH. The spray needle produces small droplets of solution which, during several bursts into smaller droplets, finally leaves the charge on the peptide itself. This is how peptides are ionised and transformed into the gaseous phase. ESI produces predominantly multiply charged (two or more protons per peptide) peptides compared with predominantly singly charged peptides generated using MALDI (see Section 1.1.1). This results in variations of charge states for every peptide, dividing the total intensity of a peptide on the different charge states, an issue that reduces sensitivity and increases spectrum complexity. This issue is relevant to the results and development of the software described in the main body of this thesis.

Most standard analyses apply data dependent acquisition (DDA) methods. Under these conditions ESI-QToF (hybrid quadrupole time-of-flight) instruments are configured to alternate between MS and MS/MS mode repeatedly throughout the elution. This produces one MS scan of all the peptides present at that time point (survey scan) followed by fragmentation scans of typically the 2-5 most intense peptides present in the survey scan before a new cycle starts over. This allows for automatic fragmentation of several peptides during elution and provides a comprehensive set of data for identification of the proteome. For further reading, an excellent introduction to QToF mass spectrometry has been published by Chernushevich *et al.*(Chernushevich et al, 2001).

1.1.3 On-line peptide separation using high performance liquid chromatography in proteomics

HPLC had its breakthrough during the 1970s and is today one of the most widely used analytical separation techniques. It has the ability to separate, identify and quantitate most, if not all, compounds that are present in a sample dissolved in liquid. In proteomics HPLC is used for peptide and/or protein separation, either off-line or on-line coupled to a mass spectrometer. Off-line HPLC allows for the use of non MS-friendly, non-high-throughput methods like salt-containing buffers, extensive gradients and multi-dimensional orthogonal fractionation. In on-line HPLC the mass spectrometer is directly coupled to the HPLC via the ESI source providing separation of peptides based on peptide properties like hydrophobicity before analysis in the mass spectrometer. This allows the analysis of very complex peptide mixtures such as cell lysates and provides an extra separation dimension to mass spectrometry. The peptides are separated leading to reduction in complexity. This allows the mass spectrometer to select more peptides for fragmentation and increases the number of total peptides fragmented. Mass spectrometry data acquired with a HPLC in front is termed LC-MS data. Furthermore, the mass, retention time and elution profile of every peptide can be utilized to extract information contributing to the verification of the peptide sequence. This has been utilized in the development of “accurate mass tag” (AMT) databases where the peptide mass and normalised retention time is stored with the peptide identity (May et al, 2007; Smith et al, 2002). An unknown peptide can be matched with this database by its retention time and mass, leading to sequence identification. Furthermore, peptides with PTMs can be distinguished by a difference in retention time. For example, a phosphorylated peptide tends to elute later in reverse-phase chromatography than its non-phosphorylated counter peptide. We exploited this characteristic in the presented studies and this will be discussed later.

It is important if not crucial to have reliable, reproducible chromatography if these co-ordinates are to be used as a comparative means of analysis from run to run. If not, technical issues can cause retention time drifts between two samples and result in poor comparison if not corrected using LC-alignment algorithms.

1.1.4 Liquid chromatography alignment

The software developed in this thesis is designed to compare successive liquid chromatography (LC) runs and therefore relies heavily on optimal chromatography conditions as well as the ability to generate reproducible chromatograms from run to run. This is not trivial as there are a number of issues which need to be considered. In practice there is always drift between two or more LC-MS datasets, either they are produced as successive samples or produced at completely different time points. However, the difference can be minimized by limiting the time between the analyses and keeping as many conditions as stable as possible, for example temperature and solvent composition. LC-MS data can be complicated by technical drifts in both mass-to-charge (m/z) and retention time of the peptides, the latter being the most predominant. LC drifts are caused by continuous accumulation of packing and contamination of the HPLC column, pressure and temperature fluctuations as well as instrument and flow rate stability. Inter- and intra-day variations in peptide elution time can cause poor comparison between samples, due to both linear and non-linear distortion of elution patterns. An example is shown in Figure 1-1. To compensate for HPLC drift between samples, alignment algorithms have to be applied. Several different approaches have been described in the literature and recently reviewed by Vandenbogaert *et al*. (Vandenbogaert et al, 2008). In brief, alignment approaches can be divided into two categories: Alignment of total-ion-count (TIC) or alignment of peptide features after processing and identification of the most important peptides. The aim is, independent of the method applied, to shift the retention time of peptides in one or more datasets in such a way that the datasets can be compared.

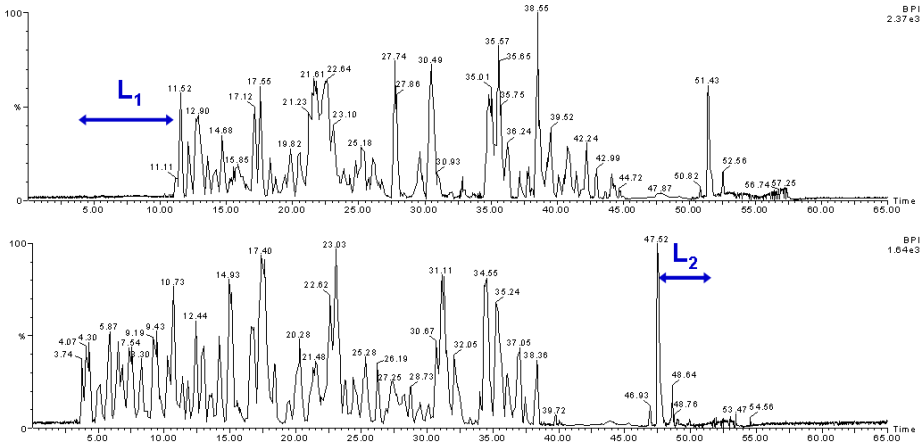

Figure 1-1. LC-alignment. Two chromatograms of the same sample where the bottom chromatogram has a 6 minutes negative drift. L_1 denotes the 6 minutes drift in the beginning of elution, while L_2 shows that the drift at the end of elution has changed from 6 minutes to 3 minutes. The drift is therefore not linear, and the peptide separation pattern throughout the chromatogram is also evidently distorted.

1.1.5 PTM detection using mass spectrometry

As mentioned above (see Section 1.1.2), the traditional operation of a HPLC coupled ESI instrument is to perform a MS survey scan followed by 2-5 MS/MS scans on the most abundant peptides. This is termed Data Dependent Acquisition (DDA) and serves as a good mean to sequence and identify as many peptides as possible during elution from the HPLC. When complex peptide mixtures are used, for example cell lysates, the number of peptides identified depends mainly on the instrument speed and the separating capacity of the column. But even with the most sophisticated state-of-the-art mass spectrometers, the low abundant peptides in a complex mixture will be lost when operating in DDA mode, leading to a systematic undersampling of peptide identifications (Figure 1-2). Proteins often suffer from sub-stoichiometric modifications; meaning that of the total amount of a specific protein only a few percent is modified in the cell. A cell lysate which has been digested and separated on HPLC prior to mass spectrometry has many peptides with a varying abundance, in which the post translationally modified peptides typically are low abundant species. In traditional DDA instrument operation, since only the 2-5 most abundant peptides are selected for fragmentation, the low abundant, and possibly post-translationally modified, peptides are often not selected for fragmentation. To address this, we wanted to develop bioinformatics software and increase the number of fragmented peptides with PTMs by targeted analysis. This software we have called POSTMan (POST-translational Modification analysis). Along similar lines other bioinformatic routes have also been chosen to address the same goal.

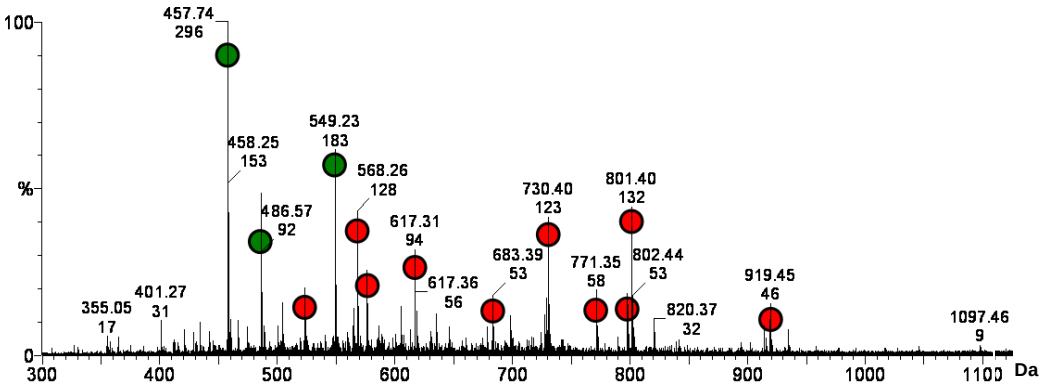

Figure 1-2. Peptide undersampling during mass spectrometry analysis operating in DDA mode. The three most abundant peptides (green circles) are chosen for fragmentation where the low abundant peptides (red circles) are lost. Peptides carrying PTMs are usually in the "red population" due to sub-stoichiometric modification.

1.2 Current bioinformatics software developed for proteomics and PTM detection

Bioinformatics has an enormous task in proteomics. From the second the peptide is recorded as a signal in the mass spectrometer the software takes over from upstream processing events. Bioinformatics software need to be designed to cope with several issues: Different proprietary MS file formats, the MS signal need to be filtered from the background noise, detected as a peak, de-isotoped to identify the monoisotopic m/z-value of the peptide, charge de-convoluted to find the correct mass of the peptide and, if to be compared, aligned in respect to retention time, intensity normalised and matched with another peptide within the same or a different sample. Furthermore, software search engines assign sequence identities to the peptides and cluster them together to provide protein information which can be further analysed by cluster analyses which groups the MS inferred proteins together in protein complexes and pathways using further software.

Within this myriad of bioinformatics software there are existing programs designed to optimize detection of PTMs in mass spectrometry data. For example the search engine Mascot (Perkins et al, 1999) has the ability to include a set of PTMs in the search for a peptide sequence identity, in addition to error tolerant searching. Other softwares utilize mass differences between peptides to detect PTMs. Programs like P-Mod (Hansen et al, 2005) matches MS/MS spectra to in silico generated spectra using sequence information,

ModifiComb (Savitski et al, 2006) matches identified MS/MS spectra to unidentified MS/MS spectra to reveal modified versions of the same peptide and Peptoscope (Potthast et al, 2007) introducing the Mass Distance Fingerprint (MDF), all take advantage of mass differences. The MDF in Peptoscope is determined by looking at mass differences of peptide precursors to globally assess modifications present in the sample. However, the current limitation of the previously listed programs is that they all operate on MS_2 level information (that is the information about only those peptides that were selected for fragmentation, usually only the abundant peptides). The programs listed utilize either the precursor mass of these peptides or the fragment data in the MS/MS spectra. This MS_2 data represents a subset, the most abundant, of all features present in an LC-MS run and is, for this reason and reasons mentioned in Section 1.1.3, likely not to include the majority of possibly post-translationally modified peptides. Software capable of using MS_1 level information (that is information about all peptides present in the sample, regardless of fragmentation) to extract possible post-translationally modified peptides and then target these peptides for fragmentation would not be restricted to only the most abundant peptides, and would be ideal as a tool to identify PTMs.

1.3 Biological applications of increased PTM detection

PTMs regulate protein function, stability, degradation, activation and cell localisation (Sims & Reinberg, 2008; Tate, 2008). They also represent a huge potential in understanding protein pathways in the context of protein and protein complex regulation. Developing proteomic and bioinformatic tools to interrogate the vast array of existing PTMs (there are currently over 300 known PTMs which regulate protein structure and function) which are known to contribute to protein function and in addition revealing new PTMs has immense implications for cell biological and clinical research. POSTMan was originally designed to interrogate MS data to reveal PTMs such as acetylated peptides. The program however has broader applications, enabling the identification of several PTMs. In the course of these experiments we were able to identify a novel, and at the time uncharacterised PTM, propionylation. We therefore sought to address the possibility that propionylation occurs *in vivo* and ask the question, how prevalent is this PTM as a regulator of protein function? These studies used as a model protein the tumour suppressor protein p53, and we applied POSTMan as well as biochemical methods to the identification of propionylated lysine

residues on p53. Propionyl-lysine specific antibodies have been raised and provide an important reagent for global proteomic investigation of propionylation as a regulatory PTM. The following sections therefore highlight relevant aspects of PTMs and the regulation of p53 and finally introduce current limited data describing propionylation as an *in vivo* protein PTM.

1.3.1 The tumour suppressor protein p53

The tumour suppressor protein p53 is a transcription factor of 44 kDa, short-lived, non-abundant and shown to be highly regulated by PTMs (Sims & Reinberg, 2008). It was for this reason we chose to use p53 as a model protein for analysis by POSTMan. p53 is involved in cell proliferation, angiogenesis, senescence, DNA repair, cell cycle arrest and apoptosis and is therefore a key regulator of cell fate. p53 is the most frequently mutated gene in human cancers. In 50% of all human cancers p53 is lost or expressed as an inactive mutant (Toledo & Wahl, 2006). Its tumour suppressor function is based on evidence from genetic experiments showing that p53 knock-out mice develop tumours with high frequency (Levine, 1997). p53 was first characterised as “The guardian of the genome” (Lane, 1992) due to its anti-proliferative response to DNA damage or checkpoint failure. p53 can also be dangerous to the cell: It is capable of killing cells via transcription-dependent and transcription-independent manners in the nucleus and mitochondria respectively (Mihara et al, 2003; Vousden & Lu, 2002). Transcription-dependent p53 mediated apoptosis includes activation of pro-apoptotic genes like the BH3-only (Bcl2 homology 3) proteins Puma (p53 upregulated modulator of apoptosis) and Noxa, Bax (Bcl2 associated X protein), p53-AIP (p53-regulated apoptosis-inducing protein), Apaf-1 (apoptotic peptidase activating factor 1), and PERP (p53 apoptosis effector related to PMP-22) and also transcriptional repression of anti-apoptotic genes like Bcl2 and IAPs (inhibitors of apoptosis) ((Johnstone et al, 2002), and references herein). The transcription-independent p53-mediated apoptosis includes translocation of a fraction of p53 to mitochondria where it can directly induce permeablisation of the outer mitochondrial membrane by complexing with BclXL and Bcl2 proteins; providing Cytochrome c release and following caspase cascade activation (Mihara et al, 2003).

p53 is heavily regulated by PTMs. In normal, unstressed cells p53 exist at low concentrations with a half-life of 5-30 minutes (Moll & Petrenko, 2003). This is mostly due to its constant degradation mediated by the p53 interacting protein MDM2 (Murine double

minute 2). MDM2 is an ubiquitin ligase and contributes to the ubiquitination of p53 and subsequent targeting for degradation by the proteasome. In response to stress signals p53 is rapidly stabilised by inhibiting this continuous degradation resulting in an increase in cellular p53 levels with an extended half-life up to several hours. The stabilisation of p53 is a hallmark of many cellular stress pathways and enhances the p53 transcriptional activity. Upon such activation p53 initiates transcription of specific genes in response to the nature of the signal; so-called p53 responsive elements. p53 receives different stress signals via PTMs on distinct amino acids within its sequence activating p53 as a transcription factor and initiating defined downstream cellular programs. For example, if the pool of ribonucleoside triphosphate gets too low or there are too few ribosomes to sustain the cell cycle, p53 is activated and cell cycle progression is arrested (Harris & Levine, 2005). Under hypoxic conditions, p53 is activated resulting in cell cycle arrest, senescence or apoptosis (Harris & Levine, 2005). In receiving these signals p53 can be phosphorylated, acetylated, methylated, ubiquitinated, sumoylated, neddylated (Riley & Maher, 2007) and propionylated (Arntzen et al. unpublished data, (Cheng et al, 2008)) in a combinatorial fashion (Figure 1-3). For example, Lys^{372} alone can either be acetylated (Gu & Roeder, 1997; Liu et al, 1999), ubiquitinated (Desterro et al, 2000; Rodriguez et al, 2000), methylated (Chuikov et al, 2004) or neddylated (Xirodimas et al, 2004) and the presence of one of these lysine modifications on this lysine precludes the conjugation by another modification. The post-translational regulation of p53 is initiated by a number of kinases (ATM (Ataxia telangiectasia mutated), ATR (Ataxia telangiectasia and Rad3-related protein), Chk1, Chk2), acetyl transferases (CBP (CREB-binding protein), p300, PCAF, TRAF (TNF receptor-associated factor)), methylases, ubiquitin ligases (MDM2, Cop-1 (constitutive photomorphogenic protein), Pirh-2 (p53-induced RING-H2 protein)), sumo ligases, and recently discovered propionyl transferases (Arntzen et al. unpublished data, (Cheng et al, 2008; Leemhuis et al, 2008)). In addition, a number of phosphatases, deacetylases, ubiquitinases and inhibitors of ubiquitin ligases also regulate p53 function by removing the PTMs. The ubiquitin ligase MDM2 is the key regulator of p53 and is involved in 60% of the seven negative and three positive feedback loops of p53.

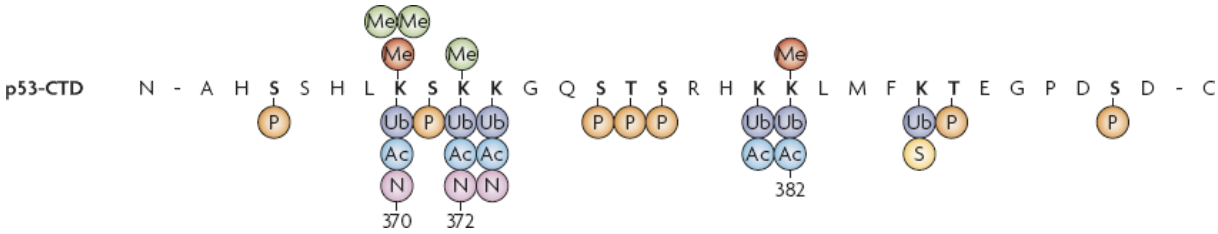

Figure 1-3. C-terminal domain of tumour suppressor protein p53. p53 is substrate for numerous PTMs: It can be phosphorylated (P), acetylated (Ac), methylated (Me), ubiquitinated (Ub), sumoylated (S), neddylated (N) and propionylated (not shown in figure). All contributing to the diverse and complex regulation of p53. Figure is borrowed from (Sims & Reinberg, 2008).

1.3.2 Propionylation as a lysine post-translational modification

The repertoire of lysine PTMs has recently been extended to include propionylation and butyrylation. Mass spectrometry was used to identify peptides carrying propionylated and butyrylated lysine residues from histone H4 (HH4) *in vitro* and *in vivo* (Chen et al, 2007). The modification of HH4 was catalysed *in vitro* by the acetyltransferases p300 and CREB-binding protein (CBP). In the same report p300 and CBP were also capable of *in vitro* propionylating and butyrylating p53 at Lys^{319} and Lys^{320}. These propionylated and butyrylated lysines from both HH4 and p53 are also known sites of acetylation in these proteins suggesting that like acetylation, propionylated and butyrylated lysine residues may have functional significance *in vivo*. In another report, the acetyltransferase P/CAF was suggested to possess propionyltransferase activity *in vitro* (Leemhuis et al, 2008). A peptide similar to the histone H3 (HH3) N-terminal tail was incubated with propionyl-CoA in the presence or absence of P/CAF. Using mass spectrometry a peptide mass shift of 56 Da corresponding to a propionyl group was observed in the presence of P/CAF but not in its absence. Furthermore, P/CAF was only able to propionylate Lys^{14} on HH3, the same site known to be acetylated by P/CAF (Schiltz et al, 1999). Enzymes that catalyse the removal of propionyl groups are at current date not known, but histone deacetylases (HDACs) from rat liver has shown depropionylation activity (Riester et al, 2004), as well as members of the sirtuin family of HDACs (Garrity et al, 2007).

1.4 Thesis aims

Detection of specific PTMs is not always straightforward. In mass-spectrometry based proteomics, several softwares have been designed to enhance PTM detection. However,

they are currently hampered by their limitation to MS_2 data. Software capable of using MS_1 data to extract possible post-translationally modified peptides would not be restricted to only the most abundant peptides, as MS_2 data is, and would therefore be a more optimal tool in PTM detection. Based on this rationale, we wanted to develop such a software and use this to elucidate PTMs on the tumour suppressor protein p53. Furthermore, propionylation was recently reported as a novel *in vivo* protein PTM. We wanted therefore to generate and characterise novel antibodies to enhance detection of *in vivo* propionylated proteins.

This master thesis had three aims:

- First aim was to develop bioinformatics software designed to enhance detection of post-translationally modified peptides using mass spectrometry. This resulted in the software POSTMan (POST-translational Modification analysis).
- Second aim was to use POSTMan to complement biochemical methods to assess the PTM pattern of the tumour suppressor protein p53, with emphasis on acetylation and propionylation, the latter being a novel PTM with currently unknown functional significance.
- The third aim was to raise specific anti-propionyllysine antibodies to complement existing proteomic tools to unravel the functional significance of propionylation as a global PTM.

