Novel Methods and Applications of Data-independent and Targeted Mass Spectrometry; Towards Robust Quantification of Molecular Signaling Events

Thierry Thomas Schmidlin

ISBN: 978-94-6295-952-1

The research in this thesis was performed in the Biomolecular Mass Spectrometry & Proteomics Group, Utrecht University, Utrecht, The Netherlands.

Lay-out and print by: ProefschriftMaken // www.proefschriftmaken.nl

Novel Methods and Applications of Data-independent and Targeted Mass Spectrometry; Towards Robust Quantification of Molecular Signaling Events

Neue Methoden und Anwendungen von daten-unabhängiger und zielgerichteter Massenspektrometrie; Studien zur robusten Quantifizierung von Signaltransduktionswegen
(mit einer Zusammenfassung in deutscher Sprache)

Nieuwe methoden en toepassingen van gegevens-onafhankelijke en doelgerichte massa spectrometrie; op weg naar robuuste kwantificatie van moleculaire signalering
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 25 juni 2018 des ochtends te 10.30 uur

door

Thierry Thomas Schmidlin

geboren op 18 augustus 1988
te Basel, Zwitserland

Promotor: Prof.dr. A.J.R.Heck
Copromotor: Dr. A.F.M. Altelaar

Contents

Chapter 1

General Introduction

1. Introduction to Proteomics

Proteins are crucial players in all cellular processes as they enable signal transduction, response to extracellular stimuli, catalysis of metabolic processes, and cell-cell interactions. Any protein's function is highly specific due to the fact that proteins are large macromolecules consisting of long chains of amino acid, which are connected by peptide bonds to form complex 3D-structures. Proteins are the dynamic and varying products of the genome. This process has been coined as the central dogma of molecular biology, as genes can be transcribed to specific mRNA templates which are further processed by ribosomes to translate this information into the structure of a protein [1].

Proteomics describes the study of the proteome, which is defined as the entirety of proteins expressed in a cell at a given condition and time [2]. Compared to the study of genomes (genomics and transcriptomics) proteomics can thus provide a different level of understanding about cellular processes and their dynamics, which genomics and transcriptomics are blind for [3]. This information can include (depending on experimental design):

1. Posttranslational modification (PTM) of the protein such as for example phosphorylation, glycosylation or ubiquitination [4].
2. Different isoforms arising from alternative splicing [5].
3. Interaction partners [6]
4. Protein degradation or posttranslational trimming events [7, 8]
5. A quantitative readout of protein abundance [9]

However, this benefit also presents a drawback. The sheer amount of modifications and splicing events found in a human proteome results in an enormous complexity, presenting a tremendous analytical challenge. On top of the sheer complexity, proteomics also needs to face that the concentration of different proteins span a huge dynamic range and that PTMs can drastically change their chemical, physical and functional properties. To date, several technologies have been established to investigate whole proteomes. This thesis exclusively focusses on mass spectrometry based proteomics, which has been proven to be the most suited technology for large-scale proteome analyses [10]. The main focus of the thesis presented here is the analysis of protein phosphorylation by a specific technique called MS-based targeted proteomics. This introduction aims to provide deeper information about the different aspects used for this research.

2. LC/MS-based Proteomics

A typical proteomics experiment consists of a core workflow starting with sample preparation followed by LC-MS/MS analysis and data analysis. Each of these steps however is highly adaptable to the specific question at hand. The following paragraphs give a short overview.

2.1. Sample Preparation and Separation

The first step of the proteomic workflow is sample preparation which aims to transfer the proteome (or a subset of it) into a LC-MS/MS compatible sample. Starting from samples such as tissue sections, cell cultures etc. the initial step comprises the extraction of the proteins from the matrix which is usually achieved by a combination of chaotropic buffers and mechanical disruption [11]. The resulting protein extract is further processed into smaller peptides that are more suited for LC-MS/MS analysis than entire proteins. Typically, this comprises the reduction and alkylation of the protein's cysteine side chains to prevent the formation of disulfide bridges and the subsequent digestion with site specific proteases, most commonly trypsin [12]. Alternatively, if the focus of the analysis are endogenously occurring peptides the enzymatic digestion can be replaced by size cut-off specific filtration [13].

Typically, the resulting peptide mixtures are high in complexity and cover a high dynamic range, exacerbating the subsequent MS/MS analysis. Several strategies have been developed to reduce the sample complexity prior to MS/MS analysis, ranging from very general separation and fractionation protocols [14] up to the specific enrichment of peptides with a specific chemical characteristic, e. g. specific side chain modifications such as phosphorylation [15]. Often the applied workflows include combinations thereof [16].

The most routine analysis includes online separation by C18 based reverse-phase low-pH chromatography directly coupled to the MS analysis [17]. This separation strategy ensures that the sample complexity presented to the mass spectrometer at each time point of the gradient is drastically reduced compared to the total sample. Likewise, the exact elution time from the column (retention time, RT) provides valuable analytical information about the peptide at hand [18]. In order to reduce the sample complexity even further, this online separation can be complemented by preceding fractionation protocols. These protocols employ offline liquid chromatography in combination with collecting fractions, with each fraction being subsequently subjected to LC-MS/MS analysis. This approach allows to drastically reduce sample complexity already before subjecting them to the online LC-MS/MS analysis [14]. To increase proteomic depth, it is important to consider the concept of orthogonality in this approach. Basically, this concept describes the relationship between the separation power of the first and second dimension. Orthogonality is defined as the effective area of this 2D separation space covered by eluting peaks and is an indication of

how different the two separation techniques are from each other [19]. In the ideal case the orthogonality reaches close to 100 %. Offline high-pH reversed-phase fractionation (HpH-RP) is a system, which has been shown to provide excellent orthogonality to the classically used C18-RP chromatography used in LC-MS/MS – and has been used exclusively for the herein presented research [16]. In this strategy, fractionation is performed on C18 based reverse-phase columns in combination with basic mobile phases (pH = 10). Highest orthogonality is achieved by applying a concatenation strategy, in which different fractions are pooled together prior to LC-MS/MS analysis [20] (Figure 1A).

An alternative to offline fractionation is the use of online two dimensional chromatography, combining both chromatographic separations in a single LC-MS run. An example is the multidimensional protein identification technology (MudPIT) which combines strong cation exchange (SCX) and RP [21, 22]. Both phases are subsequently packed into the same capillary which is directly connected to the MS (Figure 1B). Peptides are trapped on the SCX column and displaced to the RP column by using increasing salt steps. Prior to each salt step peptides are eluted from the RP phase by an acetonitrile gradient [23] (Figure 1C).

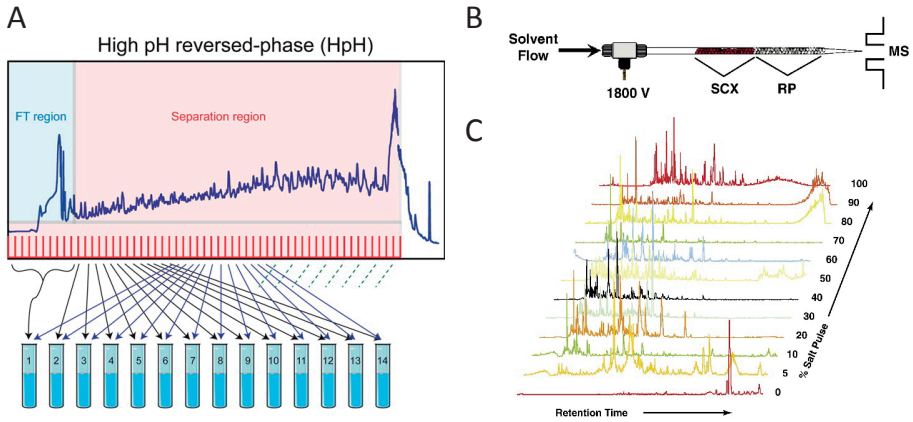


Figure 1: ***2D Chromatography****. (A) High-pH reversed-phase chromatography is often used for sample fractionation. In combination with on-line low-pH RP chromatography highest orthogonality is usually achieved by applying a concatenation strategy. In this strategy different fractions across the gradient are pooled together prior to LC-MS/MS analysis (adapted from Batth* et al. *[16]). (B) MudPIT enables online two-dimensional chromatography by subsequently packing RP and SCX phase into the same capillary. (C) Salt pulses of increasing concentrations are used to displace the peptides from the SCX column to the RP column where they are separated by an acetonitrile gradient (adapted from di Palma* et al. *[23]).*

In contrast to the above mentioned protocols that generally reduce sample complexity, other protocols aim to specifically enrich for different classes of (mostly very low abundant) peptides. A classic example is the enrichment of phosphopeptides. Phosphorylated

peptides generally occur at very low stoichiometry and ionize poorly [24]. However their unique chemical properties enable for very specific enrichment facilitating the separate analysis of the proteome and the phosphoproteome [25]. The most commonly used strategies for enrichment of phosphopeptides include metal affinity chromatography and immunoprecipitation, which will be discussed in more detail in **Chapter 1.4.1**. Specific enrichment protocols have been developed for many more subclasses of peptides, with the most common being acetylated peptides [26], glycopeptides [27, 28] and ubiquitinated peptides [29].

2.2. MS and MS/MS Analysis

The MS measurements of peptides determines their mass-over-charge ratio (*m/z*). Despite the high mass accuracy of the current mass spectrometer generation, this information usually does not give enough information to reliably identify the peptide at hand. Therefore, key to the determination of peptide sequences is the use of appropriate fragmentation events in order to obtain sufficient sequence information [30]. To date MS instruments come as hybrid instruments, however all of them rely on three crucial components, an ion source, a mass analyzer and a detector. The ion source facilitates the transfer of analyte molecules to gas-phase ions, the mass analyzer separates those ions based on their *m/z,* which are then recorded by the detector. The following paragraphs describe these general principles in more detail and provide an overview of the types of MS instruments used in this thesis.

Ionization

A mass spectrometer essentially measures the mass (m) of a molecule in relation to its charge (z). Thus, the peptides need to be ionized and transferred to the gas phase prior to MS analysis. Throughout this thesis, the method of choice for ionization is electrospray ionization (ESI) directly at the interface between chromatography and MS.

In principle ESI applies a high voltage between the column emitter and the MS inlet. This high voltage results in an accumulation of charges at the tip of the capillary creating a Taylor cone, constantly releasing charged droplets similar to a spray [31]. These droplets are usually small and contain many charged species resulting in a high charge density within the droplet, which continuously increases due to the evaporation of the solvent. Once the electrostatic repulsion exceeds the surface tension a so called Coulombic Explosion generates free ions by a yet not fully understood mechanism [32].

Several factors can be implemented in order to improve the formation of single ions in the gas phase. Most importantly all MS instruments employ heating to 150-350°C on the inlet to facilitate the evaporation of solvent in order to ease the formation of free ions. Additionally, it has been shown, that reducing the initial droplet size aids ionization and reduces competitive ion suppression effects. The reduction of droplet size is called nano-

electrospray ionization and results in unparalleled sensitivity, which is especially exploited in proteomics research [33].

Mass analyzers and filters

Within the mass spectrometer, ion optics are used to guide ion packages through the instrument from source to detector. The most common ones used nowadays are described in the following paragraphs:

1. **Quadrupole Mass Filters (Q):** A quadrupole is an assembly of four parallel metal rods kept at equal distances, in which the two pairs of opposite rods are electrically connected. It can serve three purposes: (1) as a simple ion guiding system (2) as a filter for very specific *m/z* ions (3) as a mass analyzer. When used as an ion guiding system, RF voltages are applied to the rods. This causes the ions to be radially confined adopting a cork-screw like motion called "secular motion" on their path through the quadrupole. The ion motion is proportional to the RF amplitude and the ion mass. Low mass ions experience more motion than larger ions and higher RF induce more ion motion. Upon low RF voltages all ions can pass through the quadrupole, enabling its use as an ion guiding system. When used as a mass filter, DC and RF voltages are applied to the rods. In the x-z dimension RF voltages result in secular motion of the ions on their path through the quadrupole. The lower the mass of the ions, the more secular motion is induced. Adjustment of the RF amplitudes can thus be used to create a "high mass pass filter", in which ions with a mass lower than a certain threshold experience to much motion and will be sent to an unstable trajectory causing them to be ejected from the quadrupole. The higher the RF voltage, the higher the high mass pass filter. In the y-z direction, a DC voltage is applied, attracting the ions and pulling them out of the RF induced secular motion. For small ions the radial confinement force however is strong enough to overcome the attracting force of the DC voltage, pulling them back into the central path. Larger ions, however have too much inertia and cannot be pushed back into the central path anymore upon experiencing the attracting force of the DC voltage. This generates a "low mass pass filter". When used as a mass filter, quadrupolar RF and DC voltages are adjusted in a way that only ions, which are big enough for radial confinement and small enough that the radial confinement force is larger than the attractive DC force, can adopt a stable trajectory. This results in a clearly defined isolation window [34]. (Figure 2A)
2. **Linear Ion Trap (LIT):** LITs are used to trap ions in a confined space. Similar to quadrupoles, the linear ion traps consists of four parallel metal rods. A main RF is applied to these rods, at equal amplitude but opposite sign for the two pairs, allowing to confine the ions radially. To enable axial confinement of the ions, LITs are equipped with either short quadrupoles or metal plates at both ends of the quadrupolar rods. They are used to create a "potential well" by applying a DC offset (see Figure 2B).

When trapped the main RF exerts a force onto the ions which induces ion motion. Hence, the ions go into a so called "secular motion". The larger the ions, the more force is required, resulting in smaller ions to move faster than larger ones. This can be exploited for a process called resonance ejection. Two of the four center rods are equipped with ejection slits. They are called exits rods, are opposing each other, and an additional AC is applied to them. Once the main RF causes the secular motion of a specific ion to match the frequency applied to the exit rod, that ion will start to resonate and hence eventually acquire enough energy to be ejected through the ejection slit. In a scan, the main RF is ramped up continuously, causing ions of different *m/z* to eject sequentially. Likewise, resonance ejection can be used to achieve ion isolation, by ejecting all ions except for the *m/z* of interest [34].

One major challenge of ion trap instruments is the so called space charge effect. The ion cloud trapped in the LIT contains ions of the same charge that repel each other interfering with their trajectories. With higher ion numbers, these effects get increasingly pronounced leading to peak shifts and broadening [35], hence limiting the maximum ion load of the LIT (Figure 2B).

3. **Time of Flight (TOF):** In a TOF MS the ions' m/z are measured by an arrival-time measurement. The ions are accelerated in an electric field of known strength resulting in all ions of the same charge state having the same kinetic energy. As $E_k = 0.5mv^2$ ions with a higher mass will have a lower velocity than ions with a higher mass. Therefore, measuring the time an ions needs to travel a predefined distance allows accurate *m/z* determination [36] (Figure 2C).
4. **Orbitrap**: In Orbitrap mass analyzers the *m/z* is determined by ion frequency during an oscillatory motion. The Orbitrap consists of a barrel shaped outer electrode and a spindle-shaped inner electrode. Ions are trapped due to a balance between electrostatic attraction to the inner electrode and their initial centrifugal forces. This results in a harmonic oscillation of the ions along the longitudinal axis of the spindle. The frequency of the ions moving along the spindle axis exclusively depends on the ions' *m/z*. The ion motion results in an image current (transient) on the outer electrode that can be transferred to an *m/z* spectrum by fast Fourier transformation [37] (Figure 2D).

Most modern MS instruments are hybrid instruments. They employ multiple mass analyzers and filters that are strategically coupled, and ions can be moved between them through ion optics. The geometry of assembling the individual mass analyzers, filters and fragmentation cells highly influences the choice of experiments an instrument is suited for. The following paragraphs contain a short description of the type of hybrid instruments used throughout this thesis:

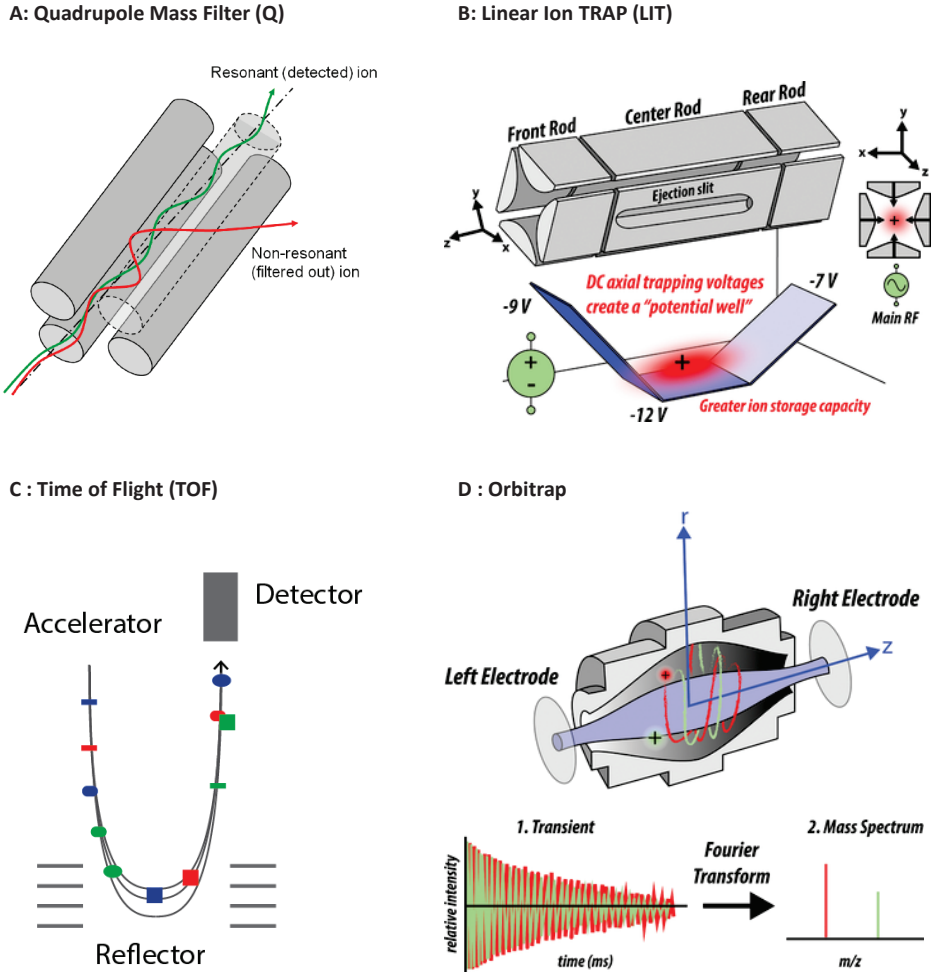
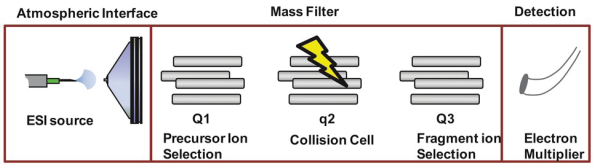


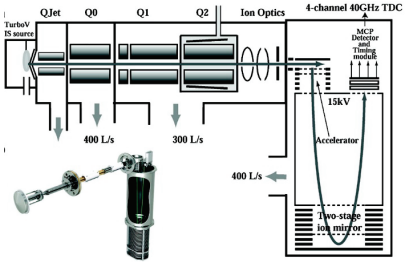
Figure 2: ***Mass analyzers and filters.*** *(A) Schematics of a quadrupole mass analyzer. (B) Architecture of a linear ion trap (LIT) showing the main building blocks as well as the principle of axial and radial confinement (adapted from Savaryn* et al. *[34]). (C) Schematics of a time of flight analyzer (D) Architecture of an Orbitrap. The oscillation frequency of ions along the spindle axis is detected by electrodes in form of a transient signal that can be transformed to a mass spectrum by Fourier transformation (adapted from Savaryn* et al. *[34])*

1. **Triple Quadrupole:** The triple quadrupole MS consists of two quadrupole mass analyzers (Q1 and Q3) in series connected by a RF-only quadrupole (Q2) which can be used for collision-induced dissociation. At the back end of Q3 an electron multiplier is used for ion detection. Triple quadrupole instruments are primarily used for selected reaction monitoring (see below) in which no full scan spectra are acquired, but the quadrupoles are merely used as mass filters [38] (Figure 3A).

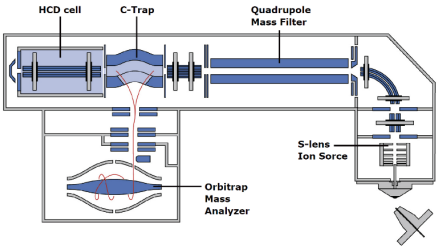
A: Triple Quadrupole



B: TripleTOF



C: Q Exactive



D: Fusion Lumos

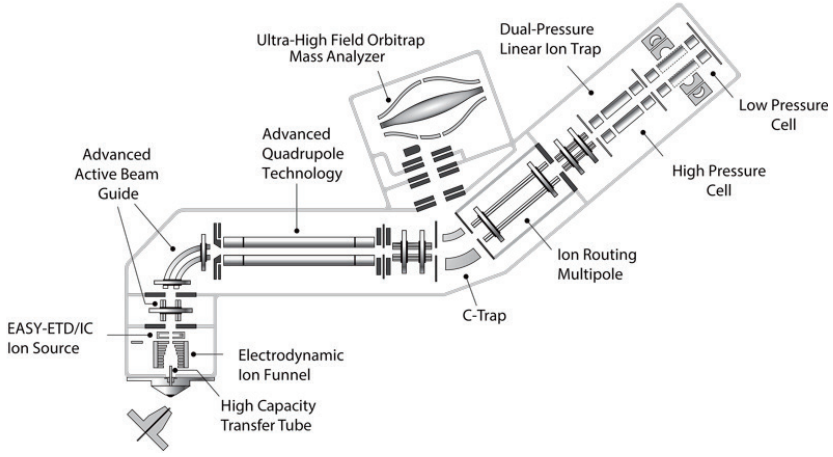


Figure 3: ***Hybrid MS instruments.*** *(A) Depiction of the architecture of a triple quadrupole instrument (adapted from Bereman* et al. *[44]). (B) Schematic representation and real image of a TripleTOF instrument (SCIEX) (adapted from Andrews* et al. *[40]). (C) Schematic representation of Q Exactive (Thermo Fischer) and (D) Fusion Lumos (Thermo Fischer).*

2. **QTRAP:** QTRAP type of instruments are very similar to triple quadrupole type of instruments with the only difference, that Q3 is replaced with a LIT. Compared to triple quadrupole instruments, this enables additional acquisition of higher resolution MS

and MS/MS spectra as well as employing an additional secondary fragmentation step (MS3) [39].

3. **TripleTOF:** TripleTOF instruments employ an analogous setup as triple quadrupole and QTRAP for Q1 and Q2, however they contain a TOF analyzer instead of Q3 [40] (Figure 3B). This facilitates the fast acquisition of full scan MS/MS spectra at high resolution, hence facilitating its use for data-dependent and data-independent acquisition (see below).
4. **Q Exactive:** This instrument employs the Orbitrap as its main mass analyzer for MS and MS/MS acquisition. Upfront ion isolation is achieved with a quadrupole and an HCD cell can be used for ion fragmentation [41] (Figure 3C).
5. **LTQ-Orbitrap series:** This instrument series employs precursor ion selection in the LIT (and in newer model such as the Fusion and the Fusion Lumos (Figure 3D) also quadruple isolation), several peptide fragmentation techniques and LIT or Orbitrap as mass analyzer. It is highly versatile in terms of acquisition method [42]. Newer generations (e. g. Lumos – Figure 3D) employ a different instrument geometry that allows for a high degree of parallel acquisitions facilitating faster data acquisition [43].

Peptide Fragmentation

In complex mixtures such as digested protein extracts from cells, the mere MS1 information is not enough evidence to conclusively identify a peptide. This however can be strengthened drastically by employing tandem MS (MS/MS). In MS/MS a first mass analyzer is used to selectively isolate one analyte of interest which is then transferred to a reaction chamber where it is fragmented. The fragment ions are then transferred to a mass analyzer and a fragmentation spectrum is recorded. The observed fragment ion pattern can be used for sequence analysis, as they ideally contain fragment ions resulting from cleavages along the backbone. The resulting ions are called as depicted in Figure 4 [45]. The most commonly used types of fragmentation in proteomics are outlined below:

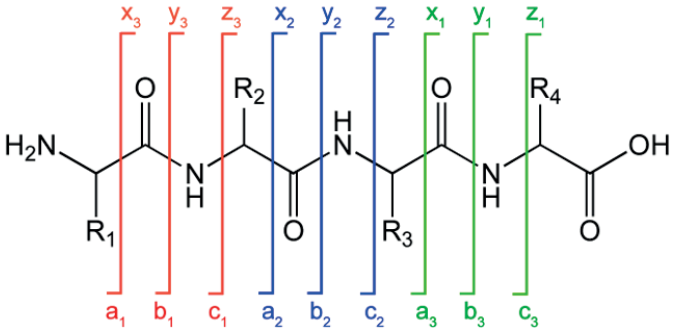


*Figure 4: **Peptide fragmentation.** Common nomenclature of peptide fragment ions as proposed by Roepstorff and Fohlman in 1984 [45].*

1. **Collision-induced dissociation (CID):** In CID fragmentation is achieved by accelerating ions to a high kinetic energy and allow them to collide with neutral inert gas molecules such as nitrogen or argon. This collision leads to the conversion of some of the kinetic energy into internal energy which results in bond breakage. The eventual location of the fragmentation event is highly dependent on amino acid composition of the peptide, potential chemical modifications on the amino acid side chains, and the actual CID parameters applied. Typically, CID is achieved in two different ways depending on the exact location within the mass spectrometer [46]:
    a. Beam-type CID: This fragmentation technique is occurring in quadrupoles (QqQ, QTof), where precursor ions are transferred into the gas filled collision cell at a high speed. This setup allows for the deposition of high (kinetic) energy onto the precursor ions prior to collision with the gas molecules.
    b. Resonance-excitation CID: Here, fragmentation is performed in an ion trap instrument (e. g. LIT). Ions are excited with a precursor specific energy while collision with the inert gas causes fragmentation. Typically, the amount of energy deposited on the ions prior to fragmentation is thus much lower than in beam-type CID.

    When applied to peptides, both CID fragmentation techniques primarily give rise to b- and y-type ions, but they differ in certain main aspects. Whereas fragment ions formed in beam-type CID can undergo further undesired fragmentation reactions, such secondary reactions usually do not occur in resonance-excitation CID as the excitation energy is not high enough. On the other hand, especially for the analysis of phosphopeptides, the use of resonance-excitation CID can be detrimental, as it primarily follows low-energy fragmentation pathways. The labile phosphate group provides such a low-energy pathway by the neutral loss of H_3PO_4, thus impeding the formation of sequence informative ions [47]. Beam-type CID is less affected by this, as the fragmentation occurs faster and secondary fragments (such as the neutral loss ion) can still undergo fragmentation.
2. **Electron-driven dissociation techniques:** In recent years, electron-driven fragmentation techniques have proven to be a powerful alternative to CID. In these methods, fragmentation is triggered through the uptake of an electron by the positively charged peptide, creating a radical ion-species and triggering its fragmentation. Electron-capture dissociation (ECD) is the first electron-driven fragmentation method that had been developed [48]. In ECD, low energy electrons react with peptide cations in the magnetic field of a FT-ICR-MS. The later introduced electron-transfer dissociation (ETD), however, utilizes ion/ion chemistry to facilitate the deployment of the electron on the analyte [49]. This is achieved by bringing the cationic analytes and a radical anion (typically a fluoranthene radical anion) into close proximity within the ion trap, resulting in the transfer of an electron from the radical anion to the analyte. The main

difference of triggering fragmentation by creating free thermal electrons (ECD) or transferring them from a radical anion (ETD) is the longevity of the reagent in the ion trap. The thermal electrons used in ECD can only maintain their thermal energy for a fraction of a microsecond when trapped in RF fields of LITs, thus making it a fragmentation technique exclusively available on FT-ICR-MS instruments. In contrast, ETD uses a RF quadrupole ion trapping device, which is a more commonly available type of instrumentation, rendering ETD the method of choice.

In comparison to collisional dissociation techniques, ETD and ECD induce fragmentation of the Cα-N bond resulting in c and z-type ions [30].

2.3. Data-dependent MS acquisition methods and data analysis

The most commonly used acquisition methods in LC-MS/MS are referred to as data-dependent acquisition (DDA) or shotgun methods. These methods contain a so-called survey scan, which records an MS scan, usually at high resolution. Based on their intensity, several peptides observed in this MS scan are subsequently subjected to fragmentation and MS/MS acquisition. To avoid continuous fragmentation of the same peptides, each peak is excluded from fragmentation for a defined time frame after being picked, a concept referred to as dynamic exclusion [10].

Data acquired from shotgun methods contain comprehensive MS1 information and thousands of MS/MS snapshots. Hence, their sequence assignment is usually performed in an automated way, using computer algorithms. The most commonly used way are database searches [50], in which hypothetical proteomes (or even genome sequences) are used to create an *in silico* proteome digests. The list of theoretically possible peptide masses is then used to match the actually measured masses against. Each hypothetical peptide sequence matching an actually measured precursor mass is then used to create hypothetical MS/MS spectra, based on the expected ion series from the chosen fragmentation technique. These hypothetical MS/MS spectra are then compared to the recorded MS/MS spectra, in order to find the best match and to identify the peptide sequence [51]. To each of these peptide spectrum matches a score is assigned reflecting its similarity [52]. Given the sheer size of the database and the number of recorded MS/MS spectra, however, random matches are prone to happen. A common strategy to control this effect is the concept of false discovery rate (FDR). Usually this is achieved by extending the database used for the data analysis with a reversed or randomized decoy database. Any match to the decoy database is considered a false positive. Typically score cut-offs are chosen in a way to keep the amount of false positive identifications from the decoy database below 1 % [53].

Traditionally DDA experiments are employed to compare different biological conditions (e. g. healthy versus disease, treated versus untreated, etc.) with each other. In this situation the need for a quantitative readout is more important than mere identification

of the analytes. However, MS is not *per se* a quantitative method, as different analytes have very different ionization efficiencies, which can additionally be influenced by their surrounding matrix. It is however possible to compare identical analytes across different samples, assuming these samples do not differ in terms of their chemical composition. A commonly used approach employs a strategy called label-free quantification. In this approach samples with different conditions are analyzed sequentially and the abundance of any given peptide is compared throughout the different runs, usually taking the integrated chromatographic peak of the peptide's MS signal as a quantitative readout [54]. However, this approach requires high reproducibility across the whole sample preparation workflow including the LC-MS/MS analysis. Additionally, basing quantification on MS1 signal is challenging in complex samples, as the MS1 space is noisy and prone to interferences [55]. Therefore, this approach requires a robust peak picking algorithm, high mass accuracy and resolution of the MS1 measurements and a high separation power of the online LC system.

2.4. MS based targeted Proteomics

An alternative to shotgun acquisition methods is the use of targeted MS-based proteomic methods. Compared to the shotgun approach, targeted methods do not rely on stochastic, intensity based peak picking algorithms. Instead they make use of hypothesis driven approaches in which dedicated methods are built for experiment specific target lists. Each of these target lists consists of individually optimized assays for every peptide of interest. This has several advantages over shotgun experiments. By circumventing the inherent bias towards high abundant peptides seen in shotgun approaches [56], targeted MS usually achieves a higher dynamic range and higher sensitivity than discovery experiments and it provides full control over which peptides are analyzed.

Selected Reaction Monitoring

Classically, MS-based targeted proteomics is performed in triple quadrupole mass spectrometers operated in Selected Reaction Monitoring (SRM) mode [57]. SRM combines sets of *m/z* filters for Q1 and Q3 representing peptide specific precursor and fragment ion pairs, so called transitions. The second quadrupole acts as a collision cell [38]. Chromatographic peak traces of typically 3-5 transitions per peptide are recorded in a predefined retention time window and their chromatographic area under the curve provides a quantitative readout. Each new set of SRM assays requires dedicated method development, optimization and validation prior to its application.

Typically, the process of developing an SRM method cycles through the following steps:

1. **Peptide Selection:** The initial peptide selection is highly dependent on the specific question at hand and is depending on the goal, e. g. the quantification of protein abundance, PTM stoichiometry or the abundance of endogenous peptides. In case the scope of the experiment is to quantify protein abundance, typically several peptides

(up to 3) are targeted for each protein of interest. The optimal procedures of selecting appropriate peptides from a protein sequence has been summarized in the concept of the proteotypic peptides [58]. It involves considering factors such as high ionization efficiency and lack of potential modification sites that could falsify quantification.
In principle, the target selection can be based purely on theoretical considerations (as is described here in **Chapter 3**), however most experiments rely on previously acquired data about the MS compatibility of the peptides at hand. These can originate directly from shotgun experiments (e. g. from online repositories) or from computational prediction tools [59].

2. **Transition Selection:** Once the list of target peptides for a specific experiment is fully designed, SRM assays for each peptide are individually developed and optimized. Usually this is performed on pure synthetic peptides of the target sequences, in order to avoid interferences from the matrix. The initial step in developing an assay is the selection of the most sensitive transitions. The intensities of individual fragment ions from the same precursor peptide can vary greatly within a MS/MS spectrum. To achieve sensitivity, it is therefore crucial to select the most intense fragment ions. This information can be obtained from shotgun datasets or by direct analysis of the synthetic peptide sequence. It is however crucial to stick to similar fragmentation types, i.e. beam-type CID [60] or HCD [61], to successfully mimic the fragmentation events occurring in a triple quadrupole.
3. **Transition Validation:** Based on the relatively low resolution (0.7 Da isolation width in Q1 and Q3) used in SRM assays, unspecific signals from other peptides with similar transitions regularly occur. Peptides with closely related sequences might by chance yield multiple identical transitions. To avoid quantifying interferences or artefact signals, multiple steps of validation are required, that are typically performed by spiking heavy labelled synthetic peptides into a complex matrix alike the one used for later experiments. Indications for specificity of an SRM assays are for example:
    i. **Perfect co-elution of multiple transitions**: If a high number of transitions have similar chromatographic behavior, they are likely originating from the same peptide precursor, thus the use of multiple transitions to measure a particular peptide can be used to evaluate assay specificity.
    ii. **Confidence through peptide identifications**: In case MS/MS identifications are available for the peptides of interests from DDA runs, this data can be used to validate the assays as well. Steps include comparison of the relative fragment ion intensities acquired in the SRM assay and the full scan MS/MS spectrum [60] and comparison of retention times between the peptide ID and the SRM peak group. In case the DDA data and the SRM measurements are acquired on different LC platforms RTs can easily be transferred by using the iRT concept [18]. The iRT is a dimensionless peptide-specific value representing the peptide's RT relative to a

standard set of peptides. This enables direct RT transfer between different LC systems by linear regression. Alternatively, peptide IDs can also be acquired directly on a triple quadrupole platform by a protocol referred to as SRM triggered MS/MS, in which – similar to DDA analyses – MS/MS acquisitions are triggered once certain SRM transitions exceed a set intensity threshold [62].