2 Materials

Table 2-1 Materials used in cell culture

Material	*Supplier*
2-propanol	Riedel-De Haën, Honeywell, Seeze, Germany
Bürker chamber	Superior Marienfeld, Lauda-Königshofen, Germany
Calcium chloride dihydrate, $CaCl_2.2H_2O$	Sigma Aldrich, St.Louis, MO, USA
CryoTube™ Vials	Nunc, Thermo Fisher Scientific, NY, USA
Dimethyl sulfoxide (DMSO)	Sigma Aldrich, St.Louis, MO, USA
Dulbecco's modified eagle's medium (DMEM)	Sigma Aldrich, St.Louis, MO, USA
Fetal bovine serum (FBS)	Invitrogen, Carlsbad, CA, USA
HEBS buffer, 2x: 274 mM NaCl 10 mM KCl 140 µM $Na_2HPO_4.2H_2O$ 1.1 mM dextrose = D+ glucose 42 mM HEPES Adjust pH to ~7 with 2 M NaOH and test for precipitate size.	 Fluka, Sigma Aldrich, St.Louis, MO, USA Merck KGaA, Darmstadt, Germany Merck KGaA, Darmstadt, Germany Sigma Aldrich, St.Louis, MO, USA Sigma Aldrich, St.Louis, MO, USA Fluka, Sigma Aldrich, St.Louis, MO, USA
HEK293T lysis buffer, 1x: 50 mM Tris(hydroxymethyl)-aminmethan 50 mM KCl 10 mM EDTA Adjust pH to 8.0 with HCl 1% Nonidet P-40 (NP-40) Complete Mini Protease inhibitor cocktail 1 µM Tricostatin A 10 mM Nicotinamide	 Merck KGaA, Darmstadt, Germany Merck KGaA, Darmstadt, Germany GE Healthcare, Chalfont St. Giles, UK Merck KGaA, Darmstadt, Germany USB corporation, Cleveland, OH, USA Roche Applied Science, Basel, Switzerland Sigma Aldrich, St.Louis, MO, USA Sigma Aldrich, St.Louis, MO, USA
HeLa lysis buffer: RIPA buffer Complete Mini Protease inhibitor cocktail 1 µM Tricostatin A 10 mM Nicotinamide	 (see below) Roche Applied Science, Basel, Switzerland Sigma Aldrich, St.Louis, MO, USA Sigma Aldrich, St.Louis, MO, USA
Lipofectamine 2000	Invitrogen, Carlsbad, CA, USA
Nicotinamide	Sigma Aldrich, St.Louis, MO, USA
Optimem	Invitrogen, Carlsbad, CA, USA
Pencillin-Streptomycin	Sigma Aldrich, St.Louis, MO, USA
Phosphate buffered saline (PBS) 10x, pH 7.0	Gibco, Invitrogen, Carlsbad, CA, USA
RIPA buffer: 50 mM Tris(hydroxymethyl)-aminmethan 150 mM NaCl Adjust pH to 8.0 with HCl 0.5% Deoxycholate (DOC) 1% Nonidet P-40 (NP-40) 0.1% SDS	 Merck KGaA, Darmstadt, Germany Fluka, Sigma Aldrich, St.Louis, MO, USA Merck KGaA, Darmstadt, Germany Fluka, Sigma Aldrich, St.Louis, MO, USA USB corporation, Cleveland, OH, USA Merck KGaA, Darmstadt, Germany
RPMI	Invitrogen, Carlsbad, CA, USA
Trichostatin A	Sigma Aldrich, St.Louis, MO, USA
Trypsin-EDTA solution	Sigma Aldrich, St.Louis, MO, USA
Water bath, Lauda Ecoline 011	Lauda, Dipl.Ing Houm AS, Oslo, Norway

Table 2-2 Cell lines

Material	*Supplier*
HEK293T cells Human embryonic kidney cell line	The German resource centre for biological material, Braunschweig, Germany
HeLa cells Human cervical cancer epithelial cell line	The German resource centre for biological material, Braunschweig, Germany

Table 2-3 Plasmids used in cell culture

Material	*Supplier*
HA-p300	Kind gift from Dr. Nullin Divecha, Paterson Institute for Cancer Research, The University of Manchester, England
pCMVβ-p300	Kind gift from Dr. Marit Bakke, Institute of Biomedicine, University of Bergen, Norway

Table 2-4 Materials used in protein gel electrophoresis

Material	*Supplier*
Acetic acid	Fluka, Sigma Aldrich, St.Louis, MO, USA
Bovine serum albumin (BSA), Cohn V fraction	Sigma Aldrich, St.Louis, MO, USA
Develop solution (silver staining): 283 mM Na_2CO_3 26 µM $Na_2S_2O_3$ 0.02% Formaldehyde	 Sigma Aldrich, St.Louis, MO, USA Sigma Aldrich, St.Louis, MO, USA Fluka, Sigma Aldrich, St.Louis, MO, USA
Ethanol	Arcus AS, Oslo, Norway
Filter, 0.45 mm	Schleicher & Schuell MicroScience GmbH, Dassel, Germany
HEK293T lysis buffer, 1x	*See Table 2-1 Materials used in cell culture*
HeLa lysis buffer	*See Table 2-2 Materials used in cell culture*
Novex mini-cell	Invitrogen, Carlsbad, CA, USA
NuPAGE® 4-12% Bis-Tris gels	Invitrogen, Carlsbad, CA, USA
NuPAGE® MOPS SDS running buffer, 20x	Invitrogen, Carlsbad, CA, USA
PhastGel™ Blue R	GE Healthcare, Chalfont St. Giles, UK
Pierce® BCA protein assay kit	Pierce biotechnology, Rockford, IL, USA
SDS-PAGE sample buffer, 1x: 80 mM Tris(hydroxymethyl)-aminmethan 50 mM DTT Adjust pH to 6.8 with HCl 2% SDS 10% Glycerol 0.025% Bromphenol blue	 Merck KGaA, Darmstadt, Germany Sigma Aldrich, St.Louis, MO, USA Merck KGaA, Darmstadt, Germany Merck KGaA, Darmstadt, Germany Merck KGaA, Darmstadt, Germany J.T. Baker, Mallinckrodt Baker B.V., Deventer, The Netherlands
Sea blue®Plus2 pre-stained protein standard	Invitrogen, Carlsbad, CA, USA
Silvernitrate, $AgNO_3$	Fluka, Sigma Aldrich, St.Louis, MO, USA
Sodiumthiosulphate, $Na_2S_2O_3$	Sigma Aldrich, St.Louis, MO, USA

Table 2-5 Materials used in western blotting

Material	*Supplier*
Bovine serum albumin (BSA), Cohn V fraction	Sigma Aldrich, St.Louis, MO, USA
Instant skimmed milk powder	Sainsbury's supermarkets Ltd. London, UK
Methanol, Chromasolv® (MeOH)	Sigma Aldrich, St.Louis, MO, USA
Nitrocellulose transfer membranes, Protran®	Whatman® Gmbh, Dussel, Germany
NuPAGE® Transfer buffer, 20x	Invitrogen, Carlsbad, CA, USA
Ponceau S	Sigma Aldrich, St.Louis, MO, USA
Super Signal® West pico Chemiluminescent substrate	Pierce, Rockford, IL, USA
Tris buffered saline tween, 1x: 20 mM Tris(hydroxymethyl)-aminmethan 300 mM NaCl Adjust pH to 7.4 with HCl 0.2% Tween-20	 Merck KGaA, Darmstadt, Germany Fluka, Sigma Aldrich, St.Louis, MO, USA Merck KGaA, Darmstadt, Germany Merck KGaA, Darmstadt, Germany
XCell II™ blot module	Invitrogen, Carlsbad, CA, USA

Table 2-6 Antibodies used in western blotting and immunoprecipitation

Material	*Supplier*	*Dilution*	*Host*	*Method*
Anti acetyl lysine	Biomol International, LP, Plymouth Meeting, PA, USA	1:500	Rabbit	WB
Anti acetyl lysine	Cell Signaling Technology®, Inc., Danvers, MA, USA	1:500	Rabbit	WB
Anti acetyl lysine	ImmuneChem pharmaceuticals, Burnaby, Canada	1:500	Rabbit	WB
Anti acetyl lysine	Upstate Cell Signaling solutions, Lake Placid, NY, USA	1:500	Rabbit	WB
Anti actin	Sigma Aldrich, St.Louis, MO, USA	1:5000	Mouse	WB
Anti mouse, horse raddish peroxidase conjugated	Jackson Immuno Research Europe Ltd., Suffolk, UK	1:10 000	Goat	WB
Anti p53 (DO-1), 10x	Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA		Mouse	IP
Anti p53 (DO-12)	Invented by Jean-Christope Bourdon University of Dundee, UK and produced by Borek Vojtesek, Masaryk memorial Cancer Institute, Czech Republic	1:1000	Mouse	WB
Anti p53 acetyl Lys382	Cell Signaling Technology®, Inc., Danvers, MA, USA	1:1000	Rabbit	WB
Anti propionyl lysine	Invented by Clive D'Santos, Institute of Biomedicine, University of Bergen, Norway	1:1000	Rabbit	WB
Anti rabbit, horse raddish peroxidase conjugated	Jackson Immuno Research Europe Ltd., Suffolk, UK	1:10 000	Goat	WB

Table 2-7 Materials used in immunoprecipitation

Material	*Supplier*
Dimethyl pimelimidate dihydrochloride (DMP)	Sigma Aldrich, St.Louis, MO, USA
Ethanolamine: 0.2 M Ethanolamine Adjust pH to 9.0 with HCl	 Sigma Aldrich, St.Louis, MO, USA Merck KGaA, Darmstadt, Germany
Hamilton syringe tip, gauge 22/2"/3	Hamilton Company, Reno, NV, USA
HEK293T lysis buffer, 1x	*See Table 2-1 Materials used in cell culture*
HEK293T lysis buffer, 5x: 250 mM Tris(hydroxymethyl)-aminmethan 250 mM KCl 50 mM EDTA Adjust pH to 8.0 with HCl 5% Nonidet P-40 (NP-40)	 Merck KGaA, Darmstadt, Germany Merck KGaA, Darmstadt, Germany GE Healthcare, Chalfont St. Giles, UK Merck KGaA, Darmstadt, Germany USB corporation, Cleveland, OH, USA
Mouse IgG	Zymed Laboratories, Inc., South San Fransisco, CA, USA
Phosphate buffered saline, 10x, pH 7.0	Gibco, Invitrogen, Carlsbad, CA, USA
pH-paper, 0-14	Merck KGaA, Darmstadt, Germany
Protein G agarose beads	Invitrogen, Carlsbad, CA, USA
SDS-PAGE sample buffer	*See Table 2-3 Materials used in protein gel electrophoresis*
Sodium hydroxide, NaOH, 50%	Fluka, Sigma Aldrich, St.Louis, MO, USA
Sodium tetraborate buffer: 0.2 M $Na_2B_4O_7.10H_2O$ Adjust pH to 9.0 with NaOH	 Sigma Aldrich, St.Louis, MO, USA Fluka, Sigma Aldrich, St.Louis, MO, USA

Table 2-8 Materials used for mass shift induction

Material	*Supplier*
Bovine serum albumin (BSA), Fatty acid free	Sigma Aldrich, St.Louis, MO, USA
Dithiothreitol (DTT)	Sigma Aldrich, St.Louis, MO, USA
Human recombinant HA-SIRT1	Kind gift from Dr. Nullin Divecha, Paterson Institute for Cancer Research, The University of Manchester, England
Magnesium chloride, $MgCl_2.6H_2O$	Sigma Aldrich, St.Louis, MO, USA
N-hydroxysuccinimide acetyl ester	Kind gift from Pieter van der Veken, Department of Medicinal Chemistry, University of Antwerp, Belgium
N-hydroxysuccinimide buteryl ester	Kind gift from Pieter van der Veken, Department of Medicinal Chemistry, University of Antwerp, Belgium
N-hydroxysuccinimide propionyl ester	Kind gift from Pieter van der Veken, Department of Medicinal Chemistry, University of Antwerp, Belgium
Recombinant p53	Sigma Aldrich, St.Louis, MO, USA
SDS-PAGE sample buffer	*See Table 2-4 Materials used in protein gel electrophoresis*
Sodium carbonate buffer: 0.1 M Na_2CO_3 Adjust pH to 8.4 with HCl	 Sigma Aldrich, St.Louis, MO, USA Merck KGaA, Darmstadt, Germany
Trifluoroacetic acid (TFA)	Fluka, Sigma Aldrich, St.Louis, MO, USA
Tris-HCl: 0.2 M Tris(hydroxymethyl)-aminmethan Adjust pH to 9.6 with HCl	 Merck KGaA, Darmstadt, Germany Merck KGaA, Darmstadt, Germany
β-NAD^+	Sigma Aldrich, St.Louis, MO, USA

Table 2-9 Materials used in mass spectrometry

Material	*Supplier*
Acetonitile (ACN)	Riedel-De Haën, Honeywell, Seeze, Germany
Ammonium bicarbonate, NH_4HCO_3	Sigma Aldrich, St.Louis, MO, USA
C_{18} Empore™ membrane	3M, Maplewood, MN, USA
D10 pipet tips	Gilson, Inc., Middleton, WI, USA
Dithiothreitol (DTT)	Sigma Aldrich, St.Louis, MO, USA
Formic acid (FA)	Fluka, Sigma Aldrich, St.Louis, MO, USA
Gel loader tips	Eppendorf, Hamburg, Germany
Hamilton syringe tip, size 22/2"/3 and 16/2"/3	Hamilton company, Reno, NV, USA
Idoacetamide	Sigma Aldrich, St.Louis, MO, USA
Matrix (α-cyanno-4-hydroxycinnamic acid)	Agilent Technologies, Inc., Santa Clara, CA, USA
Methanol, Chromasolv® (MeOH)	Sigma Aldrich, St.Louis, MO, USA
Peptide calibration standard II	Bruker Daltronics Gmbh, Bremen, Germany
Scalpel	Swann Mortens Ltd. Sheffield England
Trifluoroacetic acid (TFA)	Fluka, Sigma Aldrich, St.Louis, MO, USA
Trypsin Porcine sequence grade modified	Promega, Fitchburg, WI, USA
Wash buffer: 50% ACN 25 mM NH_4HCO_3	 Riedel-De Haën, Honeywell, Seeze, Germany Sigma Aldrich, St.Louis, MO, USA

Table 2-10 Technical equipment

Material	*Supplier*
Centrifuge for cells, Rotina 35 R	Hettich Zentrifugen, Tuttlingen, Germany
Centrifuge for eppendorf tubes	Heraeus, DJB Labcare Ltd., Buckinghamshire, England
Chemiluminisence camera LAS-3000	Fujifilm Europe GmbH, Dusseldorf, Germany
Electrophoresis power supply, EPS 601	Amersham biosciences,GE Healthcare, Chalfont St. Giles, UK
Eppendorf thermomixer comfort	Eppendorf, Hamburg, Germany
Gel shaker, IKA® KS 260 basic	IKA® Werke, GmbH & Co, Staufen, Germany
Ground steel MALDI target	Bruker Daltonics Gmbh, Bremen, Germany
Holten Lamin Air, model 1.2	Thermo Fisher Scientific, Inc., Waltham, MA, USA
MALDI ToF-Tof mass spectrometer	Bruker Daltonics GmbH, Bremen, Germany
Micro wellplate reader, SpectraCount™	Packard, Wolf Laboratories Limited, York, UK
Milli-Q water system for dd.H_2O	Millipore A/S, Oslo, Norway
Nitrogen tank, Locator 8 plus	Thermolyne® Sigma Aldrich, St.Louis, MO, USA
Parafilm Laboratory film	Pechiney plastic packaging company, Chicago, IL, USA
Pipets	Gilson, Inc., Middleton, WI, USA
Q-ToF mass spectrometer	Waters Corporation, Milford, MA, USA
Rotator, model L28	Labinco B.V., Breda, The Netherlands
Steri-Cycle CO_2 cell incubator	Forma, Thermo Fisher Scientific, Inc., Waltham, MA, USA
Termaks oven	Termaks AS, Bergen, Norway
Tissue culture dishes	Sarstedt AG & Co., Nümbrecht, Germany
Ultimate 3000 nanoHPLC	Dionex Corporation, Sunnyvale, CA, USA
Vacuum centrifugator 5301 (SpeedVac)	Eppendorf, Hamburg, Germany
Vortex, bench top	Gilson, Inc., Middleton, WI, USA

Table 2-11 Software and protein databases

Material	*Version*	*Application*	*Vendor / Developer*
Chromeleon	6.8	HPLC control	Dionex Corporation, Sunnyvale, CA, USA
Flex analysis	2.4	MALDI analysis	Bruker Daltonics Gmbh, Bremen, Germany
Flex control	2.0	MALDI control	Bruker Daltonics Gmbh, Bremen, Germany
Gimp	2.2	Graphics	GNOME project
Imagereader LAS-3000	2.0	Chemiluminescence	Fujifilm Europe GmbH, Dusseldorf, Germany
Mascot	2.1.04	Database searching	Matrix science, London, UK
MassLynx	4.0	Q-ToF control	Waters Corporation, Milford, MA, USA
Masswolf	Nov06	MS conversion	Seattle Proteome Centre, Institute for Systems Biology, Seattle, WA, USA
mMass	2.4	MALDI analysis	Institute of Chemical Technology in Prague, the Czech Republic
MSight	2.A.6	MS visualization	Swiss Institute of Bioinformatics, Basel, Switzerland
msInspect	1.2	MS feature extraction	Computational Proteomics Laboratory, Fred Hutchinsons Cancer Reasearch Center, Seattle, WA, USA
Photoshop	8.0	Graphics	Adobe, San Jose, CA, USA
POSTMan	2.0	Data analysis	PROBE, Institute for Biomedicine, University of Bergen, Norway
ProteinLynx	2.2	Q-ToF analysis	Waters Corporation, Milford, MA, USA
Qt	4.3.2	Software development	Qt Software, Nokia Corporation, Helsinki, Finland
ReAdW	3.5.4	MS conversion	Seattle Proteome Centre, Institute for Systems Biology, Seattle, WA, USA
Swiss-Prot protein database	56.5	Mascot searching	Swiss Institute of Bioinformatics, Basel, Switzerland and European Bioinformatics Institute, Cambridge, UK
ImageJ	1.41o	Image analysis	National Institutes of Health, USA

3 Methods

3.1 Development of POSTMan

3.1.1 Programming language, libraries, and software license

For the development of POSTMan the programming language C++ was chosen over Java (Sun) and Visual Basic (Microsoft) due to its seamless interaction with the Qt libraries (Nokia) and the scientific open source Qwt libraries (available from http://qwt.sourceforge.net/); in addition to previous experience with the language. POSTMan's predecessor PTMfinder (unpublished data) was also developed in C++ for console use only, and had very limited functionality, but served as a proof of principle. C++ itself offers high program operating speed due to binary compilation of source files and is the language of choice for software performing a high amount of calculations. The Qt libraries allowed for nice graphical user interface without extensive programming. Programming of core functions in POSTMan was set out to Stix AS and was beyond the scope of this master thesis. As license we chose the GNU General Public License version 2 which assure open source also on derivative projects.

3.1.2 Development of the POSTMan pre-processor

The POSTMan pre-processor was designed as a standalone process rather than a part of the analysis tool. This was to ensure the possibility to operate the analysis tool while the pre-processor was running. To make POSTMan work with different instruments, several raw file-formats are accepted through converters. The pre-processor takes advantage of raw-to-mzXML converters available from Seattle Proteome Centre (http://tools.proteomecenter.org/software.php) and integrate the converters for MassLynx 4.0, MassLynx 4.1, Xcalibur, and Analyst into the software package. Bruker files needs to be manually converted into mzXML using CompassXport (Bruker) before they can be used with POSTMan. Peak detection and feature finding is performed by passing the mzXML file into msInspect available from Fred Hutchinsons Cancer Reasearch Center (http://proteomics.fhcrc.org/CPL/msinspect.html) where all peptides are detected and assigned peptide properties. msInspect creates a tab-separated-values (tsv) file containing detected peptide features: m/z, retention time, de-convoluted mass value, charge, intensity,

and other peptide related information. These files are the input of the POSTMan analysis tool. The POSTMan pre-processor allows batch conversion and feature finding of multiple files prior to POSTMan analysis.

3.1.3 Development of the POSTMan analysis tool

The POSTMan analysis tool was created to take one or two tsv-files as input and load all peptides as spots into the Graphical plot (a plot displaying m/z versus retention time as X and Y axis respectively). The peptide intensities are reflected in the colour intensity of the spots. Because of the dynamic range of intensities in a single LC-MS-run the 10-logarithm of the intensities are used for colour discrimination. Peptides originating from reference/untreated file is coloured green while peptides originating from treated file is coloured blue. The POSTMan analysis tool was given five essential functions: 1) LC-alignment, 2) Global PTM assessment, 3) Detection of peptide pairs, 4) Filtering, and 5) Inclusion list output. These functions also represent the typical workflow of a POSTMan analysis. The pseudo code of these functions are listed in appendix A.

1) LC-alignment

POSTMan aligns LC-runs in a two way, semiautomatic, feature-based approach using whole or a subset of the data loaded. Exemplar is only multiply charged peptides with intensity above 100 counts. First the user can visually slide one run superimposed on the other in the time dimension for a coarse linear match. Secondly a non-linear computational alignment algorithm can be applied: The first file is loaded as a reference file and the second as a modifiable file. The user can now select a mass accuracy and LC-accuracy for the alignment of matching peptides between the runs. This is accomplished by creating a window X Da wide and Y seconds long around every peptide. If a peptide from the other file falls within this window it is considered a match. Then the retention time of this peptide and the retention time difference between the two peptides (delta retention time) are plotted in the alignment graph as a blue dot. When this has been done with all peptides in the dataset the time dimension is divided into bins, or time intervals. The size of these bins can be selected by the user and the smaller the bins, the finer the alignment. All peptides present in the same bin get their delta retention time averaged to give one data point. This is displayed as a red connected line in the alignment graph. Due to

the nature of feature based LC-alignment there will be some empty bins and there will also be peptides matched with wrong or multiple peptides in the other file. Smoothing of this data is therefore advantageous. Smoothing is optional and uses a moving average algorithm where the user can select the number of smoothing points and immediately view the result. When the user is satisfied with tuning the alignment parameters, the peptide retention times in the modifiable file is transformed by adding a corresponding alignment value derived from the alignment graph (positive or negative). This causes the peptides in the Graphical plot to shift accordingly. The validation of the alignment is done visually and if additional alignment is required a new iteration can be applied.

2) Global PTM assessment

The one or two files loaded can be interrogated for mass differences (delta masses) in a global blind fashion using the global PTM-finder function in POSTMan. All mass and retention time differences between all peptides are calculated and the frequency of every mass or every 20 seconds, respectively, are counted. The results are displayed as two histograms, one for delta masses and one for delta retention times. If there is an overrepresentation of one mass difference, exemplar 42 Da: The presence of acetylation or tri-methylation events can be expected. Table 3-1 shows a list of common PTMs and their respective mass.

PTM	*Monoisotopic mass (Da)*
Acetylation	42.010565
Propionylation	56.026215
Phosphorylation	79.966331
Tri-methylation	42.046950
Oxidation	15.994915

Table 3-1. The table shows a list of common PTMs and their respective monoisotopic mass. The numbers are gathered from the Unimod database of protein modifications for mass spectrometry (www.unimod.org).