4. **Transition Optimization:** Provided an initial assay has been developed that proofs to be sequence specific in the respective background, sensitivity can usually be increased by transition specific parameter optimization. Usually this optimization is limited to normalized collision energy (CE) and declustering potential (DP). CE parameters often display sharp optimums and an individual optimization can drastically increase the assay sensitivity (Figure 5). Roughly, the CE optimum is approximately linearly correlated to its mass for a given charge state [63], enabling a reasonably good prediction of the given optimal value for each peptide. However individual transitions can deviate considerably from this trend [38]. Also, optimization of the DP can increase sensitivity, however the effects are less pronounced than for the CE as DP usually provides a broad optimum.

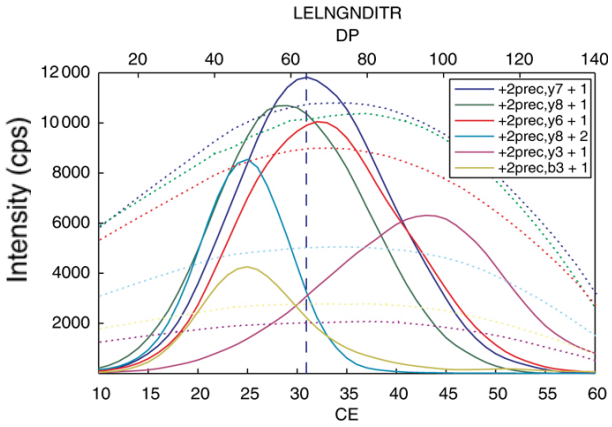


*Figure 5: **Effects of CE and DP optimization on transition intensity**. Plot shows intensity of 6 different transitions of the same peptide versus applied collision energy (solid line) or declustering potential (dashed line) (adapted from Lange* et al. *[38])*

Different kinds of software have been developed to facilitate one or more of the steps mentioned above. The most prominent and comprehensive platform for SRM assay development, refinement, validation and data analysis is Skyline [64]. It facilitates most of the workflow described above by enabling rapid visual assessment of both MS/MS spectra as well as chromatographic peak traces of SRM transitions. Likewise, it supports both individual steps of method development as well as subsequent data analysis. Further software pieces that assist individual steps of SRM method development include SSRCalc

[65], which facilitates retention time prediction and SRMCollider [66] a tool that calculates non-redundant theoretical SRM assays for targets in a given proteomic background.

A major advantage of SRM against other protein specific assays is its multiplexing capabilities. Typical SRM experiments target tens to hundreds of peptides. However, multiplexed methods have to be carefully assessed before applied to quantification experiments, as there is a tight interplay between two key parameters, the dwell time and the cycle time. The dwell time describes the acquisition time given to each transition. This time has to be long enough to get a stable signal for each transition. The minimal value for the dwell time is highly MS dependent but is usually around 10 ms [57, 67]. Since the transitions are measured consecutively, the dwell time however also affects the cycle time, which describes the gap time between two measurement points of the same transition. To ensure reliable quantification it is crucial to collect enough measurement points over the chromatographic peak. Therefore, in theory, the cycle time should be as low as possible, but in practice, collecting 10-12 measurement points over the chromatographic peak proves sufficient. So dependent on the peak width of the chromatography and the gradient length, typical cycle time values range from 1-5 seconds (Figure 6). An additional strategy that can be applied in order to increase multiplexing is the use of retention time scheduling. By ensuring that peptides are only measured in a small, specific time window of the gradient, the amount of actual measurement events for each time point in the gradient drops significantly. By optimizing gradient length, peak width, and chromatographic reproducibility, the number of transitions measured in the same run can be increased by factors of up to 20 [68].

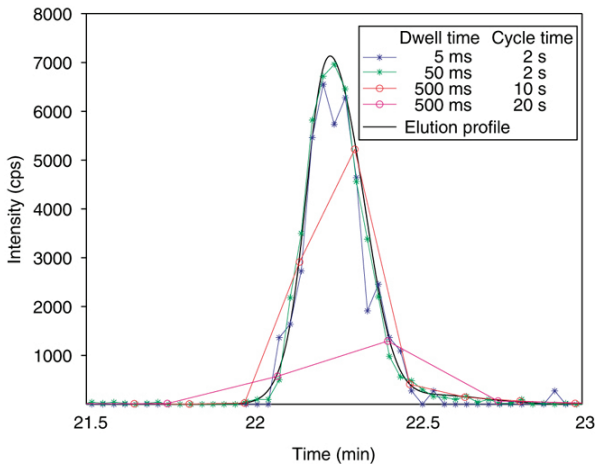


Figure 6: ***Effect of cycle time and dwell time on peak shape****. Black line shows actual elution profile of the analyte while different colors represent assumed peak shapes based on the measured data points obtained from MS methods with different dwell time and cycle time parameters (adapted from Lange* et al. *[38]).*

Once optimized, the SRM methods can be applied to detect and quantify the peptides of interest in real life samples providing a fast, relatively inexpensive and reproducible way of quantifying tens to hundreds of peptides across large numbers of samples. Compared to the method development, the data analysis is relatively straightforward as the chromatographic area under the curve of the SRM transitions can be directly related to quantitation. Depending on the required quantitative accuracy, three different quantification strategies can be applied:

1. **Label-free quantification**: In this quantification approach, the chromatographic area under the curve of a single transition or all transitions summed up is directly used as a quantitative readout. Comparing integrated peak areas enables relative quantification of peptides across multiple samples. While being relatively easy to implement and analyze, this quantification strategies does not account for factors such as fluctuating ionization efficiencies of the same ions over time or across different backgrounds, and fluctuation of LC or MS performance over time. Therefore, label-free quantification is often less accurate than other quantification strategies [69].
2. **Stable isotope-based quantification**: The addition of isotopically labelled internal standard peptides can be used to overcome the above described limitations. Usually peptides labeled with stable 13C and 15N-labeled lysine or arginine are used for each analyte of interest and spiked into the sample prior to the MS analysis. Any bias introduced in the analysis, be it in sample preparation, matrix effects or LC-MS performance will equally affect the endogenous analyte and its synthetic heavy labeled analogue. It is crucial that the subsequent acquisition method contains assays for both, the endogenous and the heavy labelled peptide and their intensity ratio is used as a base for relative quantification across different samples. This method enables the robust detection of changes in abundance as low as 10 % between different conditions [70]. In addition to being more quantitatively accurate than label free quantification, the required perfect co-elution of the transitions from both endogenous and labeled peptides provides certainty about the correct peak integration.
3. **Absolute quantification**: In relatively rare cases, relative quantification is insufficient, for instance when determining protein copy numbers. In these instances, absolute quantification is required, which is performed similarly to stable isotope-labeled quantification, but in a much more stringent way. Additional to controlling for biases and losses during sample preparation and measurement, the internal standards provide a quantitative reference point. Subsequently the absolute amount of the analyte can be deduced by the relative intensity of the light/heavy transitions. This however requires precisely quantified standard peptides (AQUA) used as internal standards and a careful control of multiple factors, such as sample loss, incomplete digestion, linear range of the MS measurement, etc. [71].

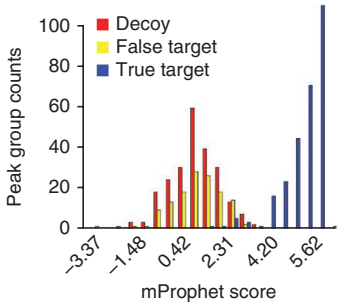


Figure 7: ***mProphet score distribution.*** *Separation of true and false target peak groups by mProphet after training of a classifier with a semisupervised learning strategy (adapted from Reiter* et al. *[72])*

Common to all three quantification approaches is the need for a well-controlled peak group integration to avoid basing quantification on non-target peak groups or interference signals. Commonly this is achieved by manual data assessment ensuring a correct definition of the peak boundaries and the exclusion of interfering transitions. However, this approach is subjective and becomes tedious upon increasing numbers of samples and/or targets. One way to automatize and control this integration is the use of the software mProphet [72]. It scores peak integration by the use of several sub-scores. These sub-scores assess different factors such as co-elution of all transitions, co-elution of the analyte with its internal standard, RT accuracy, etc. These sub-scores are subsequently integrated into an overall discriminant score. However, the specific weight that each sub-score has to be given in order to maximize the separation of true and false peak groups is highly experiment dependent (e. g. presence or absence of internal standards, LC reproducibility, sample complexity). Therefore, mProphet relies on the concept of decoy transitions that need to be introduced on a measurement level. These decoy peak groups are used to optimize the linear combination of sub-scores in a semi-supervised learning process. In this process sets of high scoring peak groups (true signals) and decoy peak groups (false signals) are used as a training set to iteratively refine the linear combination of sub-scores. Thus, the final weight of each sub-score ensures maximized discrimination between true and false peaks resulting in a final normalized discriminant score called mProphet scores. The mProphet score is used to derive FDRs which can be easily determined by a score cutoff due to the null distribution of scores attributed to the decoy peak groups [72]. Currently, mProphet is fully implemented in Skyline, enabling automated FDR control of peak integration as well customized generation of decoy transitions [73]. However, mProphet separates identification from quantification and is primarily geared towards providing an FDR control for peak group assignment and optimizing sensitivity that is often simply hampered by the ability to distinguish true from false peak groups.

SRM-like methods

Multiple SRM like methods have been developed to enable targeted proteomics on instruments different from triple quadrupoles, especially using high resolution instruments such as QTOF or Orbitrap mass spectrometers [74, 75]. They are therefore often referred to as high resolution MRM (HR-MRM). Compared to SRM, they solely rely on precursor ion isolation in combination with acquisition of high resolution/accurate mass MS/MS spectra. Transitions are then extracted post-acquisition from the full fragmentation spectra. The method design is ultimately defined by the instrument architecture. While QTOF instruments (e. g. TripleTOF, *Sciex*) profit from a low duty cycle, they are able to analyze different precursors consecutively, achieving a similar multiplexing capability as SRM. Orbitrap based instruments (e. g. Q Exactive, *Thermo Scientific*) suffer from a longer duty cycle due to the Orbitrap readout, however, Orbitraps can accumulate several precursor ions in the C-trap before fragmenting and analyzing them simultaneously in the Orbitrap. This approach, referred to as Parallel Reaction Monitoring (PRM) [76], allows Orbitrap instrument to compete with triple quadrupoles or QTOF instruments in terms of multiplexing capability despite longer duty cycle.

Several factors have to be taken into consideration when comparing performances of SRM with those of HR-MRM or PRM. HR-MRM and PRM methods require less method development, as transitions do not need to be optimized beforehand and are supposed to surpass SRM in terms of specificity. Additionally, HR-MRM and PRM data can be subjected to database searches like shotgun files, hence confirming assay specificity by including FDR-controlled peptide IDs. However, these benefits come at the cost of lower flexibility, especially in terms of a per transition optimization, as possible in SRM. Additionally, the sensitivity of a triple quadrupole operating in SRM mode is still unparalleled by any other type of mass analyzer to date (2017) [77].

Data-independent acquisition

The most recent development in MS-based targeted proteomics is the development of data-independent acquisition (DIA) methods, also referred to as SWATH (sequential window acquisition of all theoretical spectra) [78, 79]. DIA approaches differ from both shotgun and SRM, by following patterns of broad consecutive isolation windows that are neither based on MS1 spectra as in shotgun, nor on *a priori* knowledge as in SRM. The aim of DIA methods is rather the cyclic recording of fragmentation spectra for the entire MS1 range within the chromatographic time scale. This generates complete fragment ion maps, however at the cost of precursor ion specificity. This is due to the fact that the duty cycle of even the fastest MS platforms available today can only cope with the required mass ranges and chromatographic cycle times by increasing precursor isolation widths to around 20 *m/z* (Figure 8). Computational simulations have shown, that DIA methods can compensate this loss in precursor mass accuracy by increased fragment ion mass accuracy

and can thus compete with SRM in terms of specificity [78]. However, the correct data analysis remains challenging, due to (1) highly multiplexed MS/MS spectra and (2) losing the link between fragment ion and precursor ion. The most successful approaches so far include:

1. Targeted data extraction: In this approach DIA data sets are used to be queried for the presence of specific peptides of interest based on prior knowledge acquired from spectral libraries. Information such as fragment ion signals and their relative intensities, chromatographic co-elution and retention time, precursor mass, etc. are used to specifically extract ion traces from the highly complex fragment ion maps [78]. If the information of the extracted ion traces correlates well enough with the information of the spectral library, the peptide is identified. Software, such as OpenSWATH [80], enables FDR control of the peptide identification by applying similar strategies as mProphet in SRM, but including additional, DIA specific, scoring elements such as for instance precursor and fragment mass accuracy. Quantification is usually achieved by integration of peak areas of all transitions over the chromatographic time scale. Compared to SRM the use of internal standards in DIA is less common.
2. Generation of pseudo-DDA spectra: The underlying principle of this approach is to rebuild the connection between precursor ions and fragment ions. Most algorithms achieve this by relying on peak shape similarities of their chromatographic traces. This approach enables the generation of pseudo-DDA spectra that can be analyzed by database searches. One example for this approach is the software DIAUmpire, which starts with a feature-detection algorithm in MS1 and MS/MS. Correlating features from MS1 and MS/MS spectra are then grouped together on the basis of Pearson correlation of LC elution profiles. These precursor-fragment ion groups are then used for database searches and the creation of spectral libraries, which in turn can be used for targeted data extraction used for quantification [81].

A recently introduced variant of DIA makes use of the trapping capability of the C-trap in Q Exactives. This allows them to randomly combine multiple 4 *m/z* isolation windows before fragmentation and MS/MS readout. This randomization enables to reduce signal interference as the combination of ions analyzed differs in each cycle, eventually facilitating unambiguous deconvolution. This technology essentially enables comprehensive mapping of fragment ions with similar isolation widths as used in shotgun or SRM [82].

While being still in its early years, DIA shows some great potential as it facilitates the acquisition of comprehensive fragment ion information across the entire LC gradient. This has several implications. First, it is so far the only LC-MS based proteomic technology that provides label-free MS/MS based quantification next to SRM and SRM like technologies. As MS1 signals are usually much more affected by interferences than MS/MS signals [55], DIA opens the doors to a more accurate quantification process, especially when further

improvements in instrument speed and data analysis enable the use of increasingly small isolation windows. Second, DIA files can be seen as digitized sample records that can be re-analyzed for the presence of new targets any time post-acquisition without the need of a new assay development [78, 83].

Altogether DIA has the potential to become a tool combining the depth of shotgun with the specificity, the quantitative accuracy and (dependent on future instrument development) potentially also the sensitivity of SRM in the upcoming years [77]. One of the main issues that hamper the rapid implementation of DIA to date is the need for spectral libraries. Currently these spectral libraries still need to be created by the user prior to the DIA experiment, increasing acquisition time and required sample amount compared to simple shotgun analysis. Therefore, major developments are needed, either by improving the robustness and ease-of-use of spectral-library free software tools or by creating large online spectral library resources that are easily accessible for all users aiming to perform targeted data extraction.

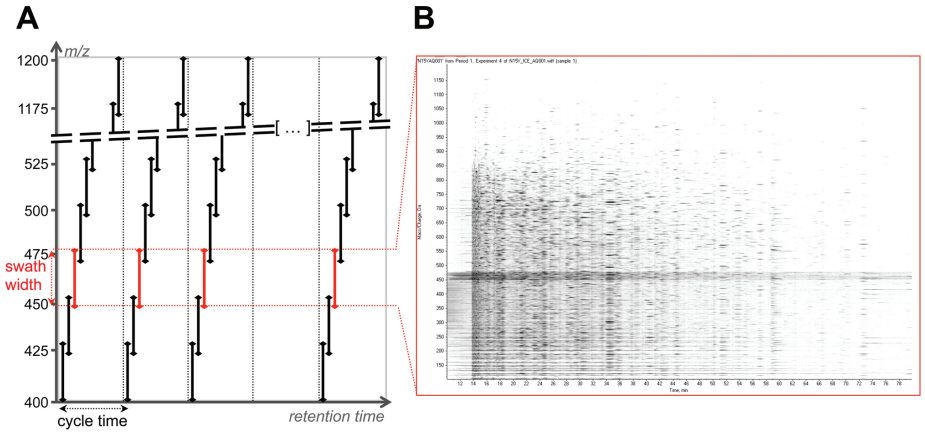


Figure 8: ***Concept of data-independent acquisition****. (A) Acquisition method of a typical DIA analysis employing large isolation windows while keeping cycle times on a chromatographic time scale. (B) Representation of a typical DIA MS/MS map (adapted from Gillet* et al. *[78])*

3. Protein Phosphorylation and its role in signaling

In living organisms, the biological functionality of proteins is not only dependent on their expression, but also on chemical modifications to amino acid side chains or the peptide termini. These chemical modifications are called post-translational modifications (PTM). PTMs can be permanent, but most of them are transient and reversible modifications that are used to fine-tune the functionality of the proteins [4]. To date the best studied PTM is protein phosphorylation [84]. The term phosphorylation refers to the covalent attachment

of a phosphate group to the hydroxyl functionality of serines, threonines and tyrosines. This modification can then alter the structural conformation which in turn can have a variety of effects on the target protein, including its activation [85], deactivation [86], changing its interactors [87] or its intracellular localization [88].

Mediator of protein phosphorylation is a group of enzymes called kinases. They facilitate the transfer of a phosphate from a high-energy, phosphate-donating molecule such as ATP to a specific substrate. The reverse reaction is mediated by a class of proteins called phosphatases. The correct interplay between different kinases and phosphatases enables the biological system to orchestrate a huge variety of biological functions such as cell growth, survival or death, proliferation, intracellular communication and adapting to external conditions and stimuli [89].

3.1. The Kinome: The Protein Kinase Complement of the human genome and its common principles of action

Historically the first evidence for protein phosphorylation by enzymes dates back to the year 1954 [90]. However, the size of the whole complement of human kinases could only be assessed in the course of the human genome project almost 60 years later. In a hallmark paper published in 2002, Manning *et al.* catalogued the protein kinase complement of the human genome [91] by searching for evidence of a conserved domain, the eukaryotic protein kinase catalytic domain or even fragments of it. With this approach they narrowed down the maximum number of human kinases to 518 members, which at this time was lower than expected [92]. Manning *et al.* subsequently divided the human kinome into 5 different groups, each subdivided into different families and subfamilies according to sequence similarity of their catalytic domain but also according to known function or domain structure. For instance, the kinases phosphorylating exclusively tyrosine residues are clustered in the Tyrosine Kinase group while the other groups contain serine/threonine-specific kinases or dual-specificity kinases.

Common to most of the human kinases is the presence of certain conserved domains with highly specific functions. Most importantly, this includes the catalytic domain, which facilitates the phosphorylation reaction and has the shape of a deep narrow cleft. It contains the ATP binding site that buries the adenosine moiety in a hydrophobic pocket and exposes the γ-phosphate to the outer edge where the transfer can take place. The substrate is placed in a region called substrate binding site (SBS) facilitating the correct positioning of the phosphorylatable residue (P0). Catalysis is mediated by opening and closing the cleft [93] and a conserved aspartic acid residue in the catalytic domain [94].

Another highly conserved domain present in many kinases is the so called activation segment [95]. It has typical lengths of 20 to 35 amino acid residues and is surrounded by two conserved tripeptide motifs, Asp-Phe-Gly (DFG) at the N-terminus and Ala-Pro-Glu

(APE) at the C-terminus. The segment in between contains several conserved secondary elements as depicted in Figure 9.

1. The **magnesium binding loop** contains the aspartate of the DFG sequence. Its carboxyl group chelates a magnesium ion that is required to position the phosphate for the phosphate transfer [96].
2. The **β9 loop** forms an antiparallel β sheet with a sequence close to the catalytic loop. This interaction is crucial to keep the catalytic loop in an active state.
3. The **activation loop** is the primary site of kinase activation. In its unphosphorylated state the kinase remains inactive and the loop is disordered. Once it becomes phosphorylated at a specific residue refolding occurs. In this state the phosphorylated residue interacts with a conserved basic pocket (RD pocket) which is directly connected to the catalytic machinery. This interaction helps the activation loop to obtain a stable fold [95]. Next to the above mentioned primary phosphorylation site many kinases have secondary phosphorylation sites in their activation loop. They are only phosphorylated once the primary site is already phosphorylated. These secondary phosphorylated residues do not interact with the RD pocket. Their contribution to the kinase activity differs greatly between different kinases. While some lack secondary phosphorylation sites others require secondary phosphorylation for full activity (e. g. ERK2 [97]). The function of these secondary phosphorylation sites also includes the recruitment of downstream interactor proteins.

   The activation loop communicates primarily through the N-terminal anchor. Even subtle change in its structure can be amplified by the anchor point and will amplify into major structural changes that determine the kinase's activation state.
4. The **P+1 loop** provides the C-terminal anchor of the activation segment. It contains the conserved Ala-Pro-Glu tripeptide.

Substrate recognition is crucial to kinase function as it ensures proper functionality within the context of all cellular processes [98]. From the growing amount of kinase crystal structures, it is known that the catalytic domain of kinases consists of a narrow groove between two lobes. This ensures that the substrate can only enter the catalytic domain in a locally unfolded way. Thus, only the few amino acids flanking the phosphorylation site interact with the kinase's active site. This interaction defines the substrate specificity, a concept named "peptide specificity" [93, 99]. The peptide specificities of many kinases have been determined by *in vitro* kinase reactions in which the purified kinase was incubated with ATP and potential targets. Cell lysates incubated with specific phosphatases [100], purified proteins [101], peptide libraries in solution [102] or peptide microarrays [103] can be used as potential targets. The combined knowledge led to the establishment of linear consensus motifs for most of the kinases. This knowledge has several implications:

1. **Predicting kinase substrates:** Known peptide specificity for kinases facilitates the prediction of potential unknown kinase targets from the genome sequence [104]. Several pieces of software try to exploit this, such as Predkin [105] or Phos3D [106].
2. **Predicting kinase activity**: Large phosphoproteomics data sets can also be analyzed in terms of overrepresentation of certain consensus motifs in order to determine which kinases are likely present in an active state [107].

Nevertheless, these data need to be considered with care, as it is known that peptide specificity is not the only element that controls phosphorylation events in the living cell. For a phosphorylation to happen in a cell, the kinase needs to be able to recruit its substrate. This can happen through direct interactions between the kinase and the substrate via domains outside of the catalytic domain, so called docking domains [108]. Or the recruitment is facilitated by mutually binding to the same scaffolding protein [109]. Examples of these domains are the src homology domains binding to phosphorylated tyrosines [110] and the D-motif that ensure interaction between MAP kinases and their substrates [111]. Another mechanism of helping substrate recruitment is the co-localization of kinase and substrate in a small subregion of the cell [99]. Conversely, a phosphorylation reaction might not take place due to masking or sequestration of the substrate protein.

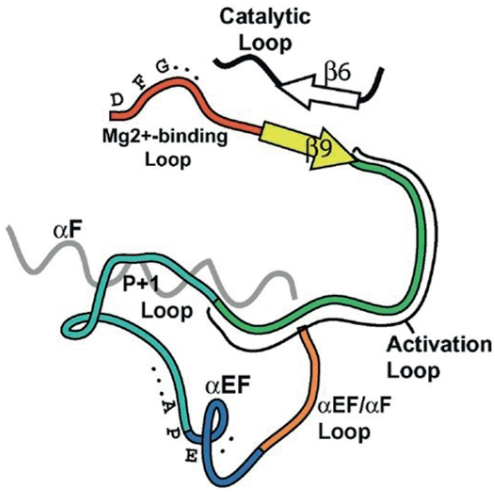


Figure 9: ***Common structure of the kinase activation segment****. The kinase activation segment is flanked by a conserved DFG sequence in the* Mg^{2+}*-binding loop and an APE sequence in the P+1 loop and contains the conserved secondary elements known as* β*9 loop and activation loop (also called t-loop) (adapted from Nolen* et al. *[95]).*

3.2. Kinase function in the cellular context

Protein phosphorylation is one of the key events enabling the cell to interact with other cells or its environment, to adapt to extracellular stimuli or even to influence its surrounding. All these regulatory processes are possible due to the concerted action of specific phosphorylation events in combination with other forms of protein modification such as ubiquitination, degradation, cleavage, etc. Specific kinase substrate interactions rarely act alone but must be seen in the cellular context in which the kinase activity has been switched on due to up-stream events while the substrate phosphorylation leads to downstream actions. Eventually these interactions result in large signaling cascades including multiple players that can influence each other, providing pathways for signal transduction capable of transferring specific (extracellular) stimuli into an appropriate response [112]. The complex organization of most of these signal transduction pathways allows the cell to fine-tune their reaction to various extracellular stimuli. While certain interactions within the network are simply built to relay the signal, the presence of positive feedback loops can amplify signals. Other nodes of the network need several independent inputs before relaying a signal onwards. Negative feedback loops can attenuate signals and branching allows different cell types to react differently to the same signal [113]. Protein phosphorylation by kinases and dephosphorylation by phosphatases are very fast events and thus often play a major role in signaling cascades required to quickly adapt to environment changes.

One of the most studied signaling network involving protein phosphorylation is the control of cell proliferation and growth. It consists of two core signaling cascades the PI3K/AKT/mTOR pathway and the MAPK/ERK pathway. They ultimately both mediate pro-survival and growth signals towards end point effectors such as transcription factors or ribosomal protein kinases.

The PI3K/AKT/mTOR pathway is initiated by growth factors or in response to energy status and amino acid levels. Growth factors can lead to activation of the lipid kinase PI3K that upon activation phosphorylates phosphatidylinositol 4,5-bisphosphate (PIP2) to phosphatidylinositol 3,4,5-trisphosphate (PIP3) leading to recruitment of the protein kinase AKT to the plasma membrane. Due to this change in location, AKT can be activated by the protein kinase PDK1 and a protein complex containing the mTOR protein kinase named mTORC2. Upon activation, AKT phosphorylates multiple downstream reactors involved in survival and proliferation, most prominently TSC2. Phosphorylation of TSC2 stops TSC2 mediated inhibition of RHEB, which in turn activates mTORC1, a multiprotein complex containing the mTOR protein kinase. In its activated state, mTORC1 phosphorylates multiple target proteins involved in proliferation, such as the initiation factor binding protein 4EBP1 and the p70 ribosomal S6 kinase, leading to ribosome biogenesis and translation initiation [114]. Inactivation of PI3K by negative feedback loops or the dephosphorylation

of PIP3 by the PTEN phosphatase can inactivate the signaling cascade and halt cell growth and proliferation (Figure 10A).

The MAPK/ERK pathway is primarily activated by polypeptide hormones, neurotransmitters, chemokines and growth factors. These external stimuli signal through receptor tyrosine kinases, G protein-coupled receptors or by directly activating protein kinase C. This leads to activation of the GTPase RAS that can activate the MAP kinase kinase kinase (MAP3K or RAF). This in turn leads to a phosphorylation cascade in which RAF phosphorylates and activates the MAP kinase kinase (MAP2K or MEK) which in turn phosphorylates the MAP kinase (MAPK or ERK), which phosphorylates effector proteins including transcription factors and ribosomal kinases (Figure 10B) [114].

Over the last decades both pathways leading to cell proliferation have been shown to be highly active in most cancers. Somatic mutations resulting in a gain of function of any component can lead to a permanently active signaling cascade, for instance seen for the MAP3K BRAF in melanoma [115]. Also, the loss of function of proteins that are normally deactivating growth and proliferation signals (tumor suppressors) can lead to constantly active proliferation signaling. The most prominent example is a loss of function in PTEN, which can inhibit PI3K/AKT/mTOR signaling through dephosphorylation of PIP3 and (though to a much lesser extend) IRS1 [116]. PTEN mutations, deletion or inactivation have been shown to be common in most forms of human cancers [117].

Most of the recently developed cancer drugs aim to circumvent these effects by deactivating the hyperactive growth signaling, trying to inhibit key components of the cascade [118]. A deep understanding of the molecular processes supporting cell growth and proliferation in cancer is thus crucial for cancer drug development.

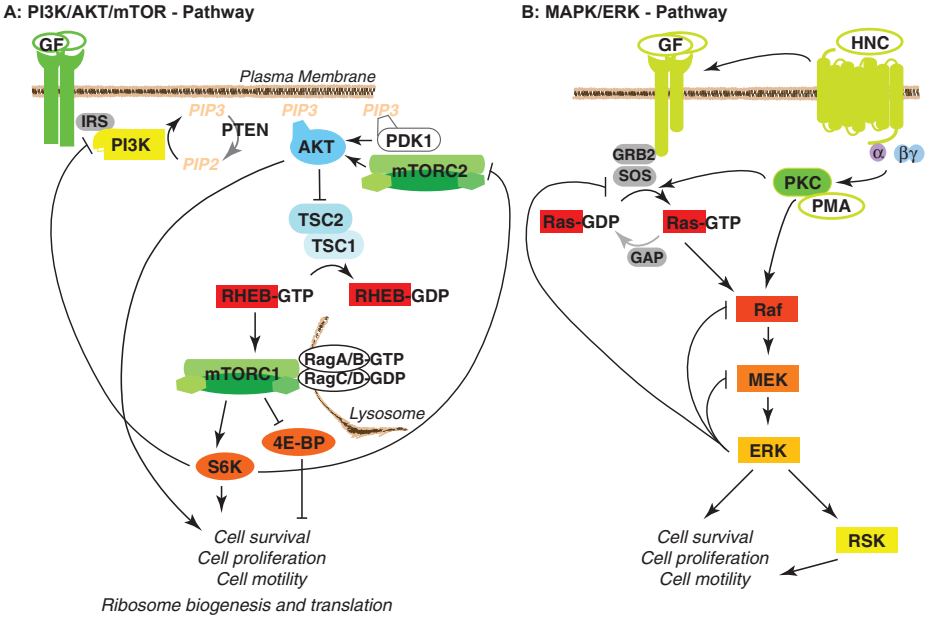


Figure 10: ***Control of cell proliferation and growth.*** *(A) Schematic depiction of the PI3K/AKT/mTOR pathway. Growth factors (GF) can bind to their cognate receptor tyrosine kinase resulting in its autophosphorylation. This results in activation of the PI3K complex which phosphorylates PIP2 to generate PIP3. This results in translocation of AKT and PDK1 to the membrane resulting in an activation of AKT by PDK1 and mTORC2. AKT phosphorylates multiple downstream targets such as TSC2, which upon being phosphorylated stops inhibiting RHEB (through GAP activity). This enables RHEB to activate mTORC1, which in turn phosphorylates multiple downstream targets leading to cell survival and proliferation. (B) Schematic depiction of the MAPK pathway. External stimuli such as e. g. growth factors (GF) or polypeptide hormones, neuropeptides and chemokines (HNC) initiate the MAPK signaling cascade through receptor tyrosine kinases or G protein-coupled receptors respectively. This in turn results in activation of the GTPase RAS upon GF activation initiating a phosphorylation cascade in which RAF phosphorylates MEK, which in turn phosphorylates ERK. PKC also leads to ERK activation, it is however not completely established if this activation occurs via RAF or RAS (adapted from Mendoza* et al. *[114]).*

4. Analysis of Protein Phosphorylation by (Targeted) Mass Spectrometry

Given the vast amount of cellular functions that are controlled by phosphorylation and dephosphorylation, the analysis of the phosphoproteome has been of major interest to research in health and disease. MS-centric methods in combination with phosphopeptide enrichment strategies have become a standard approach to study protein phosphorylation in laboratories worldwide [25]. The following sections will give a short overview of the current state of the art of MS-centric phosphopeptide analysis.

4.1. Sample handling and PTM enrichment

Despite the fact that the human proteome contains a mere 518 protein kinases, it is estimated, that up to a third of the proteins are phosphorylated at a given time [119]. However, these phosphorylation events often occur at a low stoichiometry and are hence hardly detected in proteomic analyses without prior enrichment. Therefore, phosphopeptide specific enrichment protocols have been developed throughout the last decades, with the sensitivity and selectivity enabling phosphoproteomics experiments to be conducted at high throughput and depth in a reasonably short time frame.

Protocols available for phosphopeptide enrichment from complex samples range from immobilized metal affinity chromatography (IMAC), metal oxide affinity chromatography (MOAC), chemical modification [120], up to immunoprecipitation [121].

For the experiments of this thesis, exclusively IMAC based phosphopeptide enrichments protocols were used. IMAC exploits the formation of a coordinate bond between the phosphoryl groups of the phosphopeptides and a metal ion. The metal ions are chelated to a support to form the stationary phase, typically in beads or a column. The phosphopeptide mixture is loaded at a low pH to ensure complete protonation of the peptides' carboxylic groups. Thus, the phosphate groups can bind to the metal ions while non-phosphopeptides will be washed off [122]. However, interference still occurs and moreover, the low pH conditions can result in leaching of the metal ions from the support [123]. Hence, many studies aimed to optimize the protocols resulting in the utilization of various metal ions such as Fe^{3+}, Ga^{3+}, Zr^{4+} or Ti^{4+}, different kinds of chelating solid support materials such as nitrilotriacetic acid (NTA) or phosphonates [15], and different support materials, such as monolith or mesoporous silica and magnetic beads. In this thesis we use Ti^{4+}-IMAC, which has been shown to exceed other methods in terms of selectivity [124] and exhibits a relatively high quantitative and qualitative reproducibility [125] as well as Fe^{3+}-IMAC–NTA facilitating high-throughput enrichment by automation on an AssayMAP Bravo platform while maintaining sensitivity and reproducibility [126].