The peptide decimal mass (digits after the integer value of the mass) will increase from .0000 to .9999 when the peptide mass increase (Piening et al, 2006). When the peptide mass increase the variance in decimal mass also increase and there is therefore a defined mass defect any peptide can possess. Mass differences between low mass peptides will distribute around a whole number with a small bandwidth

while high mass peptides will have large bandwidth due to the increased variance in decimal mass. Due to this fact, mass differences between random peptides will distribute around a whole number as a Gaussian distribution, and when the frequency is counted there will be an oscillating noise per whole mass due to these random matches. To estimate this patterned noise the median of all data in the histogram is calculated and a sine filter with amplitude three times the median and phase such that a maxima occurs at every whole mass and a minima every half mass is drawn. Mass differences with a frequency above this filter are likely to be overrepresented and should be investigated further. At this time of POSTMan development, the noise is not subtracted, only visualised in the delta mass plot.

3) Detection of peptide pairs

a. Pair-matching

Matching of peptides between LC-runs is the most essential function in POSTMan and the foundation for PTM detection. The user can select whether to match peptides with equal mass and retention time between runs, or peptides with a defined mass difference corresponding to a PTM of interest. During calculations the software creates a virtual window shifted X Da from every peptide (where X is the desired mass difference), z Da wide (where z is the matching threshold and reflects the accuracy of the mass spectrometer used) and y1 to y2 seconds long (where y1 and y2 are the desired retention time range in which the matching peptide should reside. This also reflects the reproducibility of the HPLC used). If a peptide from the other file falls within this window it is considered a match and displayed in the Graphical plot with a line connecting the two peptides. The two peptides are also visible as a table view where they are listed as a pair and the mass and time differences are shown for manual validation. It is important to notice that all calculations for pair-matching are performed on de-convoluted mass values provided by msInspect. This is to ensure that peptides having different charge states in the two runs still can be matched as a pair. Exemplar is an acetylated peptide: When deacetylated, the basic lysine is free to carry an extra charge and the peptide m/z may change, but not the de-convoluted mass value.

b. No-pairing and detection of unique peptides

There might be cases where only one version of a peptide, either modified or unmodified, exists in the two datasets. One example of this are small, multiply charged and multiply modified peptides, for example acetylated histone peptides. When these peptides are deacetylated they can readily carry several more charges so that the peptide m/z will shift outside the mass spectrometers lower detection limit. In these cases only the modified peptide is present in the untreated sample whereas the unmodified is not present in the treated sample. To detect these peptides POSTMan can be set to locate all unique peptides (that is peptides that is only present in one of the two files) instead of matching pairs. Incorporated is also a function to do no-pairing. This is used to apply filters to the dataset without making pairs. This is especially useful for global PTM assessment, where the data can be filtered by intensity and matched to a FASTA file before subsequent histogram analysis.

4) Filtering

A number of optional filters are available in POSTMan for filtering prior and after pair-matching to limit the number of false positive matches: *Charge state cut-off filter, relative intensity filter, intensity cut-off filter, direction of chromatographic shift filters* and *FASTA filter.* Some filters have been placed before pair-matching to reduce the number of peptides for these calculations and thus increase the overall analysis speed.

i) Charge state cut-off filter

This filter removes all peptides with charge state below a set value prior to pair-matching. In ESI instruments peptides usually carry multiple charges whereas contaminants usually have one charge. By setting this value to one these contaminants are removed from calculations prior to pair-matching. However, for alignment purposes these contaminants can be useful due to their high intensity and consistency in the LC-runs. For inclusion of contaminants this filter can be set to zero. During peak picking and feature finding in msInspect some features are assigned the charge state zero. These are called “stray ions” and usually consist of an ion spike without the isotopic envelope. By setting this filter to zero, these stray ions are removed.

ii) Relative intensity filter

This filter disrupts a matched pair if the intensity difference between the two peptides is too high. If set to the value 10, peptides with a 10-fold difference can still make a pair. This filter can be used when similar peptide intensities between runs is expected. However, due to ion suppression during the ionization process this filter should be used with caution.

iii) Intensity cut-off filter

This filter is a flat background noise filter and is used to remove peptides with low intensity prior to pair-matching. It can also be used for creating a subset of the data, exemplar displaying only peptides with intensity over 200 counts, for alignment or global PTM assessment.

iv) Direction of chromatographic shift filters

These filters are useful when the chromatographic effect of the PTM of interest is known. We have seen that acetylated peptides that lose their acetyl group tend to elute earlier in reversed phase chromatography. This has also shown to be the case with phosphopeptides. Oxidized peptides, on the other hand, elutes several minutes later than its non-oxidized counterpart. By using these filters to restrict the pair-matching to only hydrophilic or hydrophobic chromatographic shifts the number of false matches can be utterly reduced.

v) FASTA filter

The FASTA function in POSTMan allows a single protein FASTA file to be loaded into POSTMan and in silico protease digested. Parameters for digestion are: Protease specificity, fixed chemical modification on cysteines, number of missed cleavages and a set of common PTMs plus a custom PTM mass. This FASTA filter removes all peptides that do not match the theoretical masses of peptides belonging to this specific protein prior to pair-matching. This is a powerful filter to isolate a single protein for analysis in the context of a complex sample, or to remove background interfering peptides in a known protein sample.

5) Inclusion list output

The output from POSTMan is a text file, either comma- or tab-separated, containing peptide properties of choice: m/z, retention time, charge, de-convoluted mass value, mass difference, retention time difference and peptide sequence from FASTA file.

This file can be used as the basis for peptide inclusion on any kind of mass spectrometer, but requires manual input. A second file can also be compiled from POSTMan: An inclusion file compatible with MassLynx 4.0 for direct loading into Waters mass spectrometers operated by this software.

3.2 Cell culture

3.2.1 General cell culture

All work with cell culture was performed in sterile environment inside a laminar flow cell culture bench equipped with a High Efficiency Particulate Air (HEPA) filter. Incubation of cells was performed in a humidified CO_2-incubator (Forma) operated at 37 °C and 5% CO_2. For visualisation of cells an Olympus microscope was used to validate cell viability and morphology every time the culture was split. Counting of cells were performed by loading 7.5 µL cell suspension into two separate spaces in a manual cell counting chamber (Bürker) and two squares were counted in each space. The number of cells counted was multiplied by 10^4 and the mean of the four rectangles counted was used as the nominal cell count. The cell lines used in this thesis are listed in Table 3-2.

Cell line	*Split ratio / frequency*	*Medium*	*Characteristics*
HeLa	1:5 / 2 days	RPMI with 10% FBS, 1% P/S	Adherent
HEK293T	1:5 / 2 days	DMEM with 10% FBS, 1% P/S	Adherent

Table 3-2. Cell lines used in cell culture experiments

HeLa cells were split at 70-80% confluence: Washed one time with PBS prior to splitting and incubated with trypsin/EDTA for one minute. Trypsin solution was removed and fresh medium was added and dish tapped to loosen cells. This cell suspension was then split 1:5 into a new dish. HEK293T cells were split at 70-80% confluence: Washed carefully one time with PBS prior to splitting and cells were then loosened by adding fresh medium vigorously to make a cell suspension. This cell suspension was then split 1:5 into a new dish.

3.2.2 Cryostorage of cells and thawing of cells

For long time storage cells were washed once with PBS, detached with trypsin and resuspended in sterile filtered 10% Dimethyl sulfoxide (DMSO), 20% FBS and appropriate

medium. Cells were frozen at concentration 5-6 million per tube (1.8 mL). Tubes were placed for one day at -80 °C in a box containing 2-propanol for slow temperature decrease before being transferred to liquid nitrogen (-196 °C) for long time storage. When an aliquot was requested, cells were thawed rapidly on a 37 °C water bath and transferred to fresh appropriate medium. Medium was changed next day to remove any DMSO present from the storage medium. Cells were then monitored daily for recovery after cryostorage and passed at 70-80% confluence.

3.2.3 Transfection of cell lines and harvesting of cells

HeLa cells were transfected using the transfection reagent Lipofectamine 2000: On the day of transfection the cells were at 80% confluence in a 10 cm dish. 5 µg of pCMVβ-p300 was mixed with 500 µL Optimem while 10 µL of Lipofectamine was mixed with 500 µL Optimem. The two solutions were incubated separately for 5 minutes before they were mixed and incubated again at room temperature for 20 minutes. This mixture was then added drop wise to the cells. Four hours after transfection histone deacetylace (HDAC) inhibitors were added: 240 nM Trichostatin A (TSA) and 10 mM Nicotinamide (NAM). 23 hours after transfection the cells were harvested by removing the medium, washing one time with PBS and adding 200 µL of HeLa lysis buffer (Materials, Table 2-1). The dish was scraped to harvest all cells and the cell extract was incubated 10 minutes on ice before hard centrifugation (20 minutes, 16600 *x g*). The protein amount in the supernatant was determined using Bicinchoninic acid (BCA) protein assay (Section 3.3.1) before SDS-PAGE or western blotting was performed.

HEK293T cells were transfected using the Ca_2PO_4-method: Roughly 10 million cells were seeded in a 15 cm dish so that they were 70% confluent on the day of transfection. 10 µg pcDNA 3.1 (empty vector control) or 10 µg HA-p300 was incubated with 644 µL dd.H_2O and mixed with 96 µL 2.0 M $CaCl_2$. This DNA mixture was then vigorously mixed with 750 µL 2x HEBS buffer (Materials, Table 2-1) and immediately added drop wise to the cells. Precipitate size was visually validated using the microscope. 12 hours after transfection the cells were gently washed once with PBS and new medium was added. 16 hours after transfection HDAC inhibitors were added: 1 µM TSA and 10 mM NAM. Cells were harvested 36 hours after transfection by removing the medium and loosening the cells with cold PBS to make a cell suspension. The cells were then washed once with PBS and pelleted by

centrifugation (5 minutes, 86 *x g*). All PBS was removed and small residuals were wiped away with tissue paper. 1 mL of HEK293T lysis buffer (Materials, Table 2-1) was added and lysate was left on ice for 30 minutes before hard centrifugation (20 minutes, 16600 *x g*). The protein amount in the supernatant was determined using BCA protein assay (Section 3.3.1).

3.3 Protein gel electrophoresis

3.3.1 Bicinchoninic acid protein assay

Pierce® BCA protein assay kit was used to estimate the protein concentration in samples and performed as described in the user manual with one exception: Only 1, 2 or 3 µL of sample was used. The mean of these three replicates (if none were outside the linearity of the method) were used as the protein concentration. In brief, 200 µL reaction buffer (according to the manufactures instructions) was mixed with 2 µL sample in a microplate and incubated at 37 °C for 30 minutes. Absorbance was read using a spectrometer (Packard) operating at 562 nm. As standard curve, Bovine serum albumin (BSA) was prepared at concentrations 0 µg/2 µL, 2.5 µg/2 µL, 5 µg/2 µL, 10 µg/2 µL, and 15 µg/2 µL in the same buffer as the lysates.

3.3.2 One-dimensional SDS-PAGE

Protein extracts were solubilised in SDS-PAGE sample buffer and boiled at 95 °C for 5 minutes followed by centrifugation (16600 *x g*, 4 °C) for 5 minutes and loaded with a Hamilton syringe onto pre-casted SDS-PAGE 4-12% gradient gels. The first lane on all gels was set aside for a molecular weight standard. Gel electrophoresis was performed with constant voltage (200 V) for 50 minutes at room temperature with MOPS running buffer. Gels for mass spectrometric analysis were immediately incubated in Coomassie blue staining solution (Section 3.4.1) for protein visualisation, while gels for western blotting were immediately processed for transfer (Section 3.5).

3.4 Staining of gels

3.4.1 Coomassie blue

A 0.2% stock solution of Coomassie stain was prepared as follows: One PhastGel brilliant blue tablet was dissolved in 80 mL dd.H_2O and stirred for 5 minutes. 120 mL of 100% ethanol was added and stirred until all colour was solved. The solution was then filtered through a

0.45 mm filter. The 0.1% working solution was prepared by mixing one volume stock solution with one volume 20% acetic acid. Gels were soaked in Coomassie blue stain for one hour and then destained three times one hour in 30% ethanol, 10% acetic acid or 5% ethanol, 7.5% acetic acid for overnight destaining followed by four times 15 minutes washing with dd.H2O.

3.4.2 Silver

Gels for silver staining were washed three times for 15 minutes in water followed by incubation for 120 seconds in 1.3 mM sodium thiosulfate. The sodium thiosulfate was rapidly removed by three fast washes with dd.H_2O. The gels were subsequently incubated in 11.8 mM silver nitrate at 4˚C for 30 minutes. The silver nitrate was rapidly removed by three fast washes with dd.H_2O before gels were incubated in Develop solution (Materials, Table 2-4) until protein bands appeared. The reaction was stopped by incubation in 10% acetic acid for 10 minutes followed by storage in a 1% acetic acid solution.

3.5 Western blotting

Proteins were separated on 1D SDS-PAGE 4-12% gels and transferred onto nitrocellulose membranes at 4 °C with constant voltage (14 V) for 18-20 hours in Transfer buffer containing 10% MeOH. Transfer efficiency was tested by staining the gels with Coomassie blue for the presence of remaining proteins. Membranes were quickly stained with Ponceau S for 2 minutes and washed with dd.H_2O until protein bands appeared. Depending on the experiment, the membranes were cut into desired fractions, typically two lanes in size to be able to probe for several antibodies per membrane, before being washed in TBST for ten minutes. Blocking was performed with 5% non-fat dry milk at room temperature for one hour followed by a quick rinse in TBST. Primary antibodies were diluted in TBST containing 5% BSA to the concentrations listed in Materials, Table 2-6. Primary antibody incubation was for one hour at room temperature followed by three washes for five minutes in TBST. Secondary antibodies were diluted in TBST containing 0.8% BSA to the concentration 1:10000. Secondary antibody incubation was for one hour at room temperature followed by four washes for ten minutes in TBST. Membranes were incubated in Super Signal for five minutes and images were developed using a chemiluminescence sensitive camera, LAS-3000 (Fujifilm). Protein standards on blots were visualised by taking a visible-light photo of the

membranes and overlaid with the chemiluminescence image in the graphical software package Gimp version 2.2.

3.6 Anti-propionyl antibody design and characterisation

A pan-specific anti-propionyl lysine antibody was generated commercially (SBS Genetech): Two rabbits were immunised with the peptide PrKGGKGGPrKGGPrKGGKGGPrKGG (PrK = propionyl lysine) linked to Keyhole limpet hemocyanin (KLH) which is strongly immunogenic. The peptide sequence was selected with every second lysine propionylated (every one lysine being propionylated would render a too hydrophobic peptide confounding water solubility) and glycines (no side chain structure) as spacer residues between the lysines. Rabbit serum was collected separately as pre-immunization, first bleed (4 weeks), second bleed (8 weeks), and third bleed (10 week). The titer was 25000:1. The serum from the third bleed was used for antibody characterisation, labelled 7-1 for the first rabbit and 7-2 for the second rabbit. Due to an initial high background signal when testing these antibodies for western blot analysis, the composition of TBST were altered to optimize removal of unspecific antibody binding. The optimal composition of TBST (that produced lowest background signal) is listed in Materials Table 2-5, and the same composition was used for all other antibodies as well.

3.7 Immunoprecipitation

3.7.1 Cross-linking of the anti-p53 DO-1 antibody

Protein G agarose beads were washed in PBS four times and 100 µL packed beads were incubated with 10 µL 10x anti-p53 DO-1 antibody in 1 mL PBS containing 5 µL 5x lysis buffer for one hour at room temperature. Beads were subsequently washed three times in 0.2 M, pH 9.0 sodium tetraborate buffer and then incubated in 1 mL of the same buffer containing 20 mM dimethyl pimelimidate dihydrochloride for covalent cross-linking of lysines and pH adjusted with 50% NaOH to pH just above 9.0 using pH-paper. Beads were then left for one hour incubation on rotator at room temperature. Subsequently, beads were washed four times in 0.2M, pH 9.0 ethanolamine and incubated overnight in 1 mL of this buffer at 4 °C on rotator. Next day the beads were washed four times with PBS and stored in fridge (4 °C). Every time an aliquot was to be used, the beads were washed three times with PBS to remove any loose antibody.

3.7.2 Immunoprecipitation of p53

To reduce the number of unspecific proteins binding to the antibody or the protein G agarose beads a pre-clear was performed: 10 µL packed beads of pre-coupled mouse IgG on Protein G agarose beads was added to one mL of cell lysate for one hour at 4 °C. Pre-clear was only performed on samples for mass spectrometric analysis. Samples for subsequent western blotting were not pre-cleared. Cross-linked DO-1 beads were washed three times in PBS to remove any unbound antibody before 10 µL of these packed beads were added to one mL lysed cell extract. Immunoprecipitation was then performed at 4 °C overnight before the beads were washed one time with 1x HEK293T lysis buffer (Materials, Table 2-7). Another mL of 1x HEK293T lysis buffer was added to the beads and everything was transferred to a new eppendorf tube to remove unspecific proteins bound to the plastic. The last drops of liquid were removed with a Hamilton syringe before addition of SDS-PAGE sample buffer.

3.8 Induction of mass shifts

3.8.1 Chemical derivatisation of proteins

N-hydroxysuccinimide (NHS) esters were a kind gift from Pieter van der Veken, University of Antwerp. One µg of BSA or recombinant p53 was solved in 0.1 M Tris-HCl pH 9.6, 10% acetonitrile (ACN) and NHS-treated with 100x molar excess of the moles of lysines present in the protein to be derivatised.

Protein	*Amount*	*Molar mass*	*Number of lysines*	*Reagent needed*
p53	1 µg	43700 g/mol	20	46 nmol
BSA	1 µg	66000 g/mol	69	105 nmol

Table 3-3. Amount of reagent (NHS ester) needed for derivatisation of proteins

The correct amount of either *N*-hydroxysuccinimide-acetyl ester, *N*-hydroxysuccinimide-propionyl ester, or *N*-hydroxysuccinimide-buteryl ester (Table 3-3) was added and the mixture incubated at room temperature. Reaction was stopped after one hour by adding SDS-PAGE sample buffer and samples were subjected to SDS-PAGE.

3.8.2 Silent Information Regulator 2 deacetylation assay

Glycerol stocks of human recombinant deacetylase HA-tagged Silent mating type information regulation 2 homolog (SIRT1) were a kind gift from Dr. Nullin Divecha, The University of Manchester. Proteins were solubilised in 0.1 M sodium carbonate buffer pH 8.4, 5 mM $MgCl_2.6H_2O$, and 2 mM DTT containing 2 µL SIRT1 in the absence or presence of 1.5 mM β-NAD^+ and incubated for one hour at 35 °C. Samples were acidified to a final concentration of 1% trifluoroacetic acid (TFA) to stop reaction. Samples were then cleaned on C_{18} stage-tips prior to mass spectrometric analysis.

3.9 Mass spectrometry

3.9.1 In-gel digestion of proteins for mass spectrometric analysis

Bands of interest from Coomassie blue stained gels were excised into cubes of about 1 mm^3 using a scalpel and washed two times 20 minutes in Wash buffer (see Materials, Table 2-9) on shaker. Gel pieces were soaked in 100% ACN for two minutes and left for air drying. 10 mM DTT in 100 mM Ammonium bicarbonate was added to the dried gel pieces and they were incubated for 45 minutes at 56 °C on shaker to induce reduction of proteins. For cysteine alkylation gel pieces were cooled by centrifugation (16600 *x g*, 4 °C, one minute), supernatant removed and immediately added 55 mM iodacetamide in 100 mM ammonium bicarbonate. Alkylation was performed for 30 minutes in the dark at room temperature. Gel pieces were washed two times 20 minutes in Wash buffer on a shaker and air dried as before and rehydrated with 30 µL 6 ng/µL Trypsin in 50 mM ammonium bicarbonate containing 2% ACN for 30 minutes on ice prior to overnight digestion at 37 °C on a shaker. The next morning the supernatant was set aside and the remaining peptides were extracted from the gel pieces by incubation with 1% TFA for 20 minutes on a shaker. This supernatant was combined with the first. Then the gel pieces were incubated with 0.1% TFA in 60% ACN for 20 minutes on shaker and supernatants combined as previously. Supernatants were then placed in a vacuum concentrator to remove the organic content by vaporisation. Samples were finally subject for Sir2 deacetylation assay (Section 3.8.2) prior to POSTMan analysis.

3.9.2 Peptide clean-up on C_{18} stage-tips and MALDI spotting

A Hamilton syringe needle was used to plunge out reversed phase C_{18} material (3M): For QToF analysis gauge size 16/2"/3 was used and the material placed into a D10 pipette tip. For MALDI analysis gauge size 22/2"/3 was used and the material placed into a gel loader tip. The material was then activated by addition of 100% MeOH and further wetted by 60% ACN, 0.1% TFA using a disposable syringe to apply pressure and thus generate a slow liquid flow through the C_{18} material. The C_{18} material was further equilibrated in the same way with 1% TFA and samples were applied. After a slow-flow binding of peptides passing through the column, the column was washed with 0.1% TFA and air dried. Elution of peptides for QToF analysis: Peptides were eluted into eppendorf tubes with 100% ACN followed by vaporisation in a vacuum centrifugator. Peptides were then solubilised in 2% ACN in 0.2% formic acid (FA) for loading onto nanoHPLC coupled mass spectrometry. Elution of peptides for MALDI analysis: Peptides were eluted with 1.5 µL 6 mg/mL matrix (α-cyanno-4-hydroxycinnamic acid) solved in 60% ACN, 15% MeOH, 0.1% TFA straight onto a MALDI target plate. In a neighbouring spot a mass calibrator (Peptide calibration standard II) was placed for external calibration of the MALDI mass spectrometer. The target plate was left for air drying for five minutes before loaded into the MALDI instrument.

3.9.3 MALDI-ToF-ToF

A MALDI-ToF-ToF (Bruker Ultraflex II) operating in positive ion reflectron mode and equipped with a 337 nm N_2-laser was used for validation of peptide derivatisation induced mass shifts and SIRT1 induced mass shifts. After spotting onto the MALDI target plate and loading into the instrument, the standard peptide calibration II containing Bradykinin 1-7 ($[M+H]^+$ 757.3992 Da), Angiotensin II ($[M+H]^+$ 1046.5420 Da), Angiotensin I ($[M+H]^+$ 1296.6853 Da), Substance P ($[M+H]^+$ 1347.7361 Da), Bombesin ($[M+H]^+$ 1619.8230 Da), Renin Substrate ($[M+H]^+$ 1758.9326 Da), ACTH clip 1-17 ($[M+H]^+$ 2093.0868 Da), ACTH clip 18-39 ($[M+H]^+$ 2465.1990 Da) and Somatostatin 28 ($[M+H]^+$ 3147.4714 Da) was used for external calibration of the instrument, though only for adjacent samples on the target. MS_1 spectra of the samples were obtained by combining individual spectra from 600 laser shots and calibrated against the adjacent calibrator spot. Generation of figures from the obtained spectra was performed using the open source mass spectrometry tool mMass version 2.4.

3.9.4 nanoHPLC-ESI-QToF

A QToF Ultima Global (Waters) operating in positive ion mode was equipped with an Ultimate 3000 HPLC (Dionex) in front for peptide separation and subsequent fragmentation and detection in the mass spectrometer. The samples were injected onto a C_{18} trapping column in isocratic conditions (2% ACN, 0.2% FA) delivered by the loading pump operating at 25 µL/min. After three minutes the flow from the nanopump (300 nL/min) was switched in-line with the trapping column and allowed for trap column back flushing for highest peptide sensitivity. The nanopump delivered a biphasic gradient from 5% to 95% B over 44 minutes followed by 95% B for 9 minutes for complete peptide elution. Solution A was 2% ACN in 0.1% FA and solution B was 90% ACN in 0.1% FA. The QToF was operated in MS_1 mode for POSTMan analysis and in DDA mode with inclusion list generated from POSTMan for targeted peptide fragmentation (MS/MS). The scan time in MS_1 mode was 3.9 seconds to ensure high sensitivity on parent ions and scan time in DDA mode was up to 10 seconds to generate high-quality fragmentation spectra. MS/MS spectra were background subtracted, smoothed and centroided using the ProteinLynx 2.2 software with default settings, and peak lists for database search were generated.

3.9.5 Protein database interrogation

For identification of peptides the search engine Mascot (Perkins et al, 1999) was utilized with tryptic specificity and three to five missed cleavages allowed. Carbamidomethyl cysteine was set as fixed structural modification and variable structural modifications as follows: Oxidation of methionines, deamidation of glutamic and aspartic acid, acetylation and propionylation of lysines, and phosphorylation of serines and threonines where applicable. Peptide accuracy tolerance window was mainly set to 75 ppm on precursor ions and 0.5 Da on fragmentation ions, but for one particular experiment where the instrument calibration was poor an MS_1 accuracy of 200 ppm was allowed. The protein database used was the non-redundant Swiss-Prot version 56.5.

4 Results

The results in this master thesis are presented as two major parts. The first part (4.1) presents results from the development of the mass spectrometry software POSTMan. This includes the background principles incorporating mass and time shifts of peptides and demonstrates the basic features of the software using a model protein. In addition, strengths and limitations of the software are presented when the software is applied to real biological samples. The second part (4.2) focuses on the identification of PTMs residing on the tumour suppressor protein p53, specifically acetylation and propionylation. The characterisation of a novel pan-specific anti-propionyllysine antibody is covered here and in addition we make a comparison between PTM detection by western blotting and a bioinformatics approach using POSTMan leading to a number of conclusions.

Results from Section 4-1, together with additional data not presented in this master thesis, have been accepted for publication in the journal Proteomics (2009).

4.1 Detection of PTMs on p53 and other proteins using POSTMan

4.1.1 Chemical derivatisation of peptides induce a mass difference and shift in retention time readily detected by HPLC-coupled mass spectrometry

A prerequisite for POSTMan to successfully detect PTMs is the ability to induce a mass shift in a subset of the peptides under analysis. It is this mass shift, preferentially together with a shift in retention time, which is the detection tag in POSTMan. Such a mass shift can be induced by chemical or enzymatic methods which specifically alter the mass of peptides habouring a defined PTM. Two PTMs were targeted for analysis in these studies: Lysine acetylation, causing a peptide mass shift of 42.01056 Da, and lysine propionylation, causing a mass shift of 56.02622 Da. For the study of acetylated and propionylated peptides two strategies were used to induce mass shifts: The first was a chemical derivatisation using highly reactant *N*-hydroxysuccinimide (NHS) esters of either acetic or propionic acid to chemically label lysines with acetyl and propionyl groups, respectively. The second was enzymatic removal of these groups accomplished by the deacetylase SIRT1, which also showed depropionylase activity at least at the peptide level. Initial experiments were performed with bovine serum albumin (BSA) as substrate. BSA is a protein often used in

mass spectrometry: Tryptic peptides are numerous, it is cheap and readily available. It was therefore natural to establish the chemical experiments with BSA as a substrate before progressing to other proteins like p53.

4.1.1.1 Derivatisation of BSA with NHS-compounds

BSA was derivatised in the absence (control) or presence of 100x molar excess of acetyl-NHS or propionyl-NHS esters. The treated proteins were then separated by SDS-PAGE which resulted in different molecular weights (Mw) for BSA depending on the reagent used (Figure 4-1). Underivatised BSA migrated with an apparent Mw of 67 kDa (Figure 4-1, lane 1), acetylated BSA resulted in an apparent increase in Mw (Figure 4-1, lane 2) compared with the underivatised sample, while propionylated BSA ran with an intermediate Mw (Figure 4-1, lane 3). SDS-PAGE separated underivatised from derivatised BSA, and ensured that the protein sample used in downstream chemistries was free of contaminating underivatised BSA species.

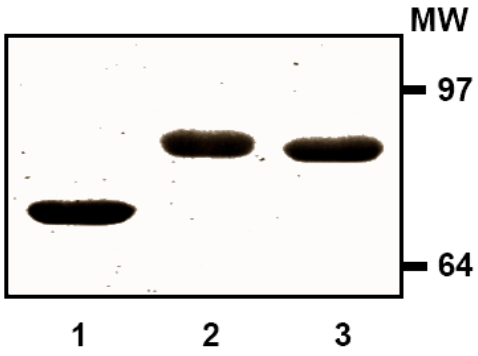

Figure 4-1. Coomassie blue stained SDS-PAGE showing underivatised BSA (lane 1), BSA derivatised with acetyl-NHS (Lane 2) and propionyl-NHS (Lane 3). Protein bands were excised from the gel and in-gel digested with trypsin.

These gel bands were excised and following trypsin cleavage of derivatised and non-derivatised BSA, several peptides were detected using MALDI-ToF mass spectrometry. A number of unique peptides were observed in the derivatised samples (Table 4-1). These peptides all contained internal lysine residues within their sequence.

BSA peptide sequence	*Theoretical mass (Da)*	*Observed mass (Da)*			
		AcBSA	*n x Ac*	*PrBSA*	*n x Pr*
SLG**K**VGTR	816.49	858.44	1	872.55	1
E**K**VLTSSAR	989.55	1031.55	1	*nd*	
AL**K**AWSVAR	1000.59	1042.59	1	1056.69	1
CASIQ**K**FGER, (1xCam)	1194.58	1236.61	1	1250.73	1
KVPQVSTPTLVEVSR	1638.93	1680.98	1	1695.06	1
LRCASIQ**K**FGERALK, (1xCam)	1775.98	1817.92	1	*nd*	
LCVLHE**K**TPVSE**K**VT**K**CCTESLVNR, (3xCam)	2986.5	3112.58	3	3154.69	3
EYEATLEECCA**K**DDPHACYSTVFD**K**LK, (3xCam)	3278.42	3320.66	1	*nd*	

Table 4-1. Internal lysine containing peptides uniquely observed when BSA was derivatised. BSA was derivatised with either acetyl-NHS or propionyl-NHS esters before in-gel trypsin cleavage. Several peptides unique to acetylated BSA (AcBSA) and propionylated BSA (PrBSA) could be detected using MALDI-ToF mass spectrometry. All unique peptides contained internal lysine residues (highlighted) and were detected in their acetylated or propionylated form respectively. Cam = carbamidomethyl-cysteine, nd = not detected.

Figure 4-2 shows the MALDI spectrum in the m/z range between 1000 and 1070 highlighting the effect of derivatisation on BSA. Underivatised BSA resulted in few prominent peptides within this range (Figure 4-2, blue spectrum). However tryptic cleavage of derivatised BSA resulted in intense ions at $[M+H]^+$ 1043.59 Da (acetylated-BSA Figure 4-2, green spectrum) and $[M+H]^+$ 1057.69 Da (propionylated-BSA Figure 4-2, orange spectrum). Sequence analysis resulted in these peptide ions matching the acetylated and propionylated mass values of the peptide ALKAWSVAR with $[M+H]^+$ 1001.59 Da respectively. This peptide contains one lysine residue at position 3 from the N-terminus. Normally trypsin would cleave at this position, unless the residue is blocked by a modification. The epsilon amino group of this lysine is likely to be the derivatised (acetylated or propionylated) residue within this sequence. We can exclude that the N-terminal amino group is derivatised as the protein was derivatised before exposure to the protease.

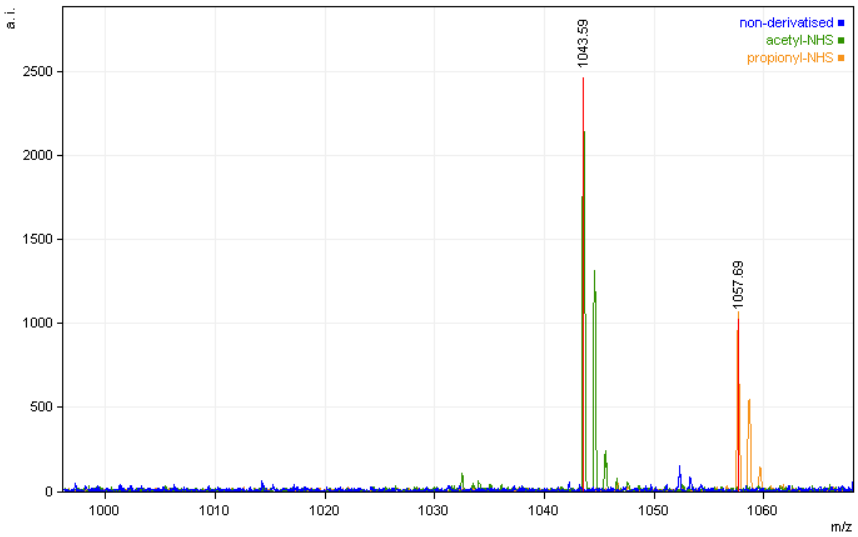

Figure 4-2. The BSA peptide ALKAWSVAR $[M+H]^+$ 1001.59 Da from derivatised and non-derivatised BSA. The blue spectrum corresponds to non-derivatised BSA: the peptide is not present. The green spectrum corresponds to acetyl-NHS derivatised BSA: the peptide is present with one acetyl group $[M+H]^+$ 1043.59 Da. The orange spectrum corresponds to propionyl-NHS derivatised BSA: the peptide is present with one propionyl group $[M+H]^+$ 1057.69 Da. The figure was generated using the open source mass spectrometry tool mMass (Strohalm et al, 2008).

4.1.1.2 The NAD dependent deacetylase SIRT1 can deacetylate and depropionylate derivatised BSA

The same BSA peptide mixture was further incubated in the presence of the deacetylase SIRT1 and its co-factor β-NAD^+. The peptide ALKAWSVAR with $[M+H]^+$ 1001.61 Da was visible in both derivatised (acetylated and propionylated) samples, whereas the labelled versions of this peptide ($[M+H]^+$ 1043.59 Da and $[M+H]^+$ 1057.69 Da for acetylated and propionylated version respectively) were not detected under these conditions (Figure 4-3, green and orange spectrum). This verifies that the deacetylase SIRT1 can remove acetyl groups (42.01 Da) and propionyl groups (56.03 Da) from peptides.

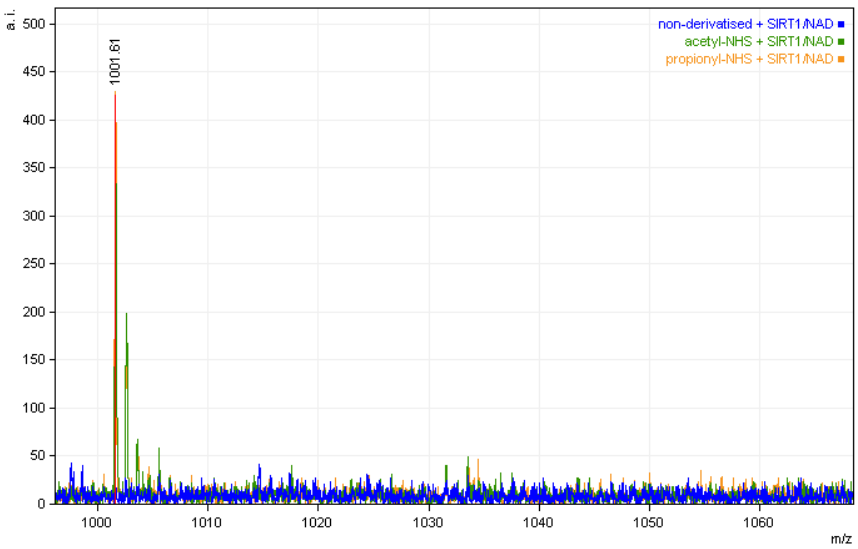

Figure 4-3. The BSA peptide ALKAWSVAR $[M+H]^+$ 1001.61 Da from NHS treated BSA and additional treatment with SIRT1. The blue spectrum corresponds to non-derivatised BSA treated with SIRT1: the peptide is not present. The green spectrum corresponds to acetyl-NHS derivatised BSA treated with SIRT1: the peptide is present without the acetyl group at $[M+H]^+$ 1001.61 Da. The orange spectrum corresponds to propionyl-NHS derivatised BSA treated with SIRT1: the peptide is present without the propionyl group at $[M+H]^+$ 1001.61 Da. The figure was generated using the open source mass spectrometry tool mMass (Strohalm et al, 2008).

4.1.1.3 Chemically acetylated and propionylated peptides have different physical characteristics to the underivatised/parent peptides.

The addition or removal of acetyl or propionyl groups changes a peptides physical properties; mass, charge state and hydrophobicity are all changed by such chemical treatment. To determine the charge state shift and hydrophobicity shift caused by addition/removal of these functional groups, HPLC-ESI mass spectrometry was used due to the presence of multiply charged peptide species generated during ESI and peptide separation based on hydrophobicity. The raw MS_1-files from acetyl-NHS derivatised BSA and acetyl-NHS derivatised BSA subsequently treated with SIRT1 for deacetylation were plotted as two dimensional maps with m/z and retention time on the X- and Y-axis respectively, and overlaid using the software MSight. Figure 4-4 shows the BSA peptide ALKAWSVAR in its acetylated and deacetylated version with retention time difference of +5.8 minutes for the acetylated version. The isotopic envelope of the peptides was also clearly visible in these plots. All acetylated peptides we encountered in this study eluted later in C_{18} chromatography than its non-acetylated counterpart.

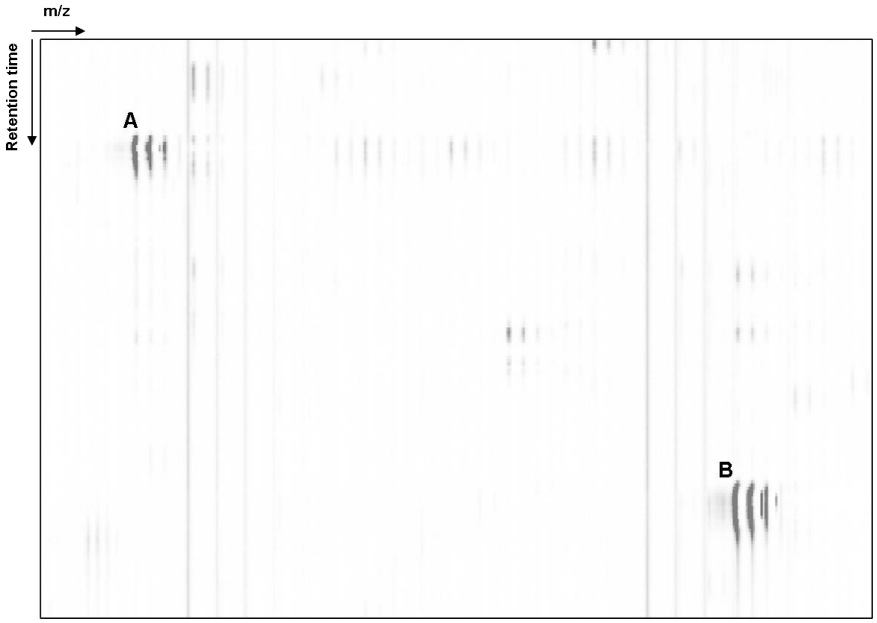

Figure 4-4. The BSA peptide ALKAWSVAR shown during elution from C_{18} HPLC-ESI mass spectrometry. MS_1-raw files from acetylated and deacetylated BSA were drawn as two dimensional maps with m/z and retention time on the X- and Y-axis respectively, and superimposed using the software MSight. The BSA peptide ALKAWSVAR is shown in its deacetylated version (A) at $[M+2H]^{2+}$ 501.29 Da and its acetylated version $[M+2H]^{2+}$ 522.29 Da. The elution time difference between the two peptides is 5.8 minutes.

When deacetylated (Figure 4-4,A), not only a shift in mass and retention time was observed, but also a reduction in intensity of this doubly charged peptide by 30% due to the presence of an additional triply charged version of the same peptide, ALKAWSVAR at $[M+3H]^{3+}$ 334.53 Da (data not shown). Deacetylation can thus also induce a shift in the peptide's charge state, most likely due to the presence of an unblocked lysine capable of carrying another charge. This raises an important issue for pairing peptides, acetylated and deacetylated, as done in POSTMan: The algorithm should have the capability to pair peptides independent of charge state.

4.1.2 POSTMan development and proof of principle

The above data demonstrate that we can specifically manipulate the acetylation and propionylation status of a set of peptides. Chemically derivatised BSA peptides can be deacetylated and depropionylated *in vitro* by the enzyme SIRT1 and we can now apply this reagent to deacetylate *in vivo* acetylated peptides. BSA was ideal as a substrate for establishing and optimising the above chemical experiments. In the following experiments

acetylated equine Cytochrome c (Sigma-Aldrich) has been used instead with the role as an *in vivo* acetylated model protein.

Comparing the SIRT1 treated peptide mixture with the original starting material i.e. with intact acetylated peptides, should reveal differences in peptide patterns in only those peptides which have been deacetylated. Based on this we developed POSTMan (described in Methods) as a tool that can automatically detect this shift in mass and retention time between two samples and create a list of candidate acetylated peptides (an inclusion file) that can be used for targeted MS/MS in a subsequent analysis. During the course of development of the software, we realised that not only acetylation and propionylation could be detected this way, but in principle any kind of PTM, as long as the mass of the modified peptides could be altered. For example, the mass of phosphorylated peptides can also be specifically altered by using alkaline phosphatase to remove the phosphate group; this is however not included in this thesis but briefly touched upon in the Discussion-section.

4.1.2.1 POSTMan workflow

The typical workflow of a POSTMan experiment includes either gel or chromatography based separation of proteins followed by proteolytic digestion of the proteins. The samples are then split in two: One kept as an untreated control and the other treated to specifically affect peptides carrying a specific PTM. The compound used can for example be an enzyme or specific chemistry that targets the PTM of interest. After HPLC-coupled ESI-MS analysis the two files are loaded into POSTMan and compared to detect mass and retention time shifts. A list of possibly modified peptides is automatically generated. Finally, these peptides are targeted for fragmentation in a second round of analysis. Figure 4-5 depicts the typical POSTMan workflow.

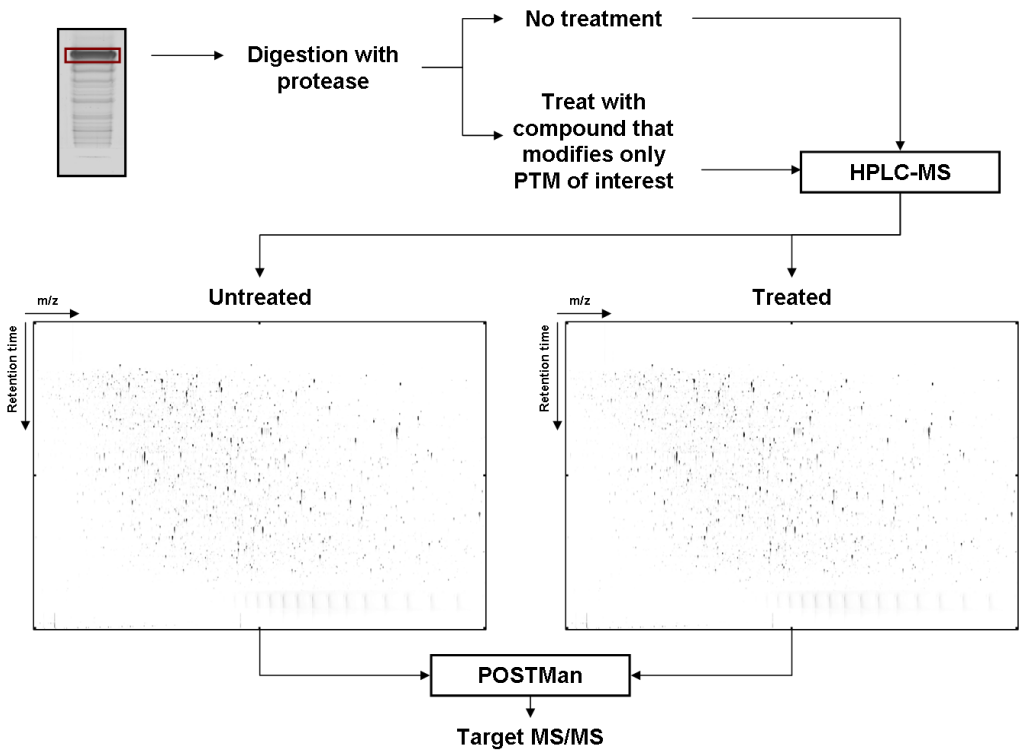

Figure 4-5. Typical workflow of a POSTMan experiment. A sample containing peptides with PTMs is proteolytically digested and split in two portions, one kept untreated and the other treated to shift mass and retention time of only the PTM-carrying peptides. Samples are analysed successively on HPLC-coupled ESI-MS operated in MS_1 mode and, after peak detection and feature finding in the POSTMan pre-processor, the files are loaded into the POSTMan analysis tool for comparison. Pairs are automatically detected and an inclusion list for targeted MS/MS is compiled. The high quality MS/MS spectra from targeted analysis confirm or disconfirm the presence of PTMs.