4.2. Analyzing phosphorylation by shotgun proteomics: current state of the art and challenges

The analysis of phosphopeptide samples with shotgun proteomics has experienced a boost in recent years. The development of new instrumentation has increased both throughput and proteomic depth. Especially instruments such as the Orbitrap Fusion, capable of parallelizing many scan functions, can operate at very high iontrap MS/MS acquisition rates [43]. In combination with isobaric labeling, Erickson *et al.* were able to quantify >38 000 phosphopeptides (11 000 phosphorylation sites) across 10 mouse tissue samples within less than 2 days of analysis time [127]. But also instruments that perform MS and MS/MS in an Orbitrap have improved substantially in scan speed. While early Orbitraps lacked the speed of an ion trap, new high field Orbitraps, as implemented in the Q Exactive HF, can

achieve acquisition rates of up to 20 Hz [128], which is in the same range as achieved with ion trap scans on an Orbitrap Fusion [129]. This has been demonstrated by Kelstrup *et al.* who were able to identify 7 600 unique phosphorylation sites from a phospho-enriched HeLa lysate within 1 hour by using a Q Exactive HF [130]. To date the main consideration of shotgun phosphoproteomics experiments is finding the right balance between throughput and depth. If the aim of the experiment is to achieve proteomics depth, fractionation can be performed prior or after phosphopeptide enrichment [16]. Likewise, the inherent biases of sample preparation protocols can be exploited by performing them in parallel for alike samples. Numerous studies successfully applied this approach. Giansanti *et al.* detected 18 430 unique phosphorylation sites in Jurkat cells by combining results obtained from digestion with five different proteases [131]. Sharma *et al.* identified 51 098 unique phosphorylation sites in HeLa cells by combining fractionation with TiOx and pTyr antibody enrichment [132]. The main drawbacks of these are the required time used for both, sample preparation and MS acquisition time (40 days of measurement for Sharma *et al.*). On the other hand, if throughput is prioritized over depth, single-shot approaches are the method of choice. Combining single-stage phosphopeptide enrichment with one-dimensional online chromatography results in limited time requirements for sample preparation and LC-MS analysis, thus enabling the rapid analysis of a large number of samples within a short time. This approach was successfully implemented for example by de Graaf *et al.* who detected 12 799 phosphorylation sites in 108 samples from differently treated Jurkat cells, by combining single-stage Ti^{4+}-IMAC with 2h LC-MS acquisition time [125]. Key to being able to acquire high throughput is the availability of robust and reproducible protocols for sample preparation, preferably in an automated way. Such examples have been presented by Post *et al.* performing phosphopeptide enrichment on Fe^{3+}-IMAC cartridges on an AssayMAP Bravo platform [126], and by Tape *et al.* combining magnetic Ti^{4+}-IMAC microspheres with a magnetic particle-handling robot [133].

Despite these technical advances multiple challenges remain associated with the analysis of phosphopeptides by shotgun MS. A first challenge is given by the fragmentation characteristics of phosphopeptides. Phosphate-groups are by nature labile and often give rise to a neutral loss of phosphoric acid (H_3PO_4 – 98Da) upon fragmentation [47]. Especially in slow heating LIT-CID experiments, the first fragmentation occurring is the cleavage of the phosphate bond, leading to a drastic reduction of sequence information from backbone fragmentations. Additionally, it has been demonstrated that protonated phosphopeptides tend to undergo intermolecular transfer of the phosphate group if they are subjected to longer activation times (> 1 ms), leading to false information about the localization [134, 135]. Other fragmentation techniques such as HCD or ETD suffer less from neutral loss events, especially ETD is primarily giving rise to sequence-informative fragments while retaining the phosphate group. This has been shown to increase the confidence of the phosphorylation site localization [136]. Similar effects can be observed by EThcD [137].

However, the most commonly used fragmentation technique used for phosphoproteomics is HCD, as it is easy to implement, fast, and due to the higher energy deposition compared to LIT-CID the neutral loss problems are largely overcome [138].

A second challenge is the confident assignment of the phosphorylation site localization on the peptide sequence. Phosphopeptides with identical amino acid sequence and the same amount of phosphorylation do not differ in MS1 space, regardless of the position of the phosphorylated residue. Hence, only MS/MS information can be exploited in order to localize the exact position of the phosphorylation site. Assuming the phosphorylation is retained at the exact location on the peptide sequence upon fragmentation, certain fragment ions can provide information about the location of the phosphorylation sites within the peptide [139]. Depending on the proximity of the respective phosphorylation sites, the peptides however can still give rise to a majority of identical fragment ions (Figure 11).

Several software solutions have been developed that tackle the issue of phosphorylation site localization, the most prominent ones being Ascore [141], phosphoRS [142], MD-score [143], and PTM-score [144]. They differ slightly in their exact function, but as a common feature they all use the identified peptide sequences to generate lists of all possible isoforms by permutation and to assign these isoforms with scores and probabilities, eventually aiming to differentiate between correct and incorrect site localizations. All four software solutions heavily rely on the presence of site-determining ions. Similar to the FDR used for database searches some of these software solutions rely on a false localization rate (FLR).

A third challenge is given by the chromatographic system. The extent to which phosphorylation site location isomers are separated by the used chromatography is highly different depending on the peptide sequence, the proximity of the phosphorylation sites, the phosphorylated residue and the quality of the chromatographic system used. While Marx *et al.* claimed that up to 60 % of phosphoisomers are chromatographically indistinguishable [145], other studies (including **Chapter 3** of this thesis) achieve decent chromatographic separation for some or even the majority of phosphorylation site location isomers [146-150]. This has potentially severe implication for the identification and quantification of phosphorylation site localization isomers. If different localization isomers co-elute perfectly, they will also be co-fragmented as they have identical precursor *m/z*, giving rise to multiplexed MS/MS spectra [151]. This can lead to the presence of site-determining ions for different isomers within the same spectrum, hampering proper identification. Likewise, every attempt of MS1 based quantification is affected, as the isomers cannot be differentiated in MS1 space. If different phosphorylation site location isomers differ only slightly in retention time, the exact settings of the dynamic elution can hinder the acquisition of MS/MS spectra for the isoform eluting later [151].

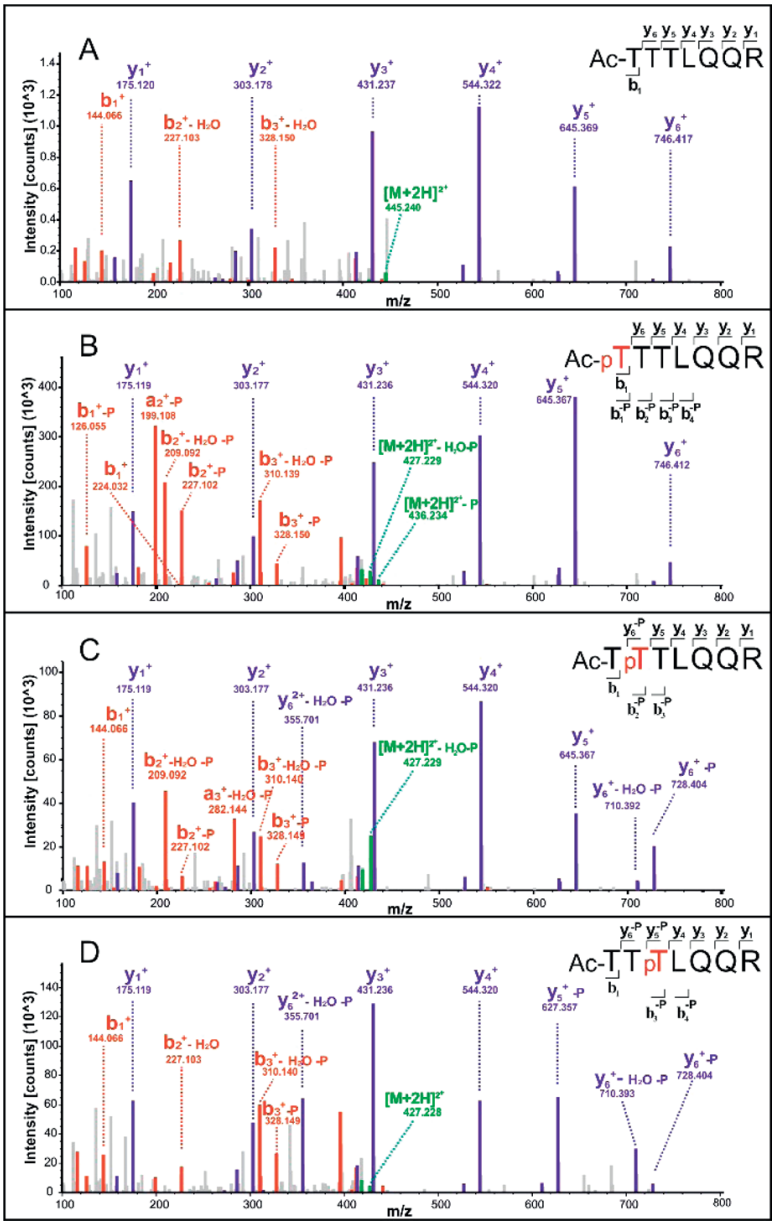


Figure 11: ***MS/MS analysis of phosphorylation site localization isomers.*** *Fragmentation patterns of different phosphoisomers result in a large number of identical fragment ions, challenging correct phosphosites localization. Panel A shows unmodified peptide of the same sequence (adapted from Angeleri* et al. *[140]).*

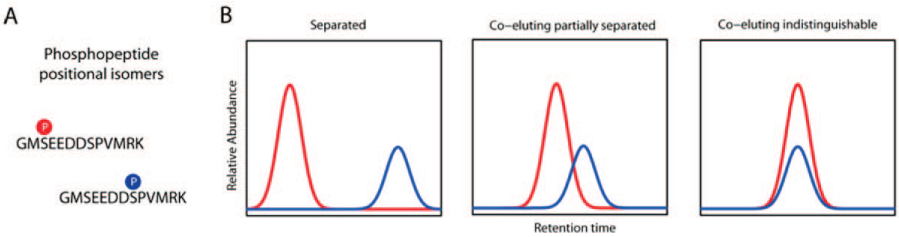


Figure 12: ***Chromatographic behavior of phosphoisomers.*** *(A) Phosphorylation site localization isomers contain the same peptide sequence and number of phosphorylated residues but differ in their position. (B) Phosphorylation site localization isomers can show different chromatographic behavior in that they are either completely or partially separated or co-elute. A lack of chromatographic separation can hamper MS/MS identification and MS1 quantification of different phosphoisomers (adapted from Courcelles* et al. *[151]).*

A fourth challenge of phosphoproteomics is hidden in the data interpretation. While the amount of detected phosphorylation sites grows constantly, only a very limited number of them is associated to a biological function. Moreover, many phosphorylation sites with known functions proof to be too low abundant to be detected in a typical LC-MS experiment. This leads to a growing discrepancy in the targets detected by proteomics and the targets directly relevant for biologists and clinicians [152]. Moreover, the lack of specific information about the functionality of the majority of detected phosphorylation sites often results in oversimplifications in the data analysis by the researches (e. g. assumptions such as: phosphorylation means activation).

4.3. Advantages of protein phosphorylation by Targeted Mass Spectrometry

Approaches using MS-based targeted proteomics have the potential to address several of the above-mentioned challenges. The hypothesis driven nature of targeted experiments already circumvents the requirement to interpret arbitrary information about uncharacterized phosphorylation sites and shifts the focus to highly accurate and reproducible quantification of smaller subsets of phosphorylation sites of interest [153, 154]. Careful assay optimization can aid to alleviate the above mentioned technical complications faced in shotgun proteomics such as neutral loss and phosphorylation site localization. Fragmentation events giving rise to sequence informative ions can be selectively optimized for each peptide of interest [140]. Continuous chromatographic traces of phosphoisomer specific fragment ions can be used for differential quantification in a site specific manner, which is not necessarily possible with MS1 based quantification unless the phosphoisomers are chromatographically perfectly separated [151].

The very few studies applying targeted MS to the analysis of phosphorylation, showcase the unique niche target analysis of phosphopeptides can provide. In the simplest case, SRM can be applied for the validation of potential biomarkers detected in discovery experiments. This has for instance been successfully applied by Narumi *et al.* in 2012 who used

a complete biomarker pipeline to human breast cancer tissues, validating four potential biomarker peptides inaccessible to antibody detection [155].

Other studies used targeted methods specifically to approach common challenges faced in the analysis of phosphopeptides by DDA. Most of them exploit the benefits of the improved reproducibility and sensitivity comparted to shotgun MS. The following sections present a few recently performed landmark studies applying SRM or DIA to phosphoproteomics.

A very early proof of principle study by Gerber *et al*. dates back to 2003 [71]. In there they demonstrate the use of AQUA peptides for the quantitative analysis of one specific phosphorylation site in HeLa cells in a cell cycle dependent manner. It marked the first attempt to use the AQUA strategy to absolutely quantify dynamic changes directly from cell lysates.

In 2010 Jin *et al.* used SRM to determine phosphorylation stoichiometry of two phosphorylation sites in the Lyn kinase. Therefore, they used synthetic peptides of known concentrations to determine the MS signal response rates for the phosphorylated and the unphosphorylated peptides of interest. Subsequently, they performed anti-Lyn IPs on HEK cells and multiple melanoma xenographs and were able to accurately determine phosphorylation stoichiometry detecting changes even at stoichiometries as low as < 1 % [156].

A first large scale SRM study in phosphopeptides has been presented in 2007 by Wolf-Yadlin *et al*. They used SRM for an in-depth analysis of 222 tyrosine phosphorylated peptides in immortalized normal human mammary epithelial cells upon EGF treatment [154]. Most subsequent studies focus on the analysis of very specific signaling pathways or biological processes. For example, Lam *et al.* performed a two-part study in which they applied SRM to quantify a total of 76 phosphorylation sites of metabolic enzymes. They performed mitochondria isolation in combination with phosphopeptide enrichment from different tissues such as mice liver and heart as well as human hearts obtained from heart transplantation patients. The study stands out in terms of carefully addressing several challenges including methionine oxidation, enzymatic cleavage efficiency and the presence of site specific ions resulting in the robust quantification of several site localization isomers [157, 158]. A similar approach focusing on PI3K-mTOR/MAPK pathway dynamics in senescent cells has been performed by de Graaf *et al.* in 2015. Here, especially the combination of highly selective phosphopeptide enrichment by Ti^{4+}-IMAC with SRM facilitates the successful detection of low abundant phosphopeptides without the need for further reduction of the sample complexity with fractionation [153].

A very elegant approach of using SRM in combination with phosphopeptide enrichment has been presented by Soste *et al.* in 2014 in their so called sentinel protein assay [159]. The aim of the study was to directly link protein or phosphoprotein abundance with biological activity or function. Careful literature mining resulted in the creation of a list of more than 500 so-called sentinels for which quantification can be directly associated to a

specific biological process or status in yeast. This resulted in a fast and accurate method that can be used to analyze responses to large sets of uncharacterized perturbations in high throughput. The set includes a total of 152 phosphopeptides that are used as quantifier for multiple processes including osmotic stress, reduced mTORC activation (e. g. after rapamycin treatment) and glucose starvation. Several of the targeted phosphorylation sites are activation loop phosphorylations in kinases. The sentinel assay proofs to be complementary to shotgun proteomics with advantages in terms of its high-confidence and direct biological interpretability. It exploits (and requires) the precision of SRM that is so far unique to hypothesis-driven MS experiments.

Recently also first studies combining phosphopeptide enrichment and DIA measurements started to appear. A first proof of principle study has been published by Parker *et al.* in 2015 [55]. In there, they combine TiO_2 phosphopeptide enrichment with DIA-MS measurement to study components of the insulin signaling pathway upon insulin stimulation in combination with different inhibitors. As a proof of principle for the applicability of DIA to phosphoproteomics they used stable isotope labelled internal standards for 15 of their target phosphopeptides. This allowed them to carefully assess the quantitative characteristics of DIA such as linear range and reproducibility, showing the enhanced quantitative accuracy of DIA over MS1 based label-free quantification. The approach is further used to extract and quantify further potentially interesting phosphorylation sites, mainly known and predicted AKT and mTOR substrates.

Altogether these studies show dedicated applications where targeted MS analysis outperforms DDA in terms of accuracy, reproducibility, robustness and differential quantification of phosphosite isomers, depending on the question at hand. It thus provides an excellent intermediate step between the unspecific screening approach provided by DDA and the highly specific characterization performed by antibody-based research.

5. Neuropeptides as alternative signaling molecules and their analysis by MS-based Proteomics

While protein phosphorylation plays a major role in intracellular signaling, various alternative forms of signaling have been discovered and extensively studied. Especially intracellular signaling in multicellular organisms requires different means of signaling as the signal needs to cover larger distances. This is primarily achieved by the secretion of molecules either locally or into the systemic circulation. To date three major classes of secreted signaling molecules have been defined, which are hormones, cytokines and neurotransmitters. The latter are crucial messengers to transmit signals within the nervous system, primarily by secretion into the synaptic cleft. Neurotransmitters can be small molecules such as serotonin or acetylcholine, but a large part of them are peptides, so called neuropeptides.

These neuropeptides play important roles in many physiological processes including energy balance, pain, and memory [160]. A dysregulation in neuropeptide signaling can thus lead to various symptoms such as obesity or diabetes, rendering them an important research topic [161, 162].

However, the biochemical analysis of neuropeptides is drastically complicated through the way of their biosynthesis. They are derived from larger precursor proteins through multiple sequential enzymatic cleavage events and post translational modifications. This can result in various neuropeptides originating from the same precursor protein that can have different biological activities depending on their PTMs (Figure 13).

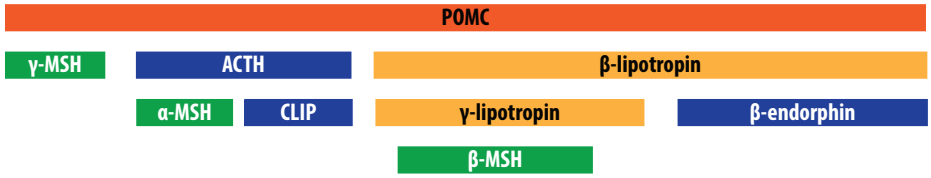


Figure 13: ***Neuropeptide biosynthesis.*** *Sequential trimming events during the biosynthesis of neuropeptides exemplified with the precursor POMC.*

The variety of possible trimming and modification events impedes the *de novo* prediction of the active neuropeptides from the genome or even the precursor protein [163]. Therefore, MS based approaches have been developed to become the methods of choice for the analysis of neuropeptides, as they have the potential to cope with the large variety at hand. Usually, they employ an approach in which neuropeptides are specifically isolated from tissue and analyzed by LC-MS/MS. A crucial step is to avoid enzymatic digestion as this results in a loss of information about the respective trimming product at hand. Compared to classical proteomic analysis, the sample preparation protocols for neuropeptidomics usually only employ peptide extraction steps in combination with molecular weight cutoff filters to separate endogenous peptides from intact proteins. Additional precautions have to be taken to avoid post-mortem degradation or enzymatic processing of the neuropeptides throughout sample preparation. This can usually be avoided through rapid boiling or irradiation of the tissue [164, 165].