4.1.2.2 POSTMan can identify acetylated peptides from the *in vivo* acetylated Cytochrome c

We used acetylated equine Cytochrome c (AcCytC) as a model protein for the POSTMan development and proof of principle. This protein has previously been used to identify immonium ions specific for and characteristic of acetylated lysine residues (Kim et al, 2002). A tryptic peptide sample of AcCytC was split in two and incubated with SIRT1 in the presence or absence of β-NAD^+. After C_{18} clean-up (see Methods 3.9.2) the peptide mixtures were analysed successively by LC-MS (Q-ToF) and loaded into the POSTMan pre-processor for data conversion and feature detection. The pre-processor output was subsequently loaded into the POSTMan analysis tool and filtered so that only peptides with intensity above 50 counts were shown. These peptides were used for aligning the samples in the retention time

dimension first manually for a coarse match by sliding one run superimposed on the other for a visual best fit, and secondly automatic with bin size 50 seconds and 6 smooth points (Figure 4-6).

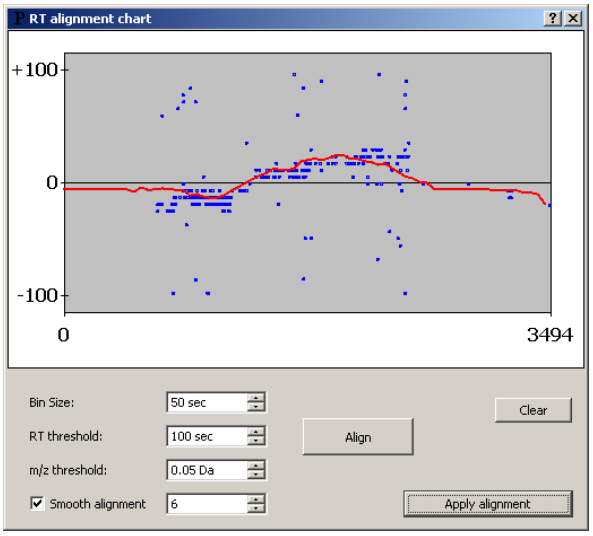

Figure 4-6. Automatic non-linear retention time alignment between the two samples AcCytC and AcCytC treated with SIRT1. The data subset used for alignment contained only peptides with intensity above 50 counts.

When "Apply alignment" was clicked, the retention time of all peptides in the treated dataset were added a value (positive or negative) according to the alignment graph (Figure 4-6, red line). Peptides with mass differences of 42.01056 Da and within a retention time window of zero to 450 seconds were paired (Figure 4-7). Peptides with intensity less than 50 counts and/or singly charged peptide ions were excluded from analysis. Furthermore, peptides with a hydrophobic shift (elutes later in C_{18} chromatography) after treatment with SIRT1 were excluded from analysis. This latter filtering step is because peptides that lose their acetyl groups tend to elute earlier than its acetylated counter peptide.

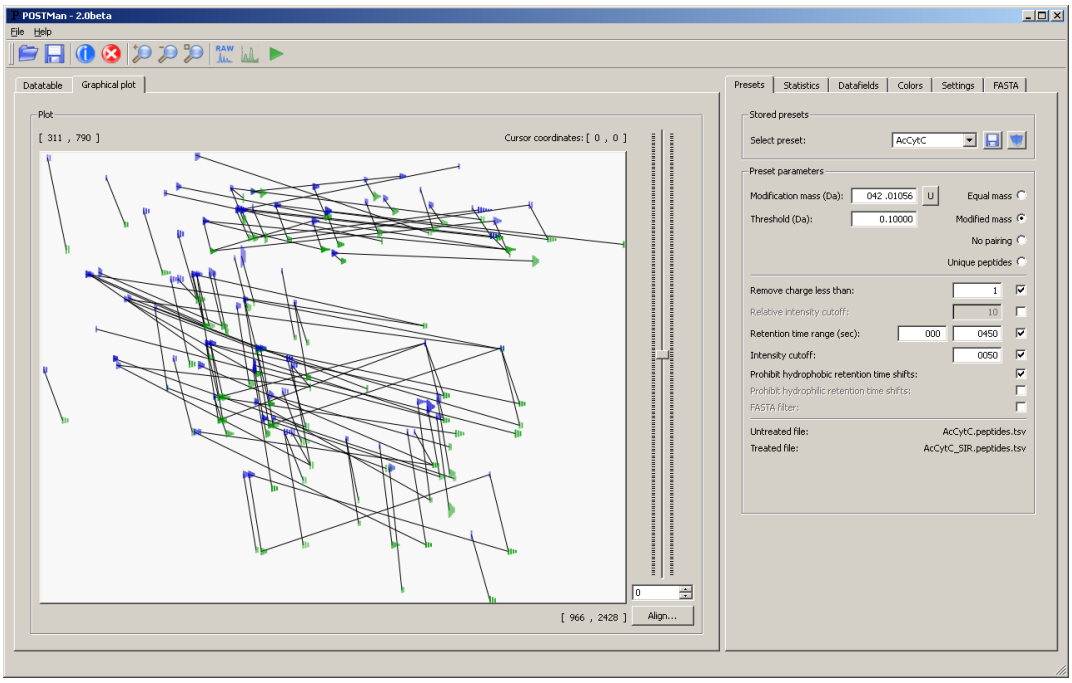

Figure 4-7. POSTMan main window showing acetylated Cytochrome c untreated (green) and after treatment with the NAD^+ dependent deacetylase SIRT1 (blue). Peptides with mass differences of 42.01056 Da and within a retention time window from 0 to 450 seconds were paired (black lines connecting peptides). Peptide species with intensity below 50 counts and/or singly charged ions were excluded from analysis. Peptide pairs with a hydrophobic shift after treatment were rejected.

This produced 132 peptide pairs and the output was saved as a MassLynx inclusion file. This file was loaded directly into the MassLynx software for targeted MS/MS analysis. 36 peptide ions were targeted successfully where 17 peptide queries identified 12 unique peptides and 9 sites of acetylation in AcCytC. A portion of the untreated sample was also analysed in conventional DDA mode for comparison where 13 peptide queries identified 10 unique peptides and the same 9 sites of acetylation. In POSTMan analysis followed by targeted MS/MS more time can be spent on every precursor ion thereby providing higher quality spectra. The quality increase of MS/MS spectra in favour of POSTMan is evident by the Mascot scores in Table 4-2.

Peptide	*Residue #*	*Site of acetylation*	*Mascot ion score*	
			DDA	*POSTMan*
HKTGPNLHGLFGR	27 - 39	28	82	82
KTGQAPGFTYTDANK	40 - 54	40	68	78
KTGQAPGFTYTDANKNK	40 - 56	40, 54	-	45
TGQAPGFTYTDANKNK	41 - 56	54	34	48
EETLMEYLENPKK	62 - 74	73	33	72, 59
YIPGTKMIFAGIK	75 - 87	80	32	72, 61
MIFAGIKK	81 - 88	87	-	33, 25
MIFAGIKKK	81 - 89	87, 88	28, 29	*nd*
MIFAGIKKKTER	81 - 92	87, 88, 89	35	*nd*
KKTEREDLIAYLK	88 - 100	88, 89	-	68
KTEREDLIAYLK	89 - 100	89	26, 46	58
KTEREDLIAYLKK	89 - 101	89, 100	-	47, 22, 30
TEREDLIAYLKK	90 - 101	100	21, 25	33
EDLIAYLKK	93 - 101	100	29	51

Table 4-2. Tryptic peptides of acetylated Cytochrome c were analysed either by DDA or using POSTMan to analyse the same sample after treatment with the NAD^+ dependant deacetylase SIRT1, combined with targeted fragmentation. Only peptides which differed by 42.01056 ± 0.1 Da were searched. This included doubly acetylated peptides where only one of the acetyl groups was removed and paired with the singly acetylated isoform e.g. peptide KTGQAPGFTYTDANKNK. DDA analysis of the sample identified 10 unique acetylated peptides and 9 sites of acetylation from a total of 13 peptide queries. POSTMan analysis identified 12 unique acetylated peptides and 9 sites of acetylation from 17 peptide queries. Targeted analysis resulted in higher Mascot scores of peptides. It is important to note that the two peptides MIFAGIKKK and MIFAGIKKKTER (nd = not determined) were not selected in the POSTMan inclusion list because they differed in mass by 84.02112 Da (2x acetylated) and 126.03168 Da (3x acetylated) respectively and not the defined mass of 42.01056 Da used in this search. Searching for masses 84.02112 Da and 126.03168 Da however does recognise these peptides (data not shown).

Among the peptides selected for targeted MS/MS was the peptide EETLMEYLENPKK (see Table 4-2). This peptide shifted charge state from $[M+2H]^{2+}$ to $[M+3H]^{3+}$ after treatment with SIRT1 due to the removal of an acetyl group and thus the presence of a free lysine. This produced an m/z shift of 291.45 Da instead of 42.01056 Da; however POSTMan was still able to create a peptide pair utilizing the de-convoluted mass value rather than the m/z ratio for the peptides (Figure 4-8).

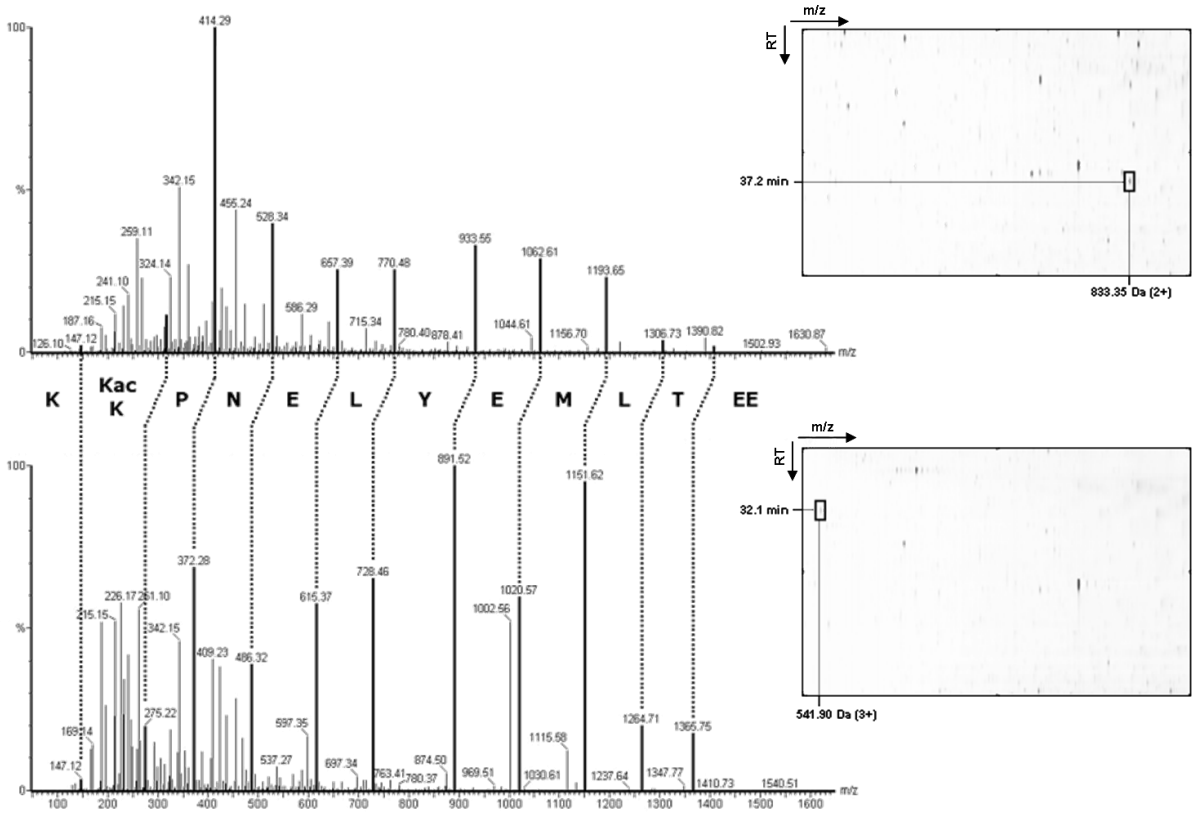

Figure 4-8. The peptide EETLMEYLENPacKK from acetylated Cytochrome c (AcCytC) shifted charge state from $[M+2H]^{2+}$ to $[M+3H]^{3+}$ after treatment with the NAD^{+} dependant deacetylase SIRT1. This produced an m/z shift of 291.45 Da and retention time difference of 5.1 minutes; far off the 42.01056 Da that was the POSTMan search parameter. Still POSTMan could make a pair of these peptides by utilizing their de-convoluted mass values instead of the m/z ratios. The y-series fragment ions from y1 to y11 were identified in both MS/MS spectra and highlighted with bold. The site of acetylation can readily be seen at Lys12 (Lys73 in AcCytC). RT = retention time.

4.1.3 Detection of acetylation sites on histones using POSTMan

In order to see if POSTMan would be able to find acetylated peptide species in the background of a cell lysate, human epithelial HeLa cells were transfected with the acetyltransferase p300 in the presence of the Histone deacetylase (HDAC) class I and II inhibitor Tricostatin A (TSA) and the HDAC class III inhibitor Nicotinamide (NAM). Lysates from both conditions were separated by SDS-PAGE and further immunoblotted with a pan-specific anti-acetyllysine antibody (ImmuneChem). This resulted in an increase in acetylation events as shown in Figure 4-9, lane 5 and 6. Five bands corresponding to acetylated proteins as detected by the antibody with similar molecular weight on the coomassie stained gel as on the western blot (see Figure 4-9, rectangles) were excised and in-gel digested with trypsin.

Peptide samples were split in two and treated with SIRT1 in the presence or absence of β-NAD^+ while a third equivalent portion of untreated material was set aside for DDA analysis. Samples were analysed by LC-MS QToF, operated in DDA mode, and by POSTMan analysis. For samples 1, 2 and 3 the peptide complexity was extremely high, which is consistent with the fact that these samples were extracted from the densest region on the gel with several proteins per molecular weight unit. No acetylated proteins could be detected in these three samples, either with DDA or POSTMan analysis. POSTMan analysis did however match several peptides with mass differences of 42.01056 Da as candidate acetylated peptides, but targeted MS/MS was unable to confirm this. This underlines that even with the use of sophisticated proteomic software, sample pre-fractionation must be performed in order to find the modified peptides present at sub-stoichiometric levels when using mass spectrometry.

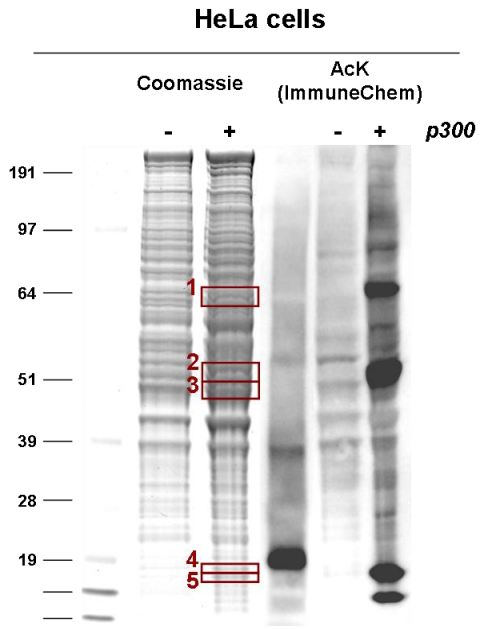

Figure 4-9. HeLa cells were transfected with the acetyltransferase p300 in the presence of histone deacetylase inhibitors Tricostatin A and Nicotinamide. An immunoblot with pan-specific anti-acetyllysine antibody (ImmuneChem) shows several proteins being acetylated in response to p300 overexpression. Coomassie stained proteins with corresponding molecular weight to the acetylated proteins detected by western analyses were excised, trypsin digested and analysed by LC-MS in DDA mode and with POSTMan.

In samples 4 and 5, the peptide complexity was lower and DDA analysis identified 6 and 7 unique proteins respectively with 2 peptides or more. Amongst these, two proteins were identified with acetylated peptides and presented here. In sample 4, Histone H3-like protein

was identified by two peptides where the doubly acetylated peptide **ac**KQLAT**ac**KAAR at $[M+2H]^{2+}$ 535.82 Da identified two sites of acetylation: Lys19 and Lys24. Furthermore, in sample 5, Histone H2B type 1 was identified by three peptides where the quadruply acetylated peptide **ac**KGS**ac**K**ac**KAVT**ac**KAQK at $[M+2H]^{2+}$ 721.43 Da identified four sites of acetylation: Lys13, Lys16, Lys17 and Lys21. POSTMan analysis of these samples did however not detect these two acetylpeptides as pairs with their deacetylated versions due to very large charge state differences. The peptide **ac**KQLAT**ac**KAAR from Histone H3-like protein, when deacetylated by SIRT1 to KQLATKAAR, could not be detected at its theoretic mass $[M+2H]^{2+}$ 493.81 Da or $[M+3H]^{3+}$ 329.54 Da which would be necessary for POSTMan to detect them as a pair. This is most likely because this peptide now can carry four charges and would ionise outside of the mass range of the mass spectrometer. This is also the case with the peptide **ac**KGS**ac**K**ac**KAVT**ac**KAQK from Histone H2B type 1: Its deacetylated peptide version KGSKKAVTKAQK expected at $[M+2H]^{2+}$ 637.41 Da, $[M+3H]^{3+}$ 425.27 Da or $[M+4H]^{4+}$ 319.21 Da was not detected. This peptide could readily carry five charges and now falls outside the range of the mass spectrometer. However, at a later stage and by applying the "find unique peptides" function in POSTMan (see Method Section 3.1.3.3b), these two peptides were recognised as potentially acetylated candidates (data not shown), but this function in POSTMan had not yet been developed at the time of analysis.

4.1.4 Detection of multiple PTMs on p53 using POSTMan

A proteomic study of acetylation and propionylation sites on the tumour suppressor protein p53 in response to p300 and SIRT1 overexpression were performed by Clive D'Santos on a Fourier Transform (FT) mass spectrometer (see appendix B). In brief, HEK293 cells were either transfected with empty vector as control (sample A), p300 (sample B) or a combination of p300 and SIRT1 (sample C). p53 was immunoprecipitated using anti-p53 DO-1 antibody and separated by SDS-PAGE. Bands corresponding to p53 were excised and in-gel digested with trypsin. The peptides were finally analysed on an FT mass spectrometer. The raw files were converted to POSTMan input files using the POSTMan pre-processor and further filtered by msInspect to only keep peptides with intensity above 90 counts (50 for sample C due to less material), charge between 2 and 5, and the minimum number of detected peaks per peptide was 2. This was done using msInspect and not POSTMan due to the extreme number of detected features from FT data and the nature of programming used

in the two programs. The samples were then subsequently analysed in POSTMan for global PTM assessment and only peptide masses that matched an in silico digest of p53 with the following parameters were considered: Tryptic enzyme specificity with 3 missed cleavages allowed, fixed structural modification on cysteine: carbamidomethyl, and variable structural modifications: Acetylation and propionylation of lysines, phosphorylation of serines and threonines, oxidation of methionines. The remaining peptides were counted in a histogram of delta masses with intervals of 0.010 Da (Figure 4-10).

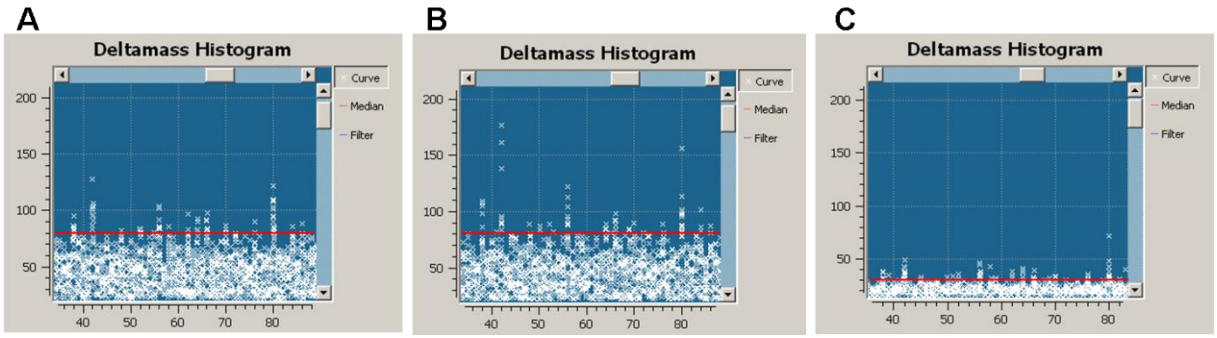

Figure 4-10. HEK293 cells were transfected with empty vector as control (A), the acetyltransferase p300 (B) and a combination of the acetyltransferase p300 and the deacetylase SIRT1 (C). The deltamasses of peptides matching an in silico digest of p53 were counted in a histogram with intervals of 0.010 Da. The random noise was estimated to 80 counts for sample A and B, and 30 counts for sample C (red lines).

The number of detected peptides with delta masses corresponding to acetylation, propionylation and phosphorylation were counted, background (noise) subtracted and normalised to the number of total peptides matching the same peptides determined by in silico digestion of p53. The linear noise estimate (a straight line connecting the sine wave maxima described in Methods Section 3.1.3.2) was 80, 80 and 30 for sample A, B and C respectively (Figure 4-10, red line). The histogram results showed a noticeable increase of 2.8% acetylation events, 1.0% propionylation events and 2.0% phosphorylation events in response to p300 overexpression (Table 4-3).

	A	*B*	*C*
p300	-	+	+
SIRT1	-	-	+
Acetylation, 42 Da	48 / 2.3%	97 / 5.1%	19 / 1.6%
Propionylation, 56 Da	24 / 1.2%	42 / 2.2%	16 / 1.4%
Phosphorylation, 80 Da	42 / 2.0%	76 / 4.0%	42 / 3.6%
Total peptides matching the in silico digest	2086	1915	1162

Table 4-3. HEK293 cells were transfected with empty vector as control (A), the acetyltransferase p300 (B) and a combination of the acetyltransferase p300 and the deacetylase SIRT1 (C). The number of detected deltamasses of peptides matching an in silico digest of p53 were counted for acetylation, propionylation and phosphorylation and background subtracted (the first number). The count was further normalised to the number of total peptides matching the same in silico digest (the second number, in percent).