The extraction of endogenous peptides however results in samples that differ substantially from tryptic digests in terms of peptide length distribution and amino acid composition. This complicates the MS/MS analysis as well as the data analysis. While traditional fragmentation techniques such as CID are much less suited to fragment neuropeptides, only the growing use of electron driven fragmentation techniques has opened the door for deeper analyses of the neuropeptidome. The lack of tryptic digest also hampers the subsequent database analysis, as the presence of well-defined cleavage sites is a crucial step in spectrum annotation. For neuropeptide analysis the naturally occurring cleaving

mechanisms are mostly not known, therefore the only way to perform a database search is to allow for every potential cleavage. This is time consuming and usually associated with an increase in false-positive identifications.

In recent years many studies have successfully employed LC-MS/MS for the analysis of neuropeptides. In 2013 Frese *et al.* have studied neuropeptide changes upon various food intake regimes in rats. By employing a combination of HCD and ETD with label-free quantification they were able to identify more than 1 700 neuropeptides in different nervous tissues, including previously unreported trimming variants [166]. Another study published in 2013 by Lee *et al.* investigated alterations in neuropeptide levels depending on circadian rhythm. They analyzed neuropeptides from punch samples of the suprachiasmatic nucleus of rats taken at different times of the day with shotgun LC-MS and label-free quantification. A total of 190 known endogenous peptides were reproducibly quantified across all samples, 10 of which showed previously unknown circadian-related alteration [167]. Besides rats the neuropeptidome of other organisms have been investigated alike, including crabs [168, 169] and insects [170]. Beside label-free shotgun LC-MS also targeted MS-based proteomics methods have been increasingly used in the field of neuropeptidomics. A proof of principle study using DIA for neuropeptide quantification has been published by Schmerberg *et al.* in 2015 [171], investigating alterations in the neuropeptidome of crabs at different times after food intake. SRM and MRM^3 have so far been primarily used for the quantification of neuropeptides with very similar characteristics to tryptic peptides in terms of length and amino acid composition [172-175]. Extending these targeted MS-based methods towards longer neuropeptide sequences has only recently been accomplished and will be described in **Chapter 2** of this thesis.

The boost in neuropeptidomic research in recent years led to an ever-growing list of potential neuropeptides in various systems. Key to a successful exploration of this knowledge, however is a further in-depth characterization of numerous targets. At the end of the day, LC-MS can only report which peptides are present in a specific biological system under a specific condition. It can usually not reveal its biological function, nor can it differentiate between naturally occurring peptides or effects of sampling artefacts or post-mortem degradation products. Specifically, the sequential truncation of peptides during their biosynthesis can result in the presence of several dozens of prohormones that do not necessarily reflect any direct biological significance. In combination with subsequent biochemical validation, however, the field of neuropeptidomics might eventually provide valuable new knowledge for diagnostics or drug development [176].

Outline

This thesis describes method development and applications in the area of MS-based targeted proteomics geared towards robust quantification of molecular signaling events. The two core tools used throughout the following chapters are SRM and DIA.

SRM has been successfully used for decades to analyze small molecules and tryptic peptides. Nonetheless, this thesis demonstrates novel applications of this very established tool. **Chapter 2** demonstrates how the SRM technology can be exploited to quantify neuropeptides in hypothalami of mice. Throughout the chapter various steps are described that need to be implemented to successfully measure highly charged peptides on a triple quadrupole mass spectrometer. Eventually the SRM technology is exploited to quantify neuropeptide levels in animals subjected to different diets. **Chapter 3** shows a functionality driven approach to SRM. The unprecedented sensitivity of SRM in combination with highly specific phosphopeptide enrichment is exploited to quantify low abundant phosphorylation events in kinase t-loops. Due to the known functionality of these phosphorylation events, their quantitative readout can be used as a direct probe for kinase activity. Throughout the chapter, the high multiplexing capabilities of SRM are exploited to build a system wide kinase activity probe covering one third of the human kinome. This probe is then applied to various cell types (including primary and patient derived), resulting in a compendium of 66 measured t-loop phosphorylation sites.

DIA on the other hand has only become a relevant force in proteomics throughout the last 5 years. Thus, the door for the MS community is open for extensive method development, optimization, refinement, characterization, etc. Large parts of this thesis are dedicated to DIA method development and assessment. **Chapter 4** describes an in-depth performance analysis of DIA in comparison to SRM for the analysis of phosphorylation dynamics in the PI3K-mTOR and MAPK signaling pathways. Throughout the chapter key characteristics of both methods in terms of sensitivity, flexibility and quantitative accuracy are demonstrated, thus providing valuable information for successful project planning. Additionally, the added benefit of MRM^3 for the quantification of low abundant targets is illustrated. **Chapter 5** exclusively focuses on method development and characterization of DIA, specifically in combination with electron driven fragmentation methods. The chapter illustrates how the use of various fragmentation methods, window sizes and cycle times can influence the success rate and sensitivity of a DIA experiment. Additionally, it illustrates major shortcomings in detecting different phosphorylation site localization isomers when using spectral library based DIA data analysis, including an extensive discussion about potential ways to circumvent them.

Chapter 6 provides a concise outlook, containing my personal reflection on the future of mass-spectrometry and proteomics, specifically in the area of data-independent acquisition.

References

[1] Crick, F., Central dogma of molecular biology. *Nature* 1970, *227*, 561-563.

[2] James, P., Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys* 1997, *30*, 279-331.

[3] Altelaar, A. F., Munoz, J., Heck, A. J., Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 2013, *14*, 35-48.

[4] Pagel, O., Loroch, S., Sickmann, A., Zahedi, R. P., Current strategies and findings in clinically relevant post-translational modification-specific proteomics. *Expert Rev Proteomics* 2015, *12*, 235-253.

[5] Godovac-Zimmermann, J., Kleiner, O., Brown, L. R., Drukier, A. K., Perspectives in spicing up proteomics with splicing. *Proteomics* 2005, *5*, 699-709.

[6] Gingras, A. C., Gstaiger, M., Raught, B., Aebersold, R., Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* 2007, *8*, 645-654.

[7] Ding, Q., Dimayuga, E., Keller, J. N., Oxidative damage, protein synthesis, and protein degradation in Alzheimer's disease. *Curr Alzheimer Res* 2007, *4*, 73-79.

[8] Majovsky, P., Naumann, C., Lee, C. W., Lassowskat, I*., et al.*, Targeted proteomics analysis of protein degradation in plant signaling on an LTQ-Orbitrap mass spectrometer. *J Proteome Res* 2014, *13*, 4246-4258.

[9] Li, H., Han, J., Pan, J., Liu, T*., et al.*, Current Trends in Quantitative Proteomics -- an update. *J Mass Spectrom* 2017.

[10] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198-207.

[11] Pasing, Y., Colnoe, S., Hansen, T., Proteomics of hydrophobic samples: Fast, robust and low-cost workflows for clinical approaches. *Proteomics* 2017, *17*.

[12] Olsen, J. V., Ong, S. E., Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* 2004, *3*, 608-614.

[13] Altelaar, A. F., Mohammed, S., Brans, M. A., Adan, R. A., Heck, A. J., Improved identification of endogenous peptides from murine nervous tissue by multiplexed peptide extraction methods and multiplexed mass spectrometric analysis. *J Proteome Res* 2009, *8*, 870-876.

[14] Manadas, B., Mendes, V. M., English, J., Dunn, M. J., Peptide fractionation in proteomics approaches. *Expert Rev Proteomics* 2010, *7*, 655-663.

[15] Zhou, H., Ye, M., Dong, J., Corradini, E*., et al.*, Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat Protoc* 2013, *8*, 461-480.

[16] Batth, T. S., Francavilla, C., Olsen, J. V., Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res* 2014, *13*, 6176-6186.

[17] Molnár, I., Horváth, C., Separation of amino acids and peptides on non-polar stationary phases by high-performance liquid chromatography. *J Chromatogr* 1977, *142*, 623-640.

[18] Escher, C., Reiter, L., MacLean, B., Ossola, R*., et al.*, Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* 2012, *12*, 1111-1121.

[19] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. *Anal Chem* 2005, *77*, 6426-6434.

[20] Wang, Y., Yang, F., Gritsenko, M. A., Clauss, T*., et al.*, Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 2011, *11*, 2019-2026.

[21] Washburn, M. P., Wolters, D., Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001, *19*, 242-247.

[22] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., *et al.*, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999, *17*, 676-682.

[23] Di Palma, S., Hennrich, M. L., Heck, A. J., Mohammed, S., Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J Proteomics* 2012, *75*, 3791-3813.

[24] Thingholm, T. E., Jensen, O. N., Larsen, M. R., Analytical strategies for phosphoproteomics. *Proteomics* 2009, *9*, 1451-1468.

[25] Riley, N. M., Coon, J. J., Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal Chem* 2016, *88*, 74-94.

[26] Kim, S. C., Sprung, R., Chen, Y., Xu, Y., *et al.*, Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* 2006, *23*, 607-618.

[27] Yin, X., Bern, M., Xing, Q., Ho, J., *et al.*, Glycoproteomic analysis of the secretome of human endothelial cells. *Mol Cell Proteomics* 2013, *12*, 956-978.

[28] Zhu, J., Sun, Z., Cheng, K., Chen, R., *et al.*, Comprehensive mapping of protein N-glycosylation in human liver by combining hydrophilic interaction chromatography and hydrazide chemistry. *J Proteome Res* 2014, *13*, 1713-1721.

[29] Xu, G., Deglincerti, A., Paige, J. S., Jaffrey, S. R., Profiling lysine ubiquitination by selective enrichment of ubiquitin remnant-containing peptides. *Methods Mol Biol* 2014, *1174*, 57-71.

[30] Steen, H., Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 2004, *5*, 699-711.

[31] Taylor, G., Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 1964, pp. 383-397.

[32] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989, *246*, 64-71.

[33] Karas, M., Bahr, U., Dülcks, T., Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius J Anal Chem* 2000, *366*, 669-676.

[34] Savaryn, J. P., Toby, T. K., Kelleher, N. L., A researcher's guide to mass spectrometry-based proteomics. *Proteomics* 2016, *16*, 2435-2443.

[35] Douglas, D. J., Frank, A. J., Mao, D., Linear ion traps in mass spectrometry. *Mass Spectrom Rev* 2005, *24*, 1-29.

[36] Cotter, R. J., Peer Reviewed: The New Time-of-Flight Mass Spectrometry. *Anal Chem* 1999, *71*, 445A-451A.

[37] Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* 2000, *72*, 1156-1162.

[38] Lange, V., Picotti, P., Domon, B., Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* 2008, *4*, 222.

[39] Jeudy, J., Salvador, A., Simon, R., Jaffuel, A., *et al.*, Overcoming biofluid protein complexity during targeted mass spectrometry detection and quantification of protein biomarkers by MRM cubed (MRM3). *Anal Bioanal Chem* 2014, *406*, 1193-1200.

[40] Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., Muddiman, D. C., Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal Chem* 2011, *83*, 5442-5446.

[41] Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., *et al.*, Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* 2011, *10*, M111.011015.

[42] Michalski, A., Damoc, E., Lange, O., Denisov, E., *et al.*, Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* 2012, *11*, O111.013698.

[43] Senko, M. W., Remes, P. M., Canterbury, J. D., Mathur, R., *et al.*, Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal Chem* 2013, *85*, 11710-11714.

[44] Bereman, M. S., MacLean, B., Tomazela, D. M., Liebler, D. C., MacCoss, M. J., The development of selected reaction monitoring methods for targeted proteomics via empirical refinement. *Proteomics* 2012, *12*, 1134-1141.

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

[46] Waldera-Lupa, D. M., Stefanski, A., Meyer, H. E., Stühler, K., The fate of b-ions in the two worlds of collision-induced dissociation. *Biochim Biophys Acta* 2013, *1834*, 2843-2848.

[47] Boersema, P. J., Mohammed, S., Heck, A. J., Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* 2009, *44*, 861-878.

[48] Zubarev, R. A., Kelleher, N. L., McLafferty, F. W., J. Am. Chem. Soc., J. Am. Chem. Soc. 1998, pp. 3265-3266.

[49] Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 2004, *101*, 9528-9533.

[50] Nesvizhskii, A. I., Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol Biol* 2007, *367*, 87-119.

[51] Cottrell, J. S., Protein identification using MS/MS data. *J Proteomics* 2011, *74*, 1842-1851.

[52] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551-3567.

[53] Elias, J. E., Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007, *4*, 207-214.

[54] Chelius, D., Bondarenko, P. V., Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* 2002, *1*, 317-323.

[55] Parker, B. L., Yang, G., Humphrey, S. J., Chaudhuri, R., *et al.*, Targeted phosphoproteomics of insulin signaling using data-independent acquisition mass spectrometry. *Sci Signal* 2015, *8*, rs6.

[56] Michalski, A., Cox, J., Mann, M., More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* 2011, *10*, 1785-1793.

[57] Picotti, P., Rinner, O., Stallmach, R., Dautel, F., *et al.*, High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods* 2010, *7*, 43-46.

[58] Mallick, P., Schirle, M., Chen, S. S., Flory, M. R., *et al.*, Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* 2007, *25*, 125-131.

[59] Tang, H., Arnold, R. J., Alves, P., Xun, Z., *et al.*, A computational approach toward label-free protein quantification using predicted peptide detectability. *Bioinformatics* 2006, *22*, e481-488.

[60] Toprak, U. H., Gillet, L. C., Maiolica, A., Navarro, P., *et al.*, Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Mol Cell Proteomics* 2014, *13*, 2056-2071.

[61] de Graaf, E. L., Altelaar, A. F., van Breukelen, B., Mohammed, S., Heck, A. J., Improving SRM assay development: a global comparison between triple quadrupole, ion trap, and higher energy CID peptide fragmentation spectra. *J Proteome Res* 2011, *10*, 4334-4341.

[62] Unwin, R. D., Griffiths, J. R., Whetton, A. D., A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). *Nat Protoc* 2009, *4*, 870-877.

[63] Maclean, B., Tomazela, D. M., Abbatiello, S. E., Zhang, S.*, et al.*, Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal Chem* 2010, *82*, 10116-10124.

[64] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M.*, et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, *26*, 966-968.

[65] Krokhin, O. V., Craig, R., Spicer, V., Ens, W.*, et al.*, An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol Cell Proteomics* 2004, *3*, 908-919.

[66] Röst, H., Malmström, L., Aebersold, R., A computational tool to detect and avoid redundancy in selected reaction monitoring. *Mol Cell Proteomics* 2012, *11*, 540-549.

[67] Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W.*, et al.*, Global analysis of protein expression in yeast. *Nature* 2003, *425*, 737-741.

[68] Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K.*, et al.*, High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* 2007, *6*, 1809-1817.

[69] Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C.*, et al.*, Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol Cell Proteomics* 2014, *13*, 907-917.

[70] Hüttenhain, R., Soste, M., Selevsek, N., Röst, H.*, et al.*, Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci Transl Med* 2012, *4*, 142ra194.

[71] Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* 2003, *100*, 6940-6945.

[72] Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R.*, et al.*, mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods* 2011, *8*, 430-435.

[73] https://skyline.ms/wiki/home/software/Skyline/page.view?name=tutorial_peak_picking.

[74] Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., Coon, J. J., Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics* 2012, *11*, 1475-1488.

[75] Tong, L., Zhou, X. Y., Jylha, A., Aapola, U.*, et al.*, Quantitation of 47 human tear proteins using high resolution multiple reaction monitoring (HR-MRM) based-mass spectrometry. *J Proteomics* 2015, *115*, 36-48.

[76] Gallien, S., Bourmaud, A., Kim, S. Y., Domon, B., Technical considerations for large-scale parallel reaction monitoring analysis. *J Proteomics* 2014, *100*, 147-159.

[77] Vidova, V., Spacil, Z., A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Anal Chim Acta* 2017, *964*, 7-23.

[78] Gillet, L. C., Navarro, P., Tate, S., Röst, H.*, et al.*, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* 2012, *11*, O111.016717.

[79] Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A., Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* 2004, *1*, 39-45.

[80] Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L.*, et al.*, OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* 2014, *32*, 219-223.

[81] Tsou, C. C., Avtonomov, D., Larsen, B., Tucholska, M.*, et al.*, DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* 2015, *12*, 258-264, 257 p following 264.

[82] Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W.*, et al.*, Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods* 2013, *10*, 744-746.

[83] Guo, T., Kouvonen, P., Koh, C. C., Gillet, L. C.*, et al.*, Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat Med* 2015, *21*, 407-413.

[84] Day, E. K., Sosale, N. G., Lazzara, M. J., Cell signaling regulation by protein phosphorylation: a multivariate, heterogeneous, and context-dependent process. *Curr Opin Biotechnol* 2016, *40*, 185-192.

[85] Casar, B., Pinto, A., Crespo, P., Essential role of ERK dimers in the activation of cytoplasmic but not nuclear substrates by ERK-scaffold complexes. *Mol Cell* 2008, *31*, 708-721.

[86] Dougherty, M. K., Müller, J., Ritt, D. A., Zhou, M.*, et al.*, Regulation of Raf-1 by direct feedback phosphorylation. *Mol Cell* 2005, *17*, 215-224.

[87] Butch, E. R., Guan, K. L., Characterization of ERK1 activation site mutants and the effect on recognition by MEK1 and MEK2. *J Biol Chem* 1996, *271*, 4230-4235.

[88] Averna, M., de Tullio, R., Passalacqua, M., Salamino, F.*, et al.*, Changes in intracellular calpastatin localization are mediated by reversible phosphorylation. *Biochem J* 2001, *354*, 25-30.

[89] Singh, V., Ram, M., Kumar, R., Prasad, R.*, et al.*, Phosphorylation: Implications in Cancer. *Protein J* 2017.

[90] Burnett, G., Kennedy, E. P., The enzymatic phosphorylation of proteins. *J Biol Chem* 1954, *211*, 969-980.

[91] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., Sudarsanam, S., The protein kinase complement of the human genome. *Science* 2002, *298*, 1912-1934.

[92] Hunter, T., A thousand and one protein kinases. *Cell* 1987, *50*, 823-829.

[93] de Oliveira, P. S., Ferraz, F. A., Pena, D. A., Pramio, D. T.*, et al.*, Revisiting protein kinase-substrate interactions: Toward therapeutic development. *Sci Signal* 2016, *9*, re3.

[94] Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Ashford, V. A.*, et al.*, Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 1991, *253*, 407-414.

[95] Nolen, B., Taylor, S., Ghosh, G., Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell* 2004, *15*, 661-675.

[96] Adams, J. A., Kinetic and catalytic mechanisms of protein kinases. *Chem Rev* 2001, *101*, 2271-2290.

[97] Prowse, C. N., Deal, M. S., Lew, J., The complete pathway for catalytic activation of the mitogen-activated protein kinase, ERK2. *J Biol Chem* 2001, *276*, 40817-40823.

[98] Turk, B. E., Understanding and exploiting substrate recognition by protein kinases. *Curr Opin Chem Biol* 2008, *12*, 4-10.

[99] Kobe, B., Kampmann, T., Forwood, J. K., Listwan, P., Brinkworth, R. I., Substrate specificity of protein kinases and computational prediction of substrates. *Biochim Biophys Acta* 2005, *1754*, 200-209.

[100] Xue, L., Wang, W. H., Iliuk, A., Hu, L.*, et al.*, Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. *Proc Natl Acad Sci U S A* 2012, *109*, 5615-5620.

[101] Ptacek, J., Devgan, G., Michaud, G., Zhu, H.*, et al.*, Global analysis of protein phosphorylation in yeast. *Nature* 2005, *438*, 679-684.

[102] Mok, J., Kim, P. M., Lam, H. Y., Piccirillo, S.*, et al.*, Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Sci Signal* 2010, *3*, ra12.

[103] Schutkowski, M., Reimer, U., Panse, S., Dong, L.*, et al.*, High-content peptide microarrays for deciphering kinase specificity and biology. *Angew Chem Int Ed Engl* 2004, *43*, 2671-2674.

[104] Trost, B., Kusalik, A., Computational prediction of eukaryotic phosphorylation sites. *Bioinformatics* 2011, *27*, 2927-2935.

[105] Saunders, N. F., Kobe, B., The Predikin webserver: improved prediction of protein kinase peptide specificity using structural information. *Nucleic Acids Res* 2008, *36*, W286-290.

[106] Durek, P., Schudoma, C., Weckwerth, W., Selbig, J., Walther, D., Detection and characterization of 3D-signature phosphorylation site motifs and their contribution towards improved phosphorylation site prediction in proteins. *BMC Bioinformatics* 2009, *10*, 117.

[107] Ding, V. M., Boersema, P. J., Foong, L. Y., Preisinger, C.*, et al.*, Tyrosine phosphorylation profiling in FGF-2 stimulated human embryonic stem cells. *PLoS One* 2011, *6*, e17538.

[108] Pawson, T., Nash, P., Assembly of cell regulatory systems through protein interaction domains. *Science* 2003, *300*, 445-452.

[109] Faux, M. C., Scott, J. D., More on target with protein phosphorylation: conferring specificity by location. *Trends Biochem Sci* 1996, *21*, 312-315.

[110] Koytiger, G., Kaushansky, A., Gordus, A., Rush, J.*, et al.*, Phosphotyrosine signaling proteins that drive oncogenesis tend to be highly interconnected. *Mol Cell Proteomics* 2013, *12*, 1204-1213.

[111] Peti, W., Page, R., Molecular basis of MAP kinase regulation. *Protein Sci* 2013, *22*, 1698-1710.

[112] Kenny, J. J., Martínez-Maza, O., Fehniger, T., Ashman, R. F., Lipid synthesis: an indicator of antigen-induced signal transduction in antigen-binding cells. *J Immunol* 1979, *122*, 1278-1284.

[113] Kolch, W., Halasz, M., Granovskaya, M., Kholodenko, B. N., The dynamic control of signal transduction networks in cancer cells. *Nat Rev Cancer* 2015, *15*, 515-527.

[114] Mendoza, M. C., Er, E. E., Blenis, J., The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem Sci* 2011, *36*, 320-328.

[115] Davies, H., Bignell, G. R., Cox, C., Stephens, P.*, et al.*, Mutations of the BRAF gene in human cancer. *Nature* 2002, *417*, 949-954.

[116] Shi, Y., Wang, J., Chandarlapaty, S., Cross, J.*, et al.*, PTEN is a protein tyrosine phosphatase for IRS1. *Nat Struct Mol Biol* 2014, *21*, 522-527.

[117] Chen, Z., Trotman, L. C., Shaffer, D., Lin, H. K.*, et al.*, Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* 2005, *436*, 725-730.

[118] Martelli, A. M., Buontempo, F., McCubrey, J. A., Drug discovery targeting the mTOR pathway. *Clin Sci (Lond)* 2018, *132*, 543-568.

[119] Cohen, P., The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends Biochem Sci* 2000, *25*, 596-601.

[120] Nika, H., Lee, J., Willis, I. M., Angeletti, R. H., Hawke, D. H., Phosphopeptide characterization by mass spectrometry using reversed-phase supports for solid-phase β-elimination/Michael addition. *J Biomol Tech* 2012, *23*, 51-68.

[121] Abe, Y., Nagano, M., Tada, A., Adachi, J., Tomonaga, T., Deep Phosphotyrosine Proteomics by Optimization of Phosphotyrosine Enrichment and MS/MS Parameters. *J Proteome Res* 2017, *16*, 1077-1086.

[122] Yang, C., Zhong, X., Li, L., Recent advances in enrichment and separation strategies for mass spectrometry-based phosphoproteomics. *Electrophoresis* 2014, *35*, 3418-3429.

[123] Negroni, L., Claverol, S., Rosenbaum, J., Chevet, E.*, et al.*, Comparison of IMAC and MOAC for phosphopeptide enrichment by column chromatography. *J Chromatogr B Analyt Technol Biomed Life Sci* 2012, *891-892*, 109-112.

[124] Zhou, H., Ye, M., Dong, J., Han, G., *et al.*, Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. *J Proteome Res* 2008, *7*, 3957-3967.

[125] de Graaf, E. L., Giansanti, P., Altelaar, A. F., Heck, A. J., Single-step enrichment by Ti4+-IMAC and label-free quantitation enables in-depth monitoring of phosphorylation dynamics with high reproducibility and temporal resolution. *Mol Cell Proteomics* 2014, *13*, 2426-2434.

[126] Post, H., Penning, R., Fitzpatrick, M. A., Garrigues, L. B., *et al.*, Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* 2017, *16*, 728-737.

[127] Erickson, B. K., Jedrychowski, M. P., McAlister, G. C., Everley, R. A., *et al.*, Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. *Anal Chem* 2015, *87*, 1241-1249.

[128] Scheltema, R. A., Hauschild, J. P., Lange, O., Hornburg, D., *et al.*, The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol Cell Proteomics* 2014, *13*, 3698-3708.

[129] Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., *et al.*, The one hour yeast proteome. *Mol Cell Proteomics* 2014, *13*, 339-347.

[130] Kelstrup, C. D., Jersie-Christensen, R. R., Batth, T. S., Arrey, T. N., *et al.*, Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J Proteome Res* 2014, *13*, 6187-6195.

[131] Giansanti, P., Aye, T. T., van den Toorn, H., Peng, M., *et al.*, An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas. *Cell Rep* 2015.

[132] Sharma, K., D'Souza, R. C., Tyanova, S., Schaab, C., *et al.*, Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep* 2014, *8*, 1583-1594.

[133] Tape, C. J., Worboys, J. D., Sinclair, J., Gourlay, R., *et al.*, Reproducible automated phosphopeptide enrichment using magnetic TiO2 and Ti-IMAC. *Anal Chem* 2014, *86*, 10296-10302.

[134] Cui, L., Reid, G. E., Examining factors that influence erroneous phosphorylation site localization via competing fragmentation and rearrangement reactions during ion trap CID-MS/MS and -MS(3.). *Proteomics* 2013, *13*, 964-973.

[135] Palumbo, A. M., Reid, G. E., Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. *Anal Chem* 2008, *80*, 9735-9747.

[136] Wiese, H., Kuhlmann, K., Wiese, S., Stoepel, N. S., *et al.*, Comparison of alternative MS/MS and bioinformatics approaches for confident phosphorylation site localization. *J Proteome Res* 2014, *13*, 1128-1137.

[137] Frese, C. K., Zhou, H., Taus, T., Altelaar, A. F., *et al.*, Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J Proteome Res* 2013, *12*, 1520-1525.

[138] Ke, M., Shen, H., Wang, L., Luo, S., *et al.*, Identification, Quantification, and Site Localization of Protein Posttranslational Modifications via Mass Spectrometry-Based Proteomics. *Adv Exp Med Biol* 2016, *919*, 345-382.

[139] MacLean, D., Burrell, M. A., Studholme, D. J., Jones, A. M., PhosCalc: a tool for evaluating the sites of peptide phosphorylation from mass spectrometer data. *BMC Res Notes* 2008, *1*, 30.

[140] Angeleri, M., Muth-Pawlak, D., Aro, E. M., Battchikova, N., Study of O-Phosphorylation Sites in Proteins Involved in Photosynthesis-Related Processes in Synechocystis sp. Strain PCC 6803: Application of the SRM Approach. *J Proteome Res* 2016, *15*, 4638-4652.

[141] Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J., Gygi, S. P., A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 2006, *24*, 1285-1292.

[142] Taus, T., Köcher, T., Pichler, P., Paschke, C., *et al.*, Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res* 2011, *10*, 5354-5362.

[143] Savitski, M. M., Lemeer, S., Boesche, M., Lang, M., *et al.*, Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics* 2011, *10*, M110.003830.

[144] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., *et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, *127*, 635-648.

[145] Marx, H., Lemeer, S., Schliep, J. E., Matheron, L., *et al.*, A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat Biotechnol* 2013, *31*, 557-564.

[146] Ollikainen, E., Bonabi, A., Nordman, N., Jokinen, V., *et al.*, Rapid separation of phosphopeptides by microchip electrophoresis-electrospray ionization mass spectrometry. *J Chromatogr A* 2016, *1440*, 249-254.

[147] Otvos, L., Tangoren, I. A., Wroblewski, K., Hollosi, M., Lee, V. M., Reversed-phase high-performance liquid chromatographic separation of synthetic phosphopeptide isomers. *J Chromatogr* 1990, *512*, 265-272.

[148] Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., *et al.*, Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 2005, *4*, 310-327.

[149] Cunningham, D. L., Sweet, S. M., Cooper, H. J., Heath, J. K., Differential phosphoproteomics of fibroblast growth factor signaling: identification of Src family kinase-mediated phosphorylation events. *J Proteome Res* 2010, *9*, 2317-2328.

[150] Ballif, B. A., Roux, P. P., Gerber, S. A., MacKeigan, J. P., *et al.*, Quantitative phosphorylation profiling of the ERK/p90 ribosomal S6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors. *Proc Natl Acad Sci U S A* 2005, *102*, 667-672.

[151] Courcelles, M., Bridon, G., Lemieux, S., Thibault, P., Occurrence and detection of phosphopeptide isomers in large-scale phosphoproteomics experiments. *J Proteome Res* 2012, *11*, 3753-3765.

[152] Edwards, A. M., Isserlin, R., Bader, G. D., Frye, S. V., *et al.*, Too many roads not taken. *Nature* 2011, *470*, 163-165.

[153] de Graaf, E. L., Kaplon, J., Mohammed, S., Vereijken, L. A., *et al.*, Signal Transduction Reaction Monitoring Deciphers Site-Specific PI3K-mTOR/MAPK Pathway Dynamics in Oncogene-Induced Senescence. *J Proteome Res* 2015, *14*, 2906-2914.

[154] Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., White, F. M., Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci U S A* 2007, *104*, 5860-5865.

[155] Narumi, R., Murakami, T., Kuga, T., Adachi, J., *et al.*, A strategy for large-scale phosphoproteomics and SRM-based validation of human breast cancer tissue samples. *J Proteome Res* 2012, *11*, 5311-5322.

[156] Jin, L. L., Tong, J., Prakash, A., Peterman, S. M., *et al.*, Measurement of protein phosphorylation stoichiometry by selected reaction monitoring mass spectrometry. *J Proteome Res* 2010, *9*, 2752-2761.

[157] Lam, M. P., Scruggs, S. B., Kim, T. Y., Zong, C., *et al.*, An MRM-based workflow for quantifying cardiac mitochondrial protein phosphorylation in murine and human tissue. *J Proteomics* 2012, *75*, 4602-4609.

[158] Lam, M. P., Lau, E., Scruggs, S. B., Wang, D., *et al.*, Site-specific quantitative analysis of cardiac mitochondrial protein phosphorylation. *J Proteomics* 2013, *81*, 15-23.

[159] Soste, M., Hrabakova, R., Wanka, S., Melnik, A*., et al.*, A sentinel protein assay for simultaneously quantifying cellular processes. *Nat Methods* 2014, *11*, 1045-1048.

[160] Hökfelt, T., Broberger, C., Xu, Z. Q., Sergeyev, V*., et al.*, Neuropeptides--an overview. *Neuropharmacology* 2000, *39*, 1337-1356.

[161] Bodnar, R. J., Klein, G. E., Endogenous opiates and behavior: 2003. *Peptides* 2004, *25*, 2205-2256.

[162] Hook, V., Bark, S., Gupta, N., Lortie, M*., et al.*, Neuropeptidomic components generated by proteomic functions in secretory vesicles for cell-cell communication. *AAPS J* 2010, *12*, 635-645.

[163] Hummon, A. B., Richmond, T. A., Verleyen, P., Baggerman, G*., et al.*, From the genome to the proteome: uncovering peptides in the Apis brain. *Science* 2006, *314*, 647-649.

[164] Sköld, K., Svensson, M., Norrman, M., Sjögren, B*., et al.*, The significance of biochemical and molecular sample integrity in brain proteomics and peptidomics: stathmin 2-20 and peptides as sample quality indicators. *Proteomics* 2007, *7*, 4445-4456.

[165] Svensson, M., Boren, M., Sköld, K., Fälth, M*., et al.*, Heat stabilization of the tissue proteome: a new technology for improved proteomics. *J Proteome Res* 2009, *8*, 974-981.

[166] Frese, C. K., Boender, A. J., Mohammed, S., Heck, A. J*., et al.*, Profiling of diet-induced neuropeptide changes in rat brain by quantitative mass spectrometry. *Anal Chem* 2013, *85*, 4594-4604.

[167] Lee, J. E., Zamdborg, L., Southey, B. R., Atkins, N*., et al.*, Quantitative peptidomics for discovery of circadian-related peptides from the rat suprachiasmatic nucleus. *J Proteome Res* 2013, *12*, 585-593.

[168] Chen, R., Xiao, M., Buchberger, A., Li, L., Quantitative neuropeptidomics study of the effects of temperature change in the crab Cancer borealis. *J Proteome Res* 2014, *13*, 5767-5776.

[169] Li, L., Kelley, W. P., Billimoria, C. P., Christie, A. E*., et al.*, Mass spectrometric investigation of the neuropeptide complement and release in the pericardial organs of the crab, Cancer borealis. *J Neurochem* 2003, *87*, 642-656.

[170] Salisbury, J. P., Boggio, K. J., Hsu, Y. W., Quijada, J*., et al.*, A rapid MALDI-TOF mass spectrometry workflow for Drosophila melanogaster differential neuropeptidomics. *Mol Brain* 2013, *6*, 60.

[171] Schmerberg, C. M., Liang, Z., Li, L., Data-independent MS/MS quantification of neuropeptides for determination of putative feeding-related neurohormones in microdialysate. *ACS Chem Neurosci* 2015, *6*, 174-180.

[172] Jäverfalk-Hoyes, E. M., Bondesson, U., Westerlund, D., Andrén, P. E., Simultaneous analysis of endogenous neurotransmitters and neuropeptides in brain tissue using capillary electrophoresis--microelectrospray-tandem mass spectrometry. *Electrophoresis* 1999, *20*, 1527-1532.

[173] Kosanam, H., Ramagiri, S., Dass, C., Quantification of endogenous alpha- and gamma-endorphins in rat brain by liquid chromatography-tandem mass spectrometry. *Anal Biochem* 2009, *392*, 83-89.

[174] Dass, C., Kusmierz, J. J., Desiderio, D. M., Mass spectrometric quantification of endogenous beta-endorphin. *Biol Mass Spectrom* 1991, *20*, 130-138.