When SIRT1 is overexpressed together with p300 the percentage of acetylated peptides is reduced to below background. In contrast, the percentages of propionylated and phosphorylated peptides were reduced in C compared to B but remained above background. This may suggest that SIRT1 preferentially removes acetyl groups over propionyl groups, but the comparison may also be impaired by the low number of peptides overall in sample C. The percentage of phosphorylated peptides is not the same in A, B and C. This corroborates potential crosstalk between acetylation and phosphorylation in the regulation of p53; otherwise the phosphorylation percentage would serve as a convenient control between the three conditions.

4.2 Detection of PTMs on p53 and other proteins using antibodies

4.2.1 The pan-specific anti-propionyllysine antibody characterisation

Several commercial pan-specific anti-acetyllysine antibodies are available, many of them being used routinely to detect *in vivo* acetylation in cells (Kim et al, 2006; Solomon et al, 2006). Recently, propionylation was identified as a new *in vivo* PTM (Chen et al, 2007; Cheng et al, 2008), and in the light of this we wanted to assess if the commercial anti-acetyllysine antibodies cross-reacted with propionylated lysines due to the minimal structural difference between acetyl and propionyl groups. From this point of view we generated a custom, pan-specific anti-propionyllysine antibody (see Methods, Section 3.8) and characterised this reagent in terms of its specificity for propionyl groups. The affinity of this antibody with

respect to propionyl groups was compared with commercially available acetyllysine antibodies.

4.2.1.1 Pre and post rabbit immunisation

The pan-specific anti-propionyllysine antibody was used to detect propionylation in human embryonic kidney cells, HEK293T. Cells were grown in the presence of HDAC inhibitors TSA and NAM which together inhibit all three classes of HDACs. To validate the specificity of the pan-specific anti-propionyl antibody the rabbit serum pre-immunisation was used together with anti-propionyllysine antibody, 7-1. A complete cell lysate was separated by SDS-PAGE and western blotted for comparison between the pre and post immune sera. Several bands specific for the post rabbit immunisation serum could be detected (Figure 4-11, red arrows).

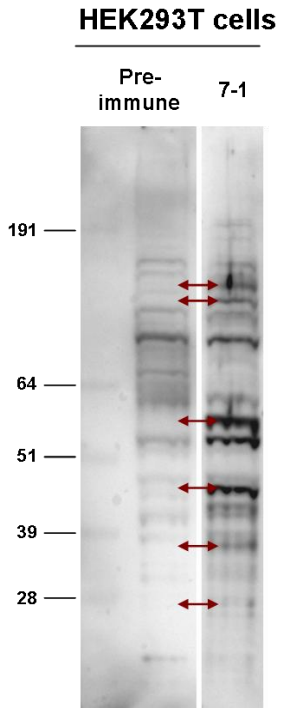

Figure 4-11. HEK293T cells were grown in the presence of the HDAC inhibitors TSA and NAM. 50 µg complete cell lysate was separated by SDS-PAGE and western blotted with sera pre- and post- rabbit immunisation to validate specificity of the anti-propionyllysine antibody 7-1. Unique bands were detected with the post immunisation serum and the most prominent are indicated with red arrows.

4.2.1.2 Antibody specificity and sensitivity

Exposed ε-amino groups of lysine residues on the recombinant tumour suppressor protein p53 (Sigma-Aldrich) were chemically derivatised with NHS-acetyl ester, NHS-propionyl ester and NHS-butyryl ester and 5 ng (114 fmol) of each of the modified proteins were separated by SDS-PAGE. Western blotting of the gel separated proteins was performed using a panel of commercial antibodies as well as two pan-specific anti-propionyllysine antibodies 7-1 and 7-2 raised in this study (see Methods, Section 3.6). A comparison of the different blots of recombinant p53, labelled and wild type (WT), is shown in Figure 4-12. For loading control 50 ng of proteins was used and the SDS-PAGE gel was stained with silver. The pan-specific commercial antibodies, with exception of the one from Upstate, showed cross-reaction with propionylated p53 and to some extent butyrylated p53. The two pan-specific anti-propionyl antibodies 7-1 and 7-2 showed cross-reaction with butyrylated p53, but no cross-reaction with acetylated p53. Upon chemical labelling all exposed lysines are expected to be labelled. It is therefore likely that the protein undergoes some conformational change due to the extensive labelling. This is also evident with the anti-p53 (DO-1) antibody. This antibody predominantly recognised WT p53 and did not recognise much of the labelled p53 variants. The anti-p53 Lys382Ac antibody should only recognise p53 when acetylated at Lys382, but showed no signal with either WT or any labelled variant. Either Lys382 was not available for chemical labelling or the antibody epitope was masked due to the heavy labelling and therefore the anti-p53 Lys382Ac antibody could not recognise it. To truly test if this antibody shows any cross-reaction to propionyl and butyryl groups on Lys382, p53 should be labelled on that particular site only, but that is not possible with this chemical approach.

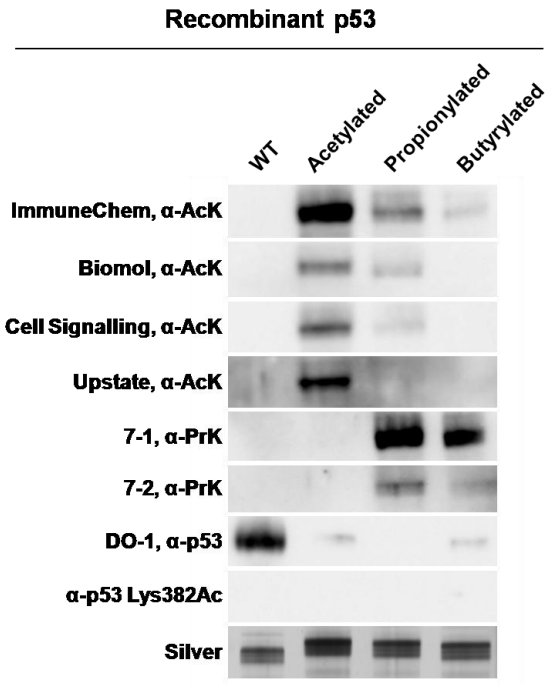

Figure 4-12. Recombinant tumour suppressor protein p53 was chemically labelled with acetyl, propionyl and butyryl groups on all exposed ε-amino groups of lysines and blotted with a panel of commercial antibodies: pan-specific anti-acetyllysine (ImmuneChem Pharmaceuticals, Biomol International, Cell Signaling Technology, Upstate Cell Signaling), anti-p53 DO-1 (Santa Cruz), anti-p53 Lys382 acetylated (Cell Signaling Technology), as well as our own two pan-specific anti-propionyllysine antibodies (7-1 and 7-2). WT = wild type, AcK = acetyllysine, PrK = propionyllysine.

All pan-specific anti-acetyllysine antibodies were used in the concentration 1:500 while the anti-propionyllysine antibodies 7-1 and 7-2 were used in 1:1000. Five ng of protein could readily be detected by the propionyl antibodies, at the same signal intensity as anti-acetyllysine from ImmuneChem. In the following experiments, only the 7-1 propionyl antibody has been used.

4.2.2 p53 immunoprecipitations

4.2.2.1 HEK293T cells overexpressing the acetyltransferase p300

HEK293T cells were transfected with either empty vector or the acetyltransferase p300. Endogenous p53 was immunoprecipitated from each of these cell samples and the immunoprecipitates were analysed by western blotting and mass spectrometry (Figure 4-13). Western blotting of the p53 immunoprecipitates with the pan-specific anti-propionyllysine antibody 7-1 resulted in an intense signal in the sample transfected with p300, but not in the

expected molecular weight range of p53. Re-blotting with anti-p53 DO-12 antibody revealed that p53 in HEK293T cells migrated in the 47 kDa area, suggesting that the signal recorded by the anti-propionyllysine antibody did not originate from p53. Further re-blotting with pan-specific anti-acetyllysine antibody showed a signal in the sample transfected with p300 in the expected p53 molecular weight range. Lys382 of p53 has been previously reported to be acetylated by p300 (Abraham et al, 2000; Gu & Roeder, 1997; Liu et al, 1999), and it would therefore be interesting to western blot these immunoprecipitates with an anti-p53 Lys382 acetylated antibody to verify if this is the case in this experiment also. In summary, western blotting showed that p53 was acetylated but not propionylated at least under these conditions.

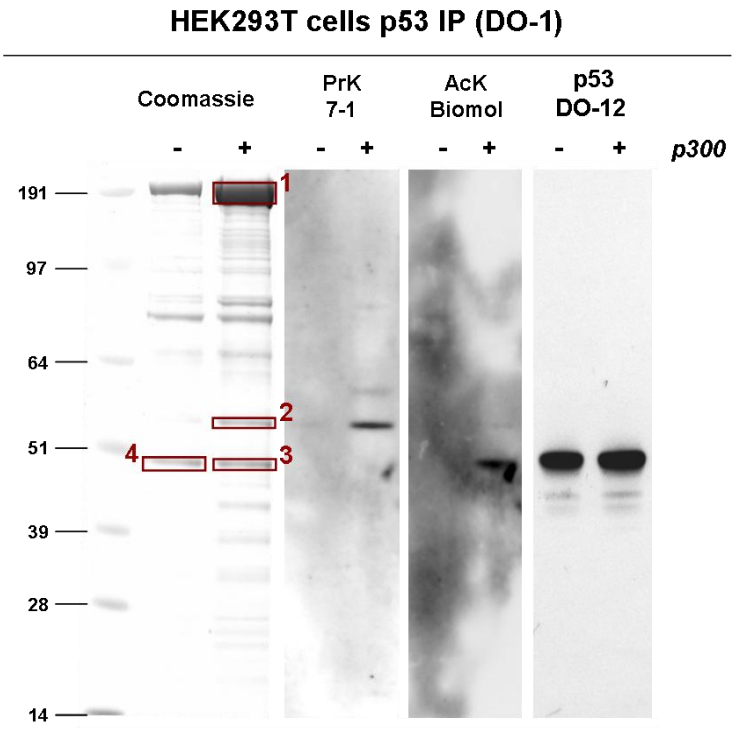

Figure 4-13. HEK293T cells were transfected with empty vector (-) or the acetyltransferase p300 (+). Endogenous p53 was immunoprecipitated (IP) and 90% of the material was separated by SDS-PAGE for mass spectrometry analysis and the remaining 10% was separated by SDS-PAGE for western immunoblotting analysis. Western blots were performed with the pan-specific anti-propionyllysine antibody (PrK, 7-1), pan-specific anti-acetyllysine antibody (AcK, Biomol International) and anti-p53 DO-12. Red numbered boxes indicate samples subjected for mass spectrometric analysis.

When the p53 immunoprecipitates were separated by SDS-PAGE, several bands were visible by coomassie staining. Two bands (1 and 2) specific for the sample transfected with p300

were excised in addition to the bands at 47 kDa (3 and 4) from both immunoprecipitates (Figure 4-13, red numbered boxes). Samples were in-gel digested and analysed by a QToF mass spectrometer operated in DDA mode for protein identification (Table 4-4). p53 was identified in samples 3 and 4 as expected compared to the western blots, but no p53-peptides identified sites of acetylation or propionylation.

Sample	*Protein Identification*	*Accession number*	*Mascot score*	*Unique peptides*	*Modifications*
1	Myosin-9	MYH9_HUMAN	4985	69	Lys1145ac, Ser1943ph
	Myosin-10	MYH10_HUMAN	3746	42	Ser1956ph
	Myosin-14	MYH14_HUMAN	1695	16	
2	Myosin-9	MYH9_HUMAN	1278	20	
	Non-POU domain-containing octamer-binding protein	NONO_HUMAN	513	9	
	Myosin-10	MYH10_HUMAN	478	6	
	UPF0027 protein C22orf28	CV028_HUMAN	399	7	
	Keratin, type II cytoskeletal 1	K2C1_HUMAN	337	4	
	Vimentin	VIME_HUMAN	129	2	
	Protein RCC2	RCC2_HUMAN	113	2	
	Aspartyl-tRNA synthetase, cytoplasmic	SYDC_HUMAN	102	2	
	Cleavage and polyadenylation specificity factor subunit 7	CPSF7_HUMAN	94	2	
3	Tumour suppressor protein p53	P53_HUMAN	541	9	
	Heterogeneous nuclear ribonucleoprotein H	HNRH1_HUMAN	274	4	
	Myosin-9	MYH9_HUMAN	246	4	
	Trifunctional enzyme subunit beta	ECHB_HUMAN	176	3	
	60S ribosomal protein L4	RL4_HUMAN	133	3	
	Elongation factor 1-gamma	EF1G_HUMAN	112	2	
	Keratin, type II cytoskeletal 1	K2C1_HUMAN	111	2	
	Alpha-enolase	ENOA_HUMAN	108	2	
	Proliferation-associated protein 2G4	PA2G4_HUMAN	95	2	
4	Tumour suppressor protein p53	P53_HUMAN	479	9	
	Alpha-enolase	ENOA_HUMAN	114	2	
	Elongation factor 1-alpha 1	EF1A1_HUMAN	95	2	

Table 4-4. Protein identification using a QToF mass spectrometer operated in DDA mode on samples from HEK293T cells transfected with empty vector or the acetyltransferase p300. Protein identifications are based on a minimum of two peptides with individual ion-score above the Mascot cut-off threshold of p=0.05. Sample numbers refer to red numbered boxes in Figure 4-13. In this particular experiment the peptide accuracy tolerance window was set to 200 ppm due to a wrong calibration in the QToF mass spectrometer.

A portion of sample 3 was also analysed by POSTMan in order to detect sites of acetylation shown by the western blot with anti-acetyllysine antibody. However, no target MS/MS spectra from a POSTMan generated output file could verify sites of acetylation. A portion of sample 2 was in the same way analysed by POSTMan in order to detect any site of propionylation, either on one of the already identified proteins, or a new protein. No sites could be verified by targeted MS/MS. This is most likely due to lower sensitivity of the mass spectrometer compared to western blotting. In addition, the fact that one rather faint band was split in four portions for: 1) Protein identification using DDA mode, 2 and 3) Untreated and treated with the deacetylase SIRT1 and analysed in MS_1 mode, and 4) Targeted analysis of candidate peptides with an inclusion list compiled from POSTMan which all limit the amount of starting material for analysis.

4.2.2.2 HEK293T cells treated with Etoposide

To elucidate the acetylation and propionylation response of p53 to the DNA damaging agent Etoposide, HEK293T cells were incubated in the presence or absence of 10 µM Etoposide for 5 hours. One hour after treatment, the HDAC inhibitors TSA and NAM were added. Cells were lysed with HEK293T lysis buffer (see Materials, Table 2-1) and endogenous p53 was immunoprecipitated using the anti-p53 DO-1 antibody. The immunoprecipitates were separated by SDS-PAGE and subjected to western blot analysis (Figure 4-14).

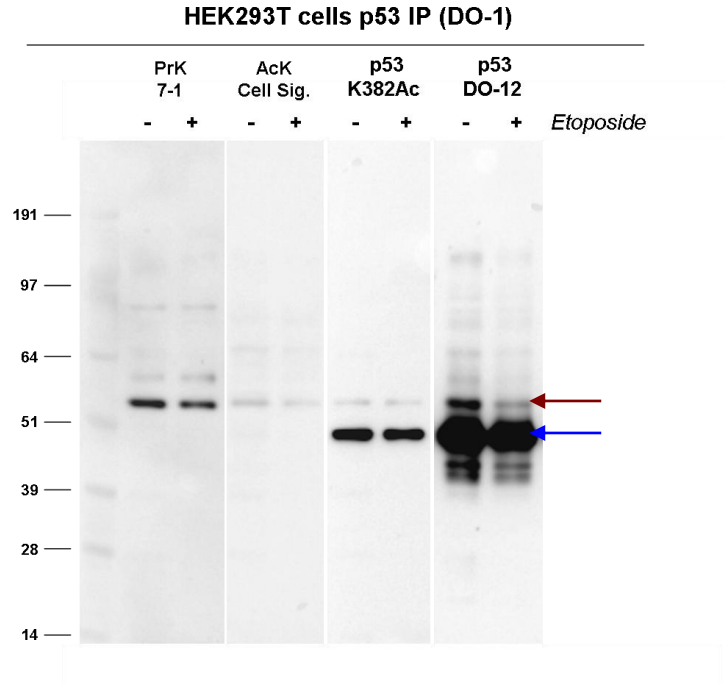

Figure 4-14. HEK293T cells were incubated in the presence or absence of the DNA damaging agent Etoposide for 5 hours. Endogenous p53 was immunoprecipitated (IP) and analysed by western blotting. Western blots were performed with the pan-specific anti-propionyllysine antibody (PrK, 7-1), pan-specific anti-acetyllysine antibody (AcK, Cell Signaling), anti-p53 Lys382 acetylated and anti-p53 DO-12. Red arrow indicates protein used for quantitation (Table 4-5) and correction for different loading amounts between the two conditions. Blue arrow indicates protein at 47 kDa (see text).

Western blotting with anti-p53 DO-12 detected a prominent band migrating below the 51 kDa (Figure 4-14, blue arrow), which showed an unequal sample loading with an increased amount in the sample from unstimulated cells. Unfortunately, no quantitation of this band on the blots could be performed due to immunoblot saturation, even with lower exposure times; otherwise the different loading could be compensated for and the correct signal be retrieved for the same band recognised by the anti-p53 Lys382Ac antibody. A signal around 53 kDa (Figure 4-14, red arrow) could be observed for all antibodies and seemed to be similar (in the same area on the gel) to that in Figure 4-13 for anti-propionyllysine. All three p53-specific antibodies (anti-p53 DO-1 used for immunoprecipitation, anti-p53 DO-12 and anti-p53 Lys382Ac) recognised this protein at a lower intensity than the one below 51 kDa, suggesting that it might be a variant of p53. This signal was not saturated and could be quantitated with the open source software ImageJ (available from http://rsb.info.nih.gov/ij/index.html) (Table 4-5). From this quantitation, when the unequal loading was corrected for, acetylation at Lys382 showed an increase by a factor of 1.7 in

response to Etoposide treatment. This corroborates previously reported data for this specific residue (Solomon et al, 2006). As for the pan-specific anti-acetyllysine antibody the increase in acetylation was by a factor of 1.5. This antibody however, is neither as specific nor sensitive as the anti- p53 Lys382Ac antibody, but points to the same conclusion; an increase in acetylation events in response to Etoposide treatment. The pan-specific anti-propionyllysine antibody also showed an increase in response to Etoposide treatment by a factor of 1.4, indicating that propionylation of p53 could be stress related (see Discussion).

Etoposide	*DO-12*	*p53 Lys382Ac*	*acK Cell.Sig.*	*prK 7.1*
-	16044	4947	6028	14958
+	9028	4863	4974	12080
+ (corrected)	16044	8642	8839	21468
ratio	1	1.7	1.5	1.4

Table 4-5. Quantitation of immunoblot signal from a protein at 53 kDa from Figure 4-14, red arrow. Immunoblot quantitation was performed with the software ImageJ using the same rectangular shape for all bands. The numbers represent the integrated density in the rectangular shape.

To elucidate acetylation and propionylation response of all proteins from the HEK293T cells to the DNA damaging agent Etoposide, the lysates after p53 immunoprecipitation were analysed by western blotting (Figure 4-15). Western blots were performed with a panel of antibodies as before: anti-actin for loading control, anti-p53 DO-12, anti-p53 Lys382 acetylated, pan-specific anti-acetyllysine and the pan-specific anti-propionyllysine antibody 7-1. The actin loading control showed equal loading and underlines the fact that immunoprecipitations can be difficult to get 1:1 (refer loading differences in Figure 4-14) even with equal amount of protein and reagents used. In response to Etoposide treatment a slight increase in p53 Lys382 acetylation can be observed, as also seen in Figure 4-14 and Table 4-5 when correcting for loading. No other visible acetylation and propionylation response to Etoposide are apparent in these blots.

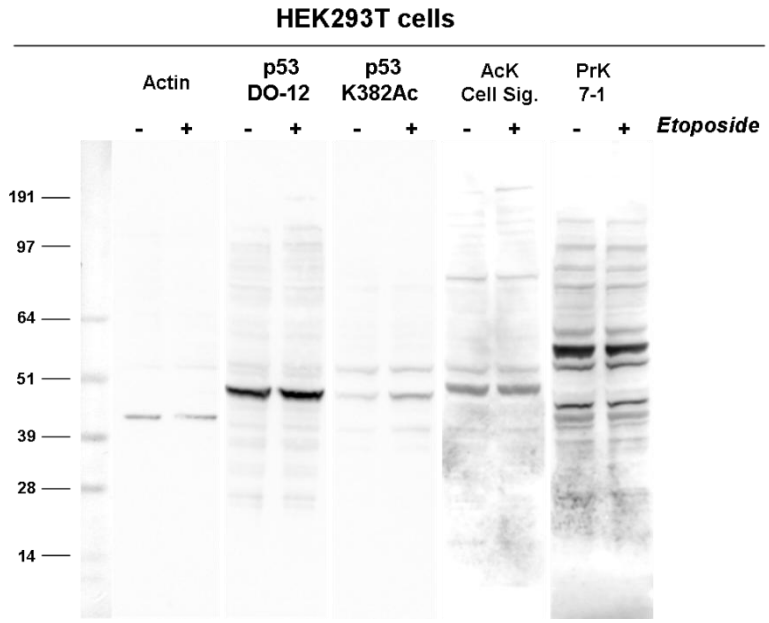

Figure 4-15. HEK293T cells grown in the presence of HDAC inhibitors were stimulated with the DNA damaging agent Etoposide for 5 hours. Endogenous p53 was immunoprecipitated and the remaining lysate was separated by SDS-PAGE followed by westen blot analysis with a panel of commercial antibodies: anti-p53 DO-12, anti-p53 Lys382 acetylated, pan-specific anti-acetyllysine (AcK, Cell Signaling) as well as the pan-specific anti-propionyllysine antibody (PrK, 7-1).

5 Discussion

Identifying and understanding the mechanisms regulating the plethora of PTMs in any given cellular system is crucial in understanding protein pathways and cellular physiology, and a systematic detection of PTMs has immense relevance to basal biology. As mentioned in the introduction, many programs for use in proteomics have been developed to contribute to PTM identification, but they currently are restricted to MS_2 data providing a sub-optimal approach for detecting PTMs. The first and main aim of this thesis was therefore to develop bioinformatics software able to enhance detection of PTMs by increasing fragmentation events and identifications of post-translationally modified peptides, utilizing MS_1 information. This will be discussed in Section 5.1.

In addition to the development of software to address PTM research from a mass spectrometry perspective, we also focused specifically on the PTMs acetylation and propionylation, and applied our software to assess these PTMs in the context of a cellular lysate and also in response to stress. To date the functional role of propionylation is not known and propionylation as an *in vivo* PTM, as well as its possible involvement in p53 regulation, was addressed and will be discussed in Section 5.2.

5.1 POSTMan as a tool in PTM detection

5.1.1 Current status and functions in POSTMan

POSTMan was developed with the specific aim of identifying post-translationally modified peptides as pairs of peptides with their unmodified counterparts. The software was designed to be as generic as possible, meaning that in principle any PTM irrespective of mass could be identified with this software. The major strength of POSTMan in comparison to other bioinformatics software used for PTM detection is that POSTMan works on MS_1 data. MS_1 data include all peptides, both possibly modified and unmodified peptides, whereas MS_2 data are traditionally limited to the most abundant fraction of the peptides, usually unmodified. Unfortunately, MS_1 data also include all contaminants and background ions, challenging the detection of PTMs. Appropriate background subtraction, peak detection, feature finding, de-isotoping and de-convolution are therefore crucial steps before analysis using POSTMan.

Furthermore, stable chromatography and/or proper LC-alignment are of enormous relevance for pairing peptides across samples, either in POSTMan for discovery of PTMs or in other applications like label-free quantitation or biomarker discovery. Distortions in peptide separation patterns between runs can lead to peptide mismatches and plenty of different approaches to remedy this have been published, as nicely reviewed in (America & Cordewener, 2008; Listgarten & Emili, 2005; Vandenbogaert et al, 2008). The method chosen for POSTMan operates on feature data and is a mixture of a manual and an automatic alignment, providing the user with control over the process. The quality of alignment of feature data is indisputably dependent on peak detection, feature finding and mass accuracy. Failure in correct assignment of feature properties, or low mass accuracy data, will impair the alignment. Using the m/z threshold setting in POSTMan the user can compensate for low mass accuracy, but this also increases the chance of matching "wrong" peptides across samples. The POSTMan automatic alignment algorithm is quite similar to the standard moving average algorithm. Instead of defining a set number of data points to use for smoothing, bins with variable number of data points are used. This means that peptide-dense regions during elution acquire the same weighting as peptide-thin regions. Whether this represents a better way of aligning the data remains to be verified, but the approach was initially chosen because of its ease of implementation. Further smoothing of the already "bin-smoothed data" was later implemented using the standard moving average algorithm. This algorithm is a simple approach where the data point ± n smooth points are averaged without weighting. For that reason the Savitzky-Golay smoothing algorithm (Savitzky & Golay, 1964) are in many cases preferred. Moving average will damage the data by removing narrow peaks, but on the other hand, LC-drifts are not likely consistent of consecutive narrow peaks, but rather a soft non-linear slope, and moving average smoothing, and/or our derivative, should therefore be applicable.

When proper pre-processing and alignment is taken into account, POSTMan can work under optimal conditions. The underlying principle of POSTMan is to compare two LC-MS runs and search for peptide pairs, either between the two individual runs or within one run. Detection of peptide differences in mass, hydrophobicity and/or charge state can be useful in biomarker discovery and, as we have applied it, in PTM analysis. In our experiments the runs to be compared were treated to induce a mass shift in a subset of the peptides within the peptide mixture. This was achieved by splitting the sample consisting of possibly

modified peptides as well as unmodified peptides, into two equal portions. One portion was thereafter treated in a way so that only the peptides with PTMs get a mass change. Any change in mass, either by the addition or subtraction of a chemical moiety also results in a change in hydrophobicity of the peptide. The mass shift can be accomplished enzymatically, for example using a deacetylase, and we took advantage of the deacetylase SIRT1 in our experiments which showed specific removal of acetyl and propionyl groups from acetylated and propionylated peptides respectively. After treatment with SIRT1 the peptides showed no residual modified peptide suggesting a complete removal of the functional groups. This chemistry served as a convenient standardisation strategy for correct identification of PTMs using POSTMan and optimisation of the software. Furthermore, removal of a specific chemical moiety can directly impact on the peptides ability to ionise during ESI. Lysine residues are excellent carriers of charge due to their basic nature, but when blocked by a PTM this ability is reduced. When the blocking PTM is removed as result of a specific treatment this ability is restored providing the peptide with more charges during ionisation. By utilizing de-convoluted mass values rather than m/z values POSTMan is still able to detect peptide pairs in these cases.

The model protein acetylated Cytochrome c (AcCytC) was used as a proof of principle. This protein has been used previously to identify acetylated sites using mass spectrometry (Kim et al, 2002). The protein comes as an acetylated protein in which over 60% of the residues are acetylated; however the specific residues are not defined. Presumably this implies that there exist a heterogeneous population of acetylated sites within this sample and we applied POSTMan to identify as many sites as possible based on the POSTMan workflow. In parallel the tryptic peptides were analysed by a DDA analysis. Nine sites (13 fragmentation events) of acetylation in AcCytC were initially identified using traditional (DDA) setup of the mass spectrometer. Using the POSTMan workflow and targeted MS/MS, the same nine sites of acetylation were detected, but with an increased number of fragmentation events of acetylated peptides (17 fragmentation events) and higher quality MS/MS spectra. These data demonstrate that POSTMan can detect true sites of acetylation, increase the number of acetylated peptides fragmented and even increase the quality, and therefore also reliability, in the MS/MS identifications.

The histogram function of POSTMan provides a new angle in PTM detection. All peptides present in one sample are compared to all peptides present in a parallel sample. All

mass differences (delta masses) and retention time differences (delta RT) are computed. If there is an overrepresentation of a specific delta mass, it is likely that this delta mass corresponds to a specific PTM, overrepresented in the dataset. An application of this could be a global PTM search, looking at all modifications at the same time. This provides unbiased search for modifications, but due to the high number of calculations and data points (including random noise) this is at present not applicable for very low abundant modifications. Modifications with high prevalence like oxidation of methionine as well as loss of ammonia can be distinguished from background (data not shown). We anticipate that further refinement of the algorithm will allow the analysis of low abundance peptides in the same manner that is presently possible with PTMs such as oxidised methionine residues. One way of accomplishing this could be to subtract the patterned noise generated by random peptide matches (see Methods, Section 3.1.3.2), but this is not implemented in the current version of POSTMan. Another approach, currently possible, is to limit the number of background peptides by filtering all data to match an in silico digest of a FASTA protein with a number of modifications allowed. This reduces the background in such a way that low abundant modifications can be observed (see Results, Section 4.1.4), but restrains the PTM detection to a predefined set of allowed modifications and only one protein examined. Under these conditions POSTMan can provide semi-quantitative information about low abundant modification events. This provides unique information about the PTM pattern of a specific protein, but it is important to stress at this point that the data being analysed is mass differences on precursor mass only, and not actual identification of modifications in MS/MS spectra. For validation of the PTMs, these peptides should be targeted for MS/MS in a subsequent run.

5.1.2 Areas for further improvement in POSTMan

The development of the software has raised a number of important issues that are currently being addressed. These areas are discussed here and provide the platform for further development of the software. The most obvious challenge of POSTMan lies in the exact same place as its major strength, the MS_1 data. This unfiltered data, consistent of peptides with multiple charge states, contaminants and background ions, represents the data foundation in POSTMan. When the number of features increases, the frequency of a specific mass difference, by random, also increase. This means in practice that the more complex the

sample, the higher the number of false-positive matches. This was most obvious in our attempts to identify acetylated peptides from HeLa cells overexpressing p300 without prior fractionation other than 1D SDS-PAGE. POSTMan analysis could recognise peptides as acetylated candidates; however we were unable to provide fragmentation data to confirm this finding. This strongly emphasises the need of pre-fractionation of complex samples even when sophisticated software is utilized.

Identification of PTMs using POSTMan requires the sample to be split in three portions: One untreated portion, one treated portion and one portion for subsequent targeted analysis by fragmentation of candidate peptides in an inclusion list. This division of sample material will of course reduce the signal of the peptides of interest and represents a trade-off. Should analysis be performed in DDA mode with hope that the modified peptides are intense enough to be selected for fragmentation? Or should the sample be divided and focus set on a subset of the peptides therein? POSTMan was never intended to replace DDA acquisitions, but rather represent an alternative tool to focus on already identified proteins and to enhance PTM detection on them. At present there is to our knowledge no alternative to this approach.

Successful targeting of peptides in subsequent analysis is a prerequisite for POSTMan analysis. This however, is not always straightforward and currently represents a challenge. Despite increasingly stable nanoflow chromatographic systems, a working LC-alignment algorithm and known issues about mass, retention time and charge state shifts, the mass spectrometer may still encounter problems performing targeted acquisitions. If the inclusion file contains co-eluting peptides or if the time set for targeted acquisition is too long, some peptides in the inclusion file will not be targeted for fragmentation. The optimal conditions for a targeted approach must be determined empirically with the mass spectrometer used and this type of analysis benefits from sensitive mass spectrometers with short duty cycle so that no peptide in the inclusion file is neglected.

5.1.3 Using POSTMan to detect phosphorylated peptides

Phosphorylation on serine and threonine residues is a labile modification often difficult to identify by mass spectrometry. In addition to the sub-stoichiometric protein modifications, sub-optimal enrichment strategies and fragmentation issues has been pointed out as reasons for poor identifications (Ulintz et al, 2009). Although POSTMan cannot improve

phosphopeptide fragmentation, it can be helpful in locating the phosphopeptides within the sample. Then, in a subsequent targeted MS/MS acquisition, the candidate phosphopeptides can spend longer time in the collision cell, or alternative collision energies can be applied, to enhance fragmentation. In recently published data (Arntzen et al, 2009) not included in this thesis, we demonstrated that targeted analysis of phosphopeptides from phenylalanine hydroxylase indeed can be achieved using POSTMan.

5.1.4 A possible use of POSTMan to differentiate between tri-methylated and acetylated peptides

High end state-of-the-art mass spectrometers are able to produce MS data with resolution around 40000-60000 in standard operation mode, whereas older instruments common to most proteomic labs, like the QToF used in this thesis, reaches only 8000–10000 in standard operation mode. The resolution required to discriminate between tri-methylation (42.046950 Da) and acetylation (42.01056 Da) is at least 27000 for a peak at 1000 Da, a number unreachable for many labs. Although high resolution instruments and immonium ion interrogation has improved the PTM differentiation (Zhang et al, 2004), POSTMan offers an additional way of determining this. In a sample containing both modified and unmodified peptides one can in many cases find both variants of a peptide, i.e. with and without the modification. In all experiments covered in this thesis the non-acetylated peptides were observed to elute earlier than their acetylated counter peptide during reverse phase chromatography. According to Black (Black & Mould, 1991), who performed calculations on hydrophobicity values for a range of modified amino acids, the hydrophobic value of acetylated lysine and tri-methylated lysine are on either side of an un-modified lysine. This implies that the retention time shifts of an acetylated peptide and a tri-methylated peptide in respect to a non-modified peptide are in opposite directions. POSTMan can filter peptide pairs based on the direction of hydrophobicity shifts, and this therefore represents a possible approach for the discrimination of these two modifications. This procedure would require the same sample to be loaded both as untreated and treated sample within POSTMan. Furthermore, to be absolutely sure which of the two PTMs it is, one could treat the sample with a deacetylase which should only produce mass shifts of 42 Da for acetylated peptides and not the tri-methylated peptides (given that the deacetylase does not possess any demethylase function).

5.1.5 POSTMan summary

The first aim of this thesis was to develop bioinformatics software able to enhance detection of post-translationally modified peptides using mass spectrometry. This includes increasing the fragmentation events of these peptides leading to more identification of peptides harbouring PTMs. In the DDA analysis of AcCytC, 13 fragmentation events originated from modified peptides, whereas 17 fragmentation events originated from modified peptides using POSTMan. In both analyses the same nine sites of acetylation were identified. POSTMan not only proved to correctly identify the modified peptides from MS_1 data, but also increased the quality of and the reliability in the MS/MS identifications. The current version of POSTMan therefore fulfils the criteria we set as the first aim of this thesis.

Although POSTMan fulfils our criteria there are a number of issues to improve upon, such as division of sample material and successful targeting of peptides. It is therefore important to stress that POSTMan was not intended to replace traditional DDA acquisitions but rather complement existing methods at PTM analysis. It has its major strengths on relatively simple protein mixtures where the identity of the protein(s) is known or can be derived quickly. This is mainly due to the powerful FASTA filtering of peptides. However, it has at present limited applications in samples which are of an extremely complex nature. Therefore, to utilize the full potential of POSTMan we have outlined a few 'good practices' that should be considered:

(i) *Experimental and technical deviations in the activity of enzyme preparations and chemistries can contribute to poor reproducibility.*

(ii) *As a rule the chromatography and analysis should be performed as a consecutive series of measurements in order to minimize technical drift at both the MS and LC level.*

(iii) *The alignment and pair-matching procedures will benefit from data derived from high mass accuracy instruments coupled with stable chromatographic systems.*

(iv) *POSTMan is ideally suited to the analysis of single proteins or relatively simple protein mixtures. We have found that analysis of complex peptide digests generally results in a high percentage of false-positive pairs, confounding subsequent targeted analysis.*

Furthermore, POSTMan is still a work in progress and developers in the proteomics community are welcome to further improve upon the software in new areas or in those

highlighted above. Since the public release of POSTMan through http://www.probe.uib.no and publication in the journal Proteomics, POSTMan has been downloaded 59 times from four continents, and we anticipate that it will be used for identification of many different PTMs ranging from small mass differences caused by citrullination to large mass differences caused by ubiquitination. In addition, due to the strengths of MS_1 data we expect that more software utilizing this MS-level information will appear in the near future.

5.2 Propionylation as an *in vivo* post-translational modification

5.2.1 Characterisation of novel pan-specific anti-propionyllysine antibodies

Since propionylation just recently was discovered to be a novel PTM (Chen et al, 2007; Cheng et al, 2008), no commercial antibodies yet exist to detect propionylated proteins. The availability of such reagents is crucial in order to interrogate individual proteins as well as complex protein mixtures for their propionylation status and how this changes in response to physiological stimuli. We wanted therefore to generate specific anti-propionyllysine antibodies to complement existing tools to unravel the functional significance of propionylation as a PTM. Due to the minimal structural difference between acetyl, propionyl and butyryl, pan-specific antibodies targeting these lysine modifications are expected to show some cross-reaction. Three of the commercially available anti-acetyllysine antibodies we tested showed cross-reaction to propionyllysine using recombinant p53 as model protein. Some of these antibodies are used routinely to identify acetylated proteins (Kim et al, 2006; Solomon et al, 2006), but with the recent data describing propionylation as a novel *in vivo* protein modification, the data described here suggest that these antibodies should now be used with a degree of caution. We established that the novel anti-propionyllysine antibodies generated here showed no cross-reaction with acetyllysine under these conditions, allowing us to use them as a tool in identifying propionylated proteins, unambiguously identifying specific bands as candidate propionylated proteins rather than acetylated proteins. However, we did observe a significant cross-reaction with butyryllysine. Butyrylation has been described in the literature as an *in vivo* modification (Chen et al, 2007), and at present we cannot exclude the possibility that we are observing butyrylated proteins.

HDACs have been suggested as proteins catalysing the removal of propionyl groups (Cheng et al, 2008; Garrity et al, 2007; Riester et al, 2004). This is itself an intriguing aspect of

dual functionality of this family of enzymes that may have important physiological consequences. In our experiments cells were therefore grown in the presence of HDAC inhibitors and probed for propionylation using these novel antibodies as well as with rabbit serum pre-immunization. These results showed several proteins specifically recognised by the post immunization serum 7-1 (see Results, Section 4.2.1.1), further verifying its use as a tool to identify propionylated proteins.

5.2.2 Modification pattern of the tumour suppressor protein p53

The second aim of this thesis was to assess the modification pattern of p53, with emphasis on acetylation and propionylation, using both a bioinformatics (POSTMan) and biochemical (immunoprecipitation/western blotting) approach. We showed, using the histogram function of POSTMan (see Results, Table 4-3), that overexpression of p300 in HEK293 cells lead to an increase of 2.8% acetylation events and 1% propionylation events on endogenous p53. p300 and CBP has shown *in vivo* propionylation of p53 as well as autoproionylation (Cheng et al, 2008). Our result corroborates this published data suggesting an *in vivo* propionyltransferase function of p300 with p53 as substrate. In the same report, Cheng and co-workers showed that depropionylation of p53 and p300 was catalysed by the NAD^+-dependent histone deacetylase (HDAC) SIRT1, a member of the sirtuin family of HDACs. In our experiments, overexpression of p300 together with SIRT1 lead to a noticeable reduction of acetylation events on p53, but not as distinct in regard to propionylation. Our data therefore corroborate Cheng's data in that depropionylation is catalysed by SIRT1, but in addition, they suggest that lysine-acetylation of p53 is a preferential substrate for SIRT1, over lysine-propionylation. Immunoprecipitations and western blots of endogenous p53 demonstrated that p300 indeed increased acetylation events on p53 (Figure 4-13, lane 6 and 7), in addition to increase propionylation events of an unknown protein about 8 kDa above p53. A certain identity of this protein could unfortunately not be verified by mass spectrometry in this experiment, but Figure 4-14 suggests that it could be a variant of p53, for example an ubiquitinated version. This would be an interesting finding, since the propionylation of this protein seems to be specific for the heavy version.

In response to genotoxic stress p53 is stabilised and depending of the severity of the lesions either DNA repair mechanisms are activated or the cell commits to apoptotic pathways. During this stabilisation of p53 key lysine residues are acetylated, for example

Lys382: This residue is an *in vivo* target for SIRT1 regulation (Vaziri et al, 2001). To date, no functional role of propionylation is known, and with the regulation of p53 by acetylation in mind we wanted to study both acetylation and propionylation of endogenous p53 in response to DNA damage. Furthermore it is clear from proteins such as p53, which are heavily modified by multiple PTMs, that there exists a protein PTM code (Sims & Reinberg, 2008). Combinations of PTMs rather than individual modifications at specific residues are most likely to be responsible for the overall regulation of the protein. This is also a potential strength in POSTMan. Initial experiments to address this utilized the DNA damaging agent Etoposide, a compound that directly binds to and inhibits DNA Topoisomerase II resulting in the accumulation of double-strand DNA breaks. Figure 4-14 and Table 4-5, when adjusted for loading amount, suggests that p53 propionylation together with acetylation increase in response to Etoposide-induced DNA damage. Western blot analysis of the lysates after p53 was immunoprecipitated (Figure 4-15) still show an Etoposide-dependent acetylation of Lys382 of p53, suggesting that not all p53 was removed by immunoprecipitation. No Etoposide-dependent propionylation however can be observed in this blot, but note that this was performed in the background of a multitude of other proteins (a cell lysate) and what was left of p53 could be below the detection limit of the pan-specific antibody. In contrast, the anti-p53 Lys328 acetylated antibody is highly specific and it is expected to pick up very low signals. But worth noting is the multitude of bands which seems to be propionylated in response to HDAC inhibitors (Figure 4-15, lane 10 and 11). This suggests that HDACs in general terms could be able to remove propionyl groups, concurrent with published data. It would be interesting to repeat these experiments and grow the cells both in the presence or absence of HDAC inhibitors, or even different classes of HDAC inhibitors, to elucidate if this truly is the case and which class of HDAC has most predominant depropionylase activity. Moreover, the propionylated proteins are not the same as those being acetylated (Figure 4-15, lane 8 and 9) raising the possibility that propionylation not only occurs instead of acetylation on the same sites, but as a distinct modification on a specific set of proteins.

Propionyl-CoA has been proposed as the likely propionyl donor for protein propionylation. *In vitro* the acetyltransferases p300, CREB-binding protein (CBP) and P/CAF were capable of propionylating histones using propionyl-CoA as substrate (Chen et al, 2007; Leemhuis et al, 2008), but whether this is also the case *in vivo* is not known nor whether

propionylation is a tightly regulated PTM with a distinct function in the cell. One hypothesis is that propionylation may reflect the cellular level of propionyl-CoA. Propionyl-CoA levels are altered by the breakdown rate of the amino acids methionine, valine and isoleucine and also the β-oxidation of odd-carbon fatty acids. Conditions that increase the pool of cellular propionyl-CoA levels, for example a high intake of dietary odd-carbon fatty acids, could result in the propionylation of target proteins such as transcription factors resulting in the transcription of specific target genes in response to these environmental changes; in this case an increase in the levels of dietary odd-carbon fatty acids. A precedent for this kind of dietary regulation exists with the Sirtuin family of deacetylases. The deacetylase activity of these enzymes is absolutely dependent on the co-factor NAD^+. NAD^+ is derived from the oxidation of fat and under conditions of starvation cellular NAD^+ levels are reduced, lowering the overall cellular activity of specific Sirtuin deacetylases. This has a direct impact on protein complexes responsible for chromatin remodelling (Sirtuins are often components of these complexes) and thus the transcription of specific genes that respond to conditions of low dietary fat intake. The long term consequences of this are an increase in longevity (ageing) in a low dietary intake environment. Indeed in model organisms (flies, worms and yeast) knockdown of specific sirtuin genes results in an increase of up to 50% in the life-span of these organisms (Chen & Guarente, 2007; Guarente & Kenyon, 2000). Along similar lines, growing cells in the presence of propionate, a short-chain fatty acid and a product of fibre fermentation in the gut, could alter the cellular level of propionyl-CoA and further effect the rate of propionylation events. This could allow us to study the specific targets and their sites of propionylation. Interestingly, the risk of developing colorectal cancer has been linked to dietary fibre intake (Bingham et al, 2003). Propionate and butyrate have been shown to inhibit HDACs, the latter being considered the most potent HDAC inhibitor. Butyrate not only inhibits HDACs leading to more acetylation, but also increases the available pool of acetyl-CoA via the β-oxidation pathway. Propionate on the other hand, is not as potent as butyrate in HDAC inhibition, but does not contribute to the acetyl-CoA pool. Both propionate and butyrate has been shown to induce hyperacetylation of histones leading to transcription of many genes including the cyclin-dependent kinase (Cdk) inhibitor p21 and cause growth arrest and differentiation in human colon carcinoma cells; pathways known to be regulated by p53 (Blottiere et al, 2003; Hinnebusch et al, 2002). Whether propionylation of p53, as a direct response to increased cellular levels of propionyl-CoA, was involved in this matter

remains to be elucidated. According to Leemhuis, histone propionylation may be a novel epigenetic mark for cellular metabolism; as acetylation, the charge neutralization by propionylation of histones may lead to changes in chromatin structure and promote gene transcription (Leemhuis et al, 2008). Our data suggests that propionylation of the non-histone protein p53 may occur in response to double-strand DNA breaks. This could point for a functional role for propionylation, contributing to the already complex regulation of p53, being a stress marker in the same way as acetylation. This could mean that propionylation on specific residues within the p53 sequence would cause a defined set of genes to be transcribed, different genes than those transcribed when p53 is acetylated, whether it is on the same residues or not.

A systematic further study of p53 propionylation is needed to validate these speculations and help answering important questions. One experiment to do would be to induce cellular stress in many different ways and screen for p53 propionylation and acetylation patterns. By stimulating cells with Etoposide, UV-radiation, gamma-radiation and H_2O_2 we could be able to set propionylation into a specific group of DNA damage responses. This would provide information on if propionylation truly is a DNA damage stress marker or not. Furthermore, stress stimuli can be induced at different time points through the cell cycle, different stages of differentiation and different organelles (for example nuclei and mitochondria) all providing useful information and helping us understanding the biological role of protein propionylation.

5.2.3 Propionylation summary

Protein propionylation as being a novel *in vivo* PTM is a completely new field of research and has an enormous potential in the years to come. A few research groups have so far focused on this subject and we anticipate that more information regarding the functional role of propionylation is to be unravelled in the near future. The second aim of this thesis was to elucidate the PTM pattern of p53 with emphasis on acetylation and propionylation. Our data indicate that propionylation of p53 may be related to genotoxic stress, but needs to be further validated with more experiments, as outlined above.

The third aim of this thesis was to raise specific anti-propionyllysine antibodies to complement existing proteomic tools to unravel the functional significance of propionylation as a global PTM. The novel antibodies generated in this study showed no cross reaction with

acetyllysine and will serve as unique tools in characterising protein propionylation in the cell. The study of low abundant protein PTMs like propionylation will require clever protein separation strategies, well designed antibodies and bioinformatics software designed for PTM characterisation as proteomic tools. Clearly these antibodies together with POSTMan have shown individually to have great potential as proteomic tools that will help in understanding the function of propionylation. This thesis has contributed with data which has opened up a completely new area of research and a number of experiments have been outlined above that will contribute to further knowledge about propionylation as a protein post-translational modification.

6 References

Abraham J, Kelly J, Thibault P, Benchimol S (2000) Post-translational modification of p53 protein in response to ionizing radiation analyzed by mass spectrometry. *J Mol Biol* **295**(4)**:** 853-864

America AH, Cordewener JH (2008) Comparative LC-MS: a landscape of peaks and valleys. *Proteomics* **8**(4)**:** 731-749

Arntzen MO, Osland CL, Raa CR, Kopperud R, Doskeland SO, Lewis AE, D'Santos CS (2009) POSTMan (POST-translational modification analysis), a software application for PTM discovery. *Proteomics* **9**(5)**:** 1400-1406

Bingham SA, Day NE, Luben R, Ferrari P, Slimani N, Norat T, Clavel-Chapelon F, Kesse E, Nieters A, Boeing H, Tjonneland A, Overvad K, Martinez C, Dorronsoro M, Gonzalez CA, Key TJ, Trichopoulou A, Naska A, Vineis P, Tumino R, Krogh V, Bueno-de-Mesquita HB, Peeters PH, Berglund G, Hallmans G, Lund E, Skeie G, Kaaks R, Riboli E (2003) Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *Lancet* **361**(9368)**:** 1496-1501

Black SD, Mould DR (1991) Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Anal Biochem* **193**(1)**:** 72-82

Blottiere HM, Buecher B, Galmiche JP, Cherbut C (2003) Molecular analysis of the effect of short-chain fatty acids on intestinal cell proliferation. *Proc Nutr Soc* **62**(1)**:** 101-106

Bonk T, Humeny A (2001) MALDI-TOF-MS analysis of protein and DNA. *Neuroscientist* **7**(1)**:** 6-12

Chen D, Guarente L (2007) SIR2: a potential target for calorie restriction mimetics. *Trends Mol Med* **13**(2)**:** 64-71

Chen Y, Sprung R, Tang Y, Ball H, Sangras B, Kim SC, Falck JR, Peng J, Gu W, Zhao Y (2007) Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol Cell Proteomics* **6**(5)**:** 812-819

Cheng Z, Tang Y, Chen Y, Kim S, Liu H, Li SS, Gu W, Zhao Y (2008) Molecular characterization of propionyllysines in non-histone proteins. *Mol Cell Proteomics*

Chernushevich IV, Loboda AV, Thomson BA (2001) An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* **36**(8)**:** 849-865

Chuikov S, Kurash JK, Wilson JR, Xiao B, Justin N, Ivanov GS, McKinney K, Tempst P, Prives C, Gamblin SJ, Barlev NA, Reinberg D (2004) Regulation of p53 activity through lysine methylation. *Nature* **432**(7015)**:** 353-360

Desterro JM, Rodriguez MS, Hay RT (2000) Regulation of transcription factors by protein degradation. *Cell Mol Life Sci* **57**(8-9)**:** 1207-1219

Garrity J, Gardner JG, Hawse W, Wolberger C, Escalante-Semerena JC (2007) N-lysine propionylation controls the activity of propionyl-CoA synthetase. *J Biol Chem* **282**(41)**:** 30239-30245

Gu W, Roeder RG (1997) Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90**(4)**:** 595-606

Guarente L, Kenyon C (2000) Genetic pathways that regulate ageing in model organisms. *Nature* **408**(6809)**:** 255-262

Hansen BT, Davey SW, Ham AJ, Liebler DC (2005) P-Mod: an algorithm and software to map modifications to peptide sequences using tandem MS data. *J Proteome Res* **4**(2)**:** 358-368

Harris SL, Levine AJ (2005) The p53 pathway: positive and negative feedback loops. *Oncogene* **24**(17)**:** 2899-2908

Hinnebusch BF, Meng S, Wu JT, Archer SY, Hodin RA (2002) The effects of short-chain fatty acids on human colon cancer cell phenotype are associated with histone hyperacetylation. *J Nutr* **132**(5)**:** 1012-1017

Johnstone RW, Ruefli AA, Lowe SW (2002) Apoptosis: a link between cancer genetics and chemotherapy. *Cell* **108**(2)**:** 153-164

Kim JY, Kim KW, Kwon HJ, Lee DW, Yoo JS (2002) Probing lysine acetylation with a modification-specific marker ion using high-performance liquid chromatography/electrospray-mass spectrometry with collision-induced dissociation. *Anal Chem* **74**(21)**:** 5443-5449

Kim SC, Sprung R, Chen Y, Xu Y, Ball H, Pei J, Cheng T, Kho Y, Xiao H, Xiao L, Grishin NV, White M, Yang XJ, Zhao Y (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* **23**(4)**:** 607-618

Lane DP (1992) Cancer. p53, guardian of the genome. *Nature* **358**(6381)**:** 15-16

Leemhuis H, Packman LC, Nightingale KP, Hollfelder F (2008) The human histone acetyltransferase P/CAF is a promiscuous histone propionyltransferase. *Chembiochem* **9**(4)**:** 499-503

Levine AJ (1997) p53, the cellular gatekeeper for growth and division. *Cell* **88**(3)**:** 323-331

Listgarten J, Emili A (2005) Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics* **4**(4)**:** 419-434

Liu L, Scolnick DM, Trievel RC, Zhang HB, Marmorstein R, Halazonetis TD, Berger SL (1999) p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage. *Mol Cell Biol* **19**(2)**:** 1202-1209

May D, Fitzgibbon M, Liu Y, Holzman T, Eng J, Kemp CJ, Whiteaker J, Paulovich A, McIntosh M (2007) A platform for accurate mass and time analyses of mass spectrometry data. *J Proteome Res* **6**(7)**:** 2685-2694

Mihara M, Erster S, Zaika A, Petrenko O, Chittenden T, Pancoska P, Moll UM (2003) p53 has a direct apoptogenic role at the mitochondria. *Molecular cell* **11**(3)**:** 577-590

Moll UM, Petrenko O (2003) The MDM2-p53 interaction. *Mol Cancer Res* **1**(14)**:** 1001-1008

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**(18)**:** 3551-3567

Piening BD, Wang P, Bangur CS, Whiteaker J, Zhang H, Feng LC, Keane JF, Eng JK, Tang H, Prakash A, McIntosh MW, Paulovich A (2006) Quality control metrics for LC-MS feature detection tools demonstrated on Saccharomyces cerevisiae proteomic profiles. *J Proteome Res* **5**(7)**:** 1527-1534

Potthast F, Gerrits B, Hakkinen J, Rutishauser D, Ahrens CH, Roschitzki B, Baerenfaller K, Munton RP, Walther P, Gehrig P, Seif P, Seeberger PH, Schlapbach R (2007) The Mass Distance Fingerprint: a statistical framework for de novo detection of predominant modifications using high-accuracy mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **854**(1-2)**:** 173-182

Riester D, Wegener D, Hildmann C, Schwienhorst A (2004) Members of the histone deacetylase superfamily differ in substrate specificity towards small synthetic substrates. *Biochem Biophys Res Commun* **324**(3)**:** 1116-1123

Riley KJ, Maher LJ, 3rd (2007) p53 RNA interactions: new clues in an old mystery. *RNA* **13**(11)**:** 1825-1833

Rodriguez MS, Desterro JM, Lain S, Lane DP, Hay RT (2000) Multiple C-terminal lysine residues target p53 for ubiquitin-proteasome-mediated degradation. *Mol Cell Biol* **20**(22)**:** 8458-8467

Roepstorff P, Fohlman J (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **11**(11)**:** 601

Savitski MM, Nielsen ML, Zubarev RA (2006) ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol Cell Proteomics* **5**(5)**:** 935-948

Savitzky A, Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* **36**(8)**:** 1627-1639

Schiltz RL, Mizzen CA, Vassilev A, Cook RG, Allis CD, Nakatani Y (1999) Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates. *J Biol Chem* **274**(3)**:** 1189-1192

Sims RJ, 3rd, Reinberg D (2008) Is there a code embedded in proteins that is based on post-translational modifications? *Nat Rev Mol Cell Biol* **9**(10)**:** 815-820

Smith RD, Anderson GA, Lipton MS, Pasa-Tolic L, Shen Y, Conrads TP, Veenstra TD, Udseth HR (2002) An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2**(5)**:** 513-523

Solomon JM, Pasupuleti R, Xu L, McDonagh T, Curtis R, DiStefano PS, Huber LJ (2006) Inhibition of SIRT1 catalytic activity increases p53 acetylation but does not alter cell survival following DNA damage. *Mol Cell Biol* **26**(1)**:** 28-38

Strohalm M, Hassman M, Kosata B, Kodicek M (2008) mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Commun Mass Spectrom* **22**(6)**:** 905-908

Tate EW (2008) Recent advances in chemical proteomics: exploring the post-translational proteome. *J Chem Biol* **1**(1-4)**:** 17-26

Toledo F, Wahl GM (2006) Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nat Rev Cancer* **6**(12)**:** 909-923

Ulintz PJ, Yocum AK, Bodenmiller B, Aebersold R, Andrews PC, Nesvizhskii AI (2009) Comparison of MS(2)-only, MSA, and MS(2)/MS(3) methodologies for phosphopeptide identification. *J Proteome Res* **8**(2)**:** 887-899

Vandenbogaert M, Li-Thiao-Te S, Kaltenbach HM, Zhang R, Aittokallio T, Schwikowski B (2008) Alignment of LC-MS images, with applications to biomarker discovery and protein identification. *Proteomics* **8**(4)**:** 650-672

Vaziri H, Dessain SK, Ng Eaton E, Imai SI, Frye RA, Pandita TK, Guarente L, Weinberg RA (2001) hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase. *Cell* **107**(2)**:** 149-159

Vousden KH, Lu X (2002) Live or let die: the cell's response to p53. *Nature reviews* **2**(8)**:** 594-604

Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL, Hochstrasser DF (1996a) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**(1)**:** 61-65

Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, Williams KL (1996b) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* **13:** 19-50

Xirodimas DP, Saville MK, Bourdon JC, Hay RT, Lane DP (2004) Mdm2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity. *Cell* **118**(1)**:** 83-97

Zhang K, Yau PM, Chandrasekhar B, New R, Kondrat R, Imai BS, Bradbury ME (2004) Differentiation between peptides containing acetylated or tri-methylated lysines by mass spectrometry: an application for determining lysine 9 acetylation and methylation of histone H3. *Proteomics* **4**(1)**:** 1-10

Appendix A: Pseudo code of algorithms in POSTMan

A.1 General algorithms

A.1.1 Making peptide pairs

A.1.1.1 Equal mass algorithm

```
//Algorithm to match peptides between untreated and treated sample as a pair
//when there is no mass difference between them

FOR all the peptides in the untreated sample
    FOR all the peptides in the treated sample
        CALCULATE mass difference
        IF ABSOLUTE(mass difference) > set mass threshold value
            REJECT this peptide combination as a peptide pair
        ELSE
            STORE this combination as a pair
        END IF
    END FOR
END FOR
```

A.1.1.2 Modified mass algorithm

```
//Algorithm to match peptides between untreated and treated sample as a pair
//when there is a defined mass difference between them

FOR all the peptides in the untreated sample
    FOR all the peptides in the treated sample
        CALCULATE mass difference
        IF mass difference > (modification mass + set mass threshold value)
        OR IF mass difference < (modification mass - set mass threshold value)
            REJECT this peptide combination as a peptide pair
        ELSE
            STORE this combination as a pair
        END IF
    END FOR
END FOR
```

A.1.2 LC-alignment algorithm

```
//Algorithm to align two LC-MS samples in the time dimension
//Needed due to non-linear drifts in LC elution time between successive samples
//Corrects the retention time of peptides in the treated file with the untreated as reference
//
//RT = retention time
//vector = container holding a point (RT) and a length/direction (ΔRT)

//Store all the alignment vectors
FOR all the peptides in the untreated sample
    FOR all the peptides in the treated sample
        IF the two peptides (treated & untreated) have similar m/z AND RT AND charge
            STORE the RT of the treated peptide
            STORE the ΔRT of the two peptides
        END IF
    END FOR
END FOR

//Sort the vectors based on RT
```

```
SORT list of vectors

//Draw all the vectors as points in the map
DRAW list of vectors

//Iterate over all the vectors and bin them
FOR all the vectors in the list
    BIN vectors in RT intervals based on the set binsize
    AVERAGE ΔRT in every bin
END FOR

//Display the averaged, binned data
DRAW line between averaged vectors, one datapoint per bin

//Optional smoothing of the averaged, binned data
//Uses the moving average algorithm
SMOOTH moving_average_algorithm(binned data)

//Transformation of retention times in treated data file
FOR all the peptides in the treated sample
    TRANSFORM RT based on alignment
END FOR
```

A.2 Filters

A.2.1 Elimination filters

A.2.1.1 Intensity cut-off filter

```
//Filter to remove peptides in both samples that have intensity below a set value

FOR all the peptides in the untreated file
    IF peptide intensity is below the set value
        REMOVE this peptide from further calculations
    END IF
END FOR
FOR all the peptides in the treated file
    IF peptide intensity is below the set value
        REMOVE this peptide from further calculations
    END IF
END FOR
```

A.2.1.2 Charge cut-off filter

```
//Filter to remove peptides in both samples that have charge below a set value

FOR all the peptides in the untreated file
    IF peptide charge is below the set value
        REMOVE this peptide from further calculations
    END IF
END FOR
FOR all the peptides in the treated file
    IF peptide charge is below the set value
        REMOVE this peptide from further calculations
    END IF
END FOR
```

A.2.1.3 FASTA filter

```
//Filter to remove peptides in both samples that do not have a mass correlating with a mass
//from an in silico digest of a loaded FASTA file

//In silico digest of the loaded FASTA file
DIGEST FASTA file based on set parameters and store all possible peptides
```

```
APPLY modifications to the peptides and store all peptide variants
REMOVE duplicates
CALCULATE mass of all peptides including their modifications

FOR all the peptides in the untreated file
    IF peptide mass does not correlate with any calculated mass from FASTA file
        REMOVE this peptide from further calculations
    END IF
END FOR
FOR all the peptides in the treated file
    IF peptide mass does not correlate with any calculated mass from FASTA file
        REMOVE this peptide from further calculations
    END IF
END FOR
```

A.2.2 Pair-matching filters

A.2.2.1 Relative intensity filter

```
//Filter to check peptide-pairs for their relative intensity
//If the intensity ratio is higher than the set value the pair is rejected

FOR all the peptides in the untreated file
    FOR all the peptides in the treated file
        IF intensity ratio between peptides > set value
            REJECT this peptide combination as a peptide pair
        ELSE
            STORE this combination as a pair
        END IF
    END FOR
END FOR
```

A.2.2.2 Retention time range filter

```
//Filter to check peptide-pairs for their retention time difference
//If the unsigned distance in retention time is outside a set range the pair is rejected

FOR all the peptides in the untreated file
    FOR all the peptides in the treated file
        IF ABSOLUTE(ΔRT) between peptides > set max value or < set min value
            REJECT this peptide combination as a peptide pair
        ELSE
            STORE this combination as a pair
        END IF
    END FOR
END FOR
```

A.2.2.3 Direction of chromatographic shift filters

```
//Filters to check peptide-pairs for the direction of RT shifts

//Prohibition of hydrophobic retention time shifts filter
//If the RT of the treated peptide is larger than the RT of the untreated the pair is rejected

FOR all the peptides in the untreated file
    FOR all the peptides in the treated file
        IF RT(treated) > RT(untreated)
            REJECT this peptide combination as a peptide pair
        ELSE
            STORE this combination as a pair
        END IF
    END FOR
END FOR
```

```
//Prohibition of hydrophilic retention time shifts filter
//If the RT of the treated peptide is smaller than the RT of the untreated the pair
//is rejected

FOR all the peptides in the untreated file
    FOR all the peptides in the treated file
        IF RT(treated) < RT(untreated)
            REJECT this peptide combination as a peptide pair
        ELSE
            STORE this combination as a pair
        END IF
    END FOR
END FOR
```

Appendix B: Supplementary methods

These methods are from the work of Dr. Clive D'Santos where endogenous p53 was studied and p300 was overexpressed, either alone or in combination with the deacetylase SIRT1. Refer Results, Section 4.1.4.

Cell culture and immunoprecipitation of p53

HEK293T cells were grown in DMEM with 10% foetal calf serum plus penicillin/streptomycin in 5% CO_2 at 37 °C and were routinely passaged at 60% confluence. Cells were transfected with 10 µg of HA-p300 alone or together with 10 µg of FLAG-SIRT1. Cells were maintained for two days and then lysed in 1 mL of lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, 5 mM KCl, 1% NP-40) in the presence of deacetylases inhibitors (10 mM NAM and 1 µM TSA). Cleared supernatants were subjected to immunoprecipitation first with an anti-p53 antibody (DO-1, Santa Cruz) overnight followed by immunoprecipitation with anti-HA antibody (clone 12CA5) for two hours. The beads from each of the immunoprecipitiations were solubilised in 35 µL of SDS-PAGE loading buffer and 5 µL of the beads subjected to SDS-PAGE for western blotting. Total levels of p300 and SIRT1 were estimated using the anti-HA or FLAG antibody, while acetylation levels of p53 were assessed using a polyclonal antibody specific for acetylated Lys382 (Upstate biotechnology). Acetylation of p300 was assessed by western blotting using a pan specific anti-acetyl lysine antibody (Cell signalling technology). The remaining immunoprecipitations were subjected to SDS-PAGE, coomassie staining and the relevant protein bands were excised and subjected to analysis by mass spectrometry.

SDS-PAGE and in-gel digestion

Proteins were separated by one-dimensional SDS-PAGE using NuPage® Novex Bis-Tris gels and NuPage MES-SDS running buffer (Invitrogen) according to the manufacturer's instructions. The gel was stained with Coomassie using Colloidal Blue Staining kit (Invitrogen). Protein bands were excised and subjected to in-gel tryptic digestion. Briefly the gel pieces were destained and washed, and after dithiothreitol reduction and iodoacetamide alkylation, the proteins were digested with porcine trypsin (modified sequencing grade; Promega, Madison, WI) overnight at 37 °C. The resulting tryptic peptides were removed and the gel pieces further extracted with washing in 10% FA. The two peptide extracts were

combined and analysed by mass spectrometry or subjected to deacetylase treatment prior to MS analysis.

nano LC-ESI LTQ-FT MS/MS

Nanoscale liquid chromatography tandem mass spectrometry (nano-HPLC-MS/MS) FTMS (fourier transform mass spectrometry) experiments were performed on an Agilent 1100 nanoflow system (Agilent Technologies) connected to a 7-Tesla Finnigan LTQ-FT mass spectrometer (Thermo Electron, Bremen, Germany) equipped with a nanoelectrospray ion source. Loading was accomplished by using a flow rate of 5 µL/min onto a homemade 2 cm fused silica pre-column (100 µm i.d.; 375 µm o.d.; Resprosil C_{18}-AQ, 3 µm (Ammerbuch-Entringen, DE) using autosampler. Sequential elution of peptides was accomplished using a linear gradient from Solution A (0.6% acetic acid) to 50% of Solution B (80% acetonitrile; 0.5% acetic acid) in 40 minutes over the pre-column inline with a homemade 20-25 cm resolving column (50 µm i.d.; 375 µm o.d.; Resprosil C_{18}-AQ, 3 µm (Ammerbuch-Entringen, DE). The mass spectrometer was operated in DDA mode to automatically switch between MS and MS/MS acquisition. Survey Full scan MS spectra (from m/z 350-1500) were acquired in the FT-ICR with resolution R=100000 at m/z 400 (after accumulation to a target value of 2000000 in the linear ion trap). The two most intense ions were sequentially isolated and fragmented in the linear ion trap using collision induced dissociation at a target value of 10000.