[175] Pailleux, F., Beaudry, F., Evaluation of multiple reaction monitoring cubed for the analysis of tachykinin related peptides in rat spinal cord using a hybrid triple quadrupole-linear ion trap mass spectrometer. *J Chromatogr B Analyt Technol Biomed Life Sci* 2014, *947-948*, 164-167.

[176] Romanova, E. V., Sweedler, J. V., Peptidomics for the discovery and characterization of neuropeptides and hormones. *Trends Pharmacol Sci* 2015, *36*, 579-586.

Chapter 2

Diet-Induced Neuropeptide Expression: Feasibility of Quantifying Extended and Highly Charged Endogenous Peptide Sequences by Selected Reaction Monitoring

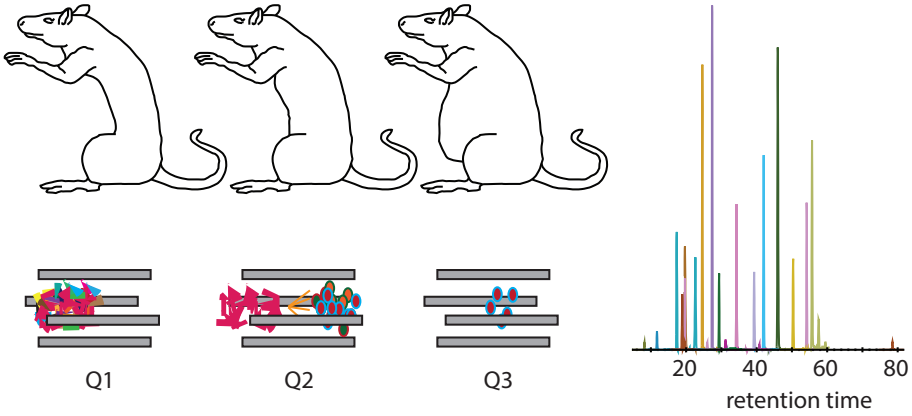
Thierry Schmidlin[1], Arjen J. Boender[2], Christian K. Frese[1], Albert J. R. Heck[1], Roger A. H. Adan[2], A. F. Maarten Altelaar[1]

Published in Analytical Chemistry 2015, Volume 87, issue 19, pages 9966-9973

[1] *Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University and Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, The Netherlands*

[2] *Department of Translational Neuroscience, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG, Utrecht, The Netherlands*

Abstract



Understanding regulation and action of endogenous peptides, especially neuropeptides, which serve as inter- and intracellular signal transmitters, is key in understanding a variety of functional processes, such as energy balance, memory, circadian rhythm, drug addiction, etc. Therefore, accurate and reproducible quantification of these bioactive endogenous compounds is highly relevant. The biosynthesis of endogenous peptides, involving multiple possible trimming and modification events, hinders the *de novo* prediction of the active peptide sequences, making MS-based measurements very valuable in determining the actual active compounds. Here, we report an extended selected reaction monitoring (SRM)-based strategy to reproducibly and quantitatively monitor the abundances of a set of 15 endogenously occurring peptides from *Rattus norvegicus* hypothalamus. We demonstrate that SRM can be extended toward reproducible detection and quantification of peptides, bearing characteristics very different from tryptic peptides. We show that long peptide sequences, producing precursors with up to five and MS2 fragment ions with up to three charges, can be targeted by SRM on a triple quadrupole instrument. Using this approach to quantify endogenous peptide levels in hypothalami of animals subjected to different diets revealed several significant changes, most notably the significant upregulation of VGF-derived signaling peptide AQEE-30 upon high caloric feeding.

1 Introduction

The advent of mass spectrometry (MS)-based proteomics has caused a significant increase in the amount of quantitative data on thousands of proteins, including dynamic post-translational modifications, across a plethora of conditions. This has given the scientific community both a powerful new tool but also a new challenge [1]. While MS driven discovery experiments provide new views on the molecular mechanisms occurring in cells, the need for validation tools for many new targets rises [2]. Whereas the production of antibodies for all targets of interest is costly and time-consuming, the development of selected reaction monitoring (SRM) resulted in a fast, robust, and relatively inexpensive tool to quantify, in principal, almost any target of interest, and especially the use of heavy labeled peptides enables straightforward assay development [3-5]. Compared to shotgun proteomics this enables more accurate and reproducible quantification of numerous proteins across large populations. However, the technology is still limited in sample complexity it can handle and requires peptides amenable to robust MS analysis and hypothesis-based assay development [2].

The workflow to establish SRM assays has been standardized in recent years. Especially the use of heavy labeled peptides made assay development and quantification of tryptic peptides more straightforward [5]. However, the use of alternative enzymes beyond trypsin is gaining popularity, especially in the field of global proteomics [6-9]. Moreover, experiments ranging from middle-down using different kind of proteases [10, 11] to top-down experiments analyzing intact proteins [12-14] and the analysis of endogenous peptides [15-18] has been the focus of an increasing number of MS-based proteomics studies. One important group of endogenous peptides are neuropeptides, which serve as inter- and intracellular neuronal signal transmitters. They are involved in a variety of processes, such as energy balance, memory, circadian rhythm, and drug addiction [19]. Their biosynthesis involves multiple steps of proteolytic cleavage of a precursor protein and subsequent posttranslational modifications [20]. The variety of possible trimming and modification events impedes the *de novo* prediction of the active neuropeptides from the genome or even the precursor protein [21]. Therefore, mass-spectrometric measurement of such peptides can be very valuable in determining the actual active compounds present [16, 22-26]. Especially, the growing use of the electron driven fragmentation technique electron transfer dissociation (ETD) [27] allows for in depth investigation of the neuropeptidome by mass spectrometry, as ETD is less influenced by peptide sequence, length, and post-translational modifications than collision-based fragmentation techniques.

Several reports describe quantification of endogenous peptides, reviewed recently by Buchberger *et al*. [28]. Because of the endogenous nature of the peptides of interest, resulting in arbitrary peptide sequences, and the maturing of label-free quantification methods, most studies utilize this approach [16, 29-31], although chemical labeling strate-

gies have been employed [32, 33] and recently also data-independent acquisition (DIA) [34]. However, to date, targeted MS-based tools for robust quantification of endogenous peptides are mostly lacking, as they show very different mass spectrometric characteristics from tryptic peptides. Monitoring target peptides by SRM already dates back several decades and has occasionally been used to quantify a limited number of neuropeptides, mainly in whole rat brain samples. In 1999, Jäverfalk-Hoyes *et al*. [35] used SRM assays to quantify Met-Enkephalin and Substance P 1-7 in rat brains, and in 2009, Kosanam *et al*. [36] used SRM to quantify endogenous α- and β-endorphins. Both of these studies monitor comparably short, trypsin like peptides using a single transition per peptide only. So far the only attempt to measure longer neuropeptides was performed by Dass *et al*. [37] using SRM in combination with fast atom bombardment to quantify β-endorphin, a peptide containing 31 amino acid residues, however, using tryptic digestion prior to the measurement.

The use of SRM for the targeted quantification of peptides evolved substantially throughout the last years. In current typical SRM assays, multiple parameters are taken into account. Several transitions per peptide are acquired and their relative intensities as well as their LC elution time and profile result in a large gain in specificity [38]. These strict SRM criteria result in extraction of quantitative information from complex biological backgrounds with high selectivity and thus confidence [5]. Additional instrumental and software developments enable a high degree of multiplexing [39]. Collision energy optimization for individual transitions additionally allows a boost in sensitivity for each peptide of interest [40], thus creating the most sensitive MS-based detection method available [41].

Taken these developments together have made SRM, used in combination with stable isotope dilution, the quantitative "gold standard" MS-based quantification method for tryptic peptides [42, 43]. Since antibody based analysis of neuropeptides is cumbersome, targeted MS analysis could be very valuable; however, selection of SRM friendly peptides is not possible. Here, we show that the general concept of SRM can be expanded to measure long and highly basic endogenous peptides with precursor and fragment charges much higher than usually targeted. While long and highly basic peptides occur in tryptic digests as well, they are mostly omitted as they often carry missed cleavages [4]. We show the capability of SRM on a triple quadrupole MS to reliably quantify peptides with up to 30 amino acid residues and multiple basic sites. The selected endogenous peptides require transitions containing precursor charges of up to 5+ and fragment charges of up to 3+. To the best of our knowledge, the study performed here is the first to show the capability of SRM to simultaneously quantify multiple endogenous peptides using high precursor charge and fragment ion charge states in hypothalami.

We selected a set of 15 peptides (Table 1) to be targeted by SRM. The peptides were chosen based on a good detectability and potential regulation patterns from our recent shotgun study [16], comparing endogenous peptide expression changes in different brain

areas of *Rattus norvegicus* in response to diverse feeding conditions. In the previous study, we detected over 1700 unique endogenous peptides by combining higher energy collisional dissociation (HCD) and ETD fragmentation. The variations between individuals of the same feeding schedule, however, were relatively high (CVs up to 124%), thus hampering the detection of significant diet-induced changes in neuropeptide abundance. To overcome this limitation and determine significant changes in peptide expression, a higher number of replicates is required, necessitating a high-throughput analysis method, like SRM. Therefore, we set up a targeted SRM method, expanding the number of biological replicates by preparing in total 24 new hypothalamic neuropeptide samples, to resolve confident quantification of selected neuropeptides.

Table 1. List of Endogenous Neuropeptides (and Their Precursor Proteins) Targeted by Selected Reaction Monitoring

precursor protein	peptide name	sequence
Cbln1	[des-Ser1]-Cerebellin	GSAKVAFSAIRSTNH
Cbln1	Cerebellin	SGSAKVAFSAIRSTNH
Gal	Galanin	GWTLNSAGYLLGPHAIDNHRSFSDKHGLT(amide)
Gnrh1	Gonadoliberin-1	-*Pyr*-HWSYGLRPG(amide)[a]
Nts	Neurotensin	-*Pyr*-LYENKPRRPYIL
Pdyn	Dynorphin A(1–8)	YGGFLRRI
Pdyn	Dynorphin A(1–13)	YGGFLRRIRPKLK
Pdyn	α-neoendorphin	YGGFLRKYPK
Pdyn	β-neoendorphin	YGGFLRKYP
Pdyn	Dynorphin B	YGGFLRRQFKVVT
Pdyn	Dynorphin A(1–17)	YGGFLRRIRPKLKWDNQ
Pmch	Neuropeptide EI	EIGDEENSAKFPI(amide)
Pnoc	Nociceptin	FGGFTGARKSARKLANQ
Sst	Anthrin	APSDPRLRQF
Vgf	AQEE-30	AQEEADAEERRLQEQEELENYIEHVLLHRP

[a] Pyr = N-terminal pyroglutamate.

2 Experimental Section

2.1 Materials

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich (Steinham, Germany). Synthetic peptides were purchased from JPT (Berlin, Germany).

2.2 Animals

The experiments were approved by the animal ethics committee of the University Medical Center Utrecht, following Dutch legislation. Male Wistar rats (Charles-River, Erkrath, Germany) were used (N = 24), ranging in initial weight from 200 to 250 g. The animals were housed individually (378 mm × 217 mm × 180 mm) in a controlled environment under a 12:12 light/dark cycle with lights on at 07:00 h. Rats were allowed to adapt to their environment for 7 days, after which they were divided into three experimental groups that were matched for body weight. Control (CTRL) and high-fat high-sucrose (HFHS) animals had ad libitum access to water and chow (Special diet services, Essex, U.K.). In addition, HFHS animals had ad libitum access to saturated fat (Vandemoortele, Eeklo, Belgium) and a 30% sucrose solution (Suiker Unie, Hoogkerk, The Netherlands). Restricted feeding schedule (RFS) animals received chow from 13:00 to 15:00 daily and had ad libitum access to water. Detailed information about the food intake and the body weight of all 24 individuals can be found in Table S-8.

2.3 Sample Preparation

All animals were decapitated between 12:15 h and 12:45 h within 10 s after they had been taken from their home cage. Immediately after decapitation, the heads were heated using a 800 W microwave (Sharp Co., Bangkok, Thailand) [26, 44] for 5 s, after which the brains were removed from the skull and the hypothalamic areas were dissected, snap-frozen in liquid nitrogen, and stored at −80 °C until further use.

2.4 Peptide Extraction

Neuropeptides were extracted from rat hypothalami by a double-extraction protocol as previously described [16]. In short, hypothalami were subjected to two subsequent extractions using acidified acetone and diluted acetic acid, respectively. For the first step, endogenous peptides were extracted with 300 µL of ice-cold acidified acetone (acetone/water/hydrochloric acid 40:6:1), lysed on ice by microtip sonication, incubated on ice for 1 h, and centrifuged for 25 min at 14 000g. Supernatants were neutralized with NaOH and dried down. For the second extraction step, endogenous peptides from the remaining pellets were extracted with 0.25% acetic acid, incubated on ice for 1 h, and centrifuged for 25 min at 14 000g. For each hypothalamus, the two extracts were pooled and passed through a 10 kDa cutoff filter (Amicon ultra YM-10, Millipore) for 45 min at 14 000g. C18 solid-phase extraction (Sep-Pak C18 cartridge 1 cc, Waters) was used for desalting prior to MS analysis.

2.5 Liquid Chromatography–Mass Spectrometry Settings

All SRM experiments were performed on a TSQ Vantage triple quadrupole mass spectrometer (Thermo Fisher). Upfront chromatographic separation was performed using an

EASY-spray system consisting of an Easy-nLC 1000 coupled to a 25 cm, 75 µm i.d. PepMap RSLC, C18, 100 Å, 2 µm particle size column (Thermo Scientific, Odense, DK). The LC was configured in one-column setup. Formic acid (0.1%, Merck, Darmstadt, Germany) in deionized water (Milli-Q, Millipore) was used as solvent A, 0.1% formic acid in acetonitrile (Biosolve, Valkenswaard, NL) as solvent B. All measurements were performed at 200 nL/min flow rate. Different gradients were used for assay development and samples. For detailed gradient information, see section S5 in the Supporting Information. Samples were loaded in 10% formic acid at a volume of 2 µL. Loading on the column occurred at 600 bar. The mass spectrometer was set to 0.7 Da peak width (fwhm) for Q1 and Q3.

HCD spectra were acquired on a Q Exactive (Thermo Fisher) coupled to an Easy-nLC 1000 (Thermo Scientific, Odense, DK) equipped with in-house packed columns. Peptides were trapped (Reprosil C18, Dr Maisch, GmbH, Ammersbuch, Germany, 3 µm, 2 cm × 100 µm) at 800 bar with 100% solvent A (0.1% formic acid in water) before separation on the analytical column (Agilent Poroshell 120 EC-C18, 2.7 µm, 40 cm × 50 µm) on a 31 min gradient from 7% to 30% solvent B (0.1% formic acid in ACN) at a flow rate of 150 nL/min.

2.6 Spectral Library Acquisition

Two different fragment spectral libraries were acquired, once on a triple quadrupole mass spectrometer and once on a Q Exactive, which has been shown to accurately reflect the fragmentation events happening in a triple quadrupole mass spectrometer [45]. The Q Exactive mass spectrometer was set up as described above and programmed in data-dependent acquisition mode. Full scans were acquired at a resolution of 35 000 at 400 *m/z*. Fragmentation was induced for the Top 10 peaks using a 10 s dynamic exclusion. Target peaks were isolated in a 1.5 Da isolation window and subjected to HCD fragmentation with a normalized collision energy of 25%. MS/MS spectra were acquired in the Orbitrap at a resolution of 17 500 at 400 *m/z*.

Triple quadrupole fragment ion spectral libraries were acquired on a Thermo TSQ Vantage. The mass spectrometer was therefore programmed in QED mode. This contains SRM type of measurements for potential transitions determined from HCD libraries or theoretically. Upon surpassing a threshold of 1000 counts per second, a fragmentation event is triggered for the corresponding precursor. For each triggered fragmentation event, an identical repeat is programmed 15 s later before a 50 s exclusion duration. MS/MS spectra were acquired in the range of 200–1250 *m/z* at a scan time of 1.1 s.

MS/MS spectra for both libraries were partly analyzed using MASCOT [46] with a customized database containing only peptides of interest to restrict the search space and partly by manual fragment annotation.

2.7 Assay Development for Selected Reaction Monitoring

The most intense fragment ions found in the spectral libraries for each peptide were directly used as transitions, multiplexing up to 10 transitions per precursor if present. Those initial SRM assays were applied to the synthetic peptide library, enabling subsequent optimization of multiple parameters such as collision energy and LC gradient. For collision energy optimization, instrument specific parameters were used as a starting point (CE = 0.03*m/z* + 2.905 for doubly charged precursors and CE = 0.038*m/z* + 2.281 for precursor charges of three and higher) to scan through different normalized collision energy values using a step size of one [40]. The chromatographic gradient was adjusted to achieve an equal distribution of the peptides across 75 min of measurement time.

2.8 Peptide Detection in Hypothalamic Samples

In total, 10% of all peptide extracts were analyzed in randomized order using a TSQ Vantage triple quadrupole mass spectrometer (Thermo Fisher) with the same setup as described above. Retention times were determined from the peptide library and each transition was run in a scheduled 8 min window. Potential carryover has been monitored by blank runs in between each sample subjected to identical measurement settings. In total, 10 fmol of fully digested bovine serum albumin has been spiked into each sample to control reproducibility of the retention time.

2.9 Quantification and Significance Analysis

Data analysis has been performed using Skyline [47]. The chromatographic quality has been assessed by visual inspection of the peak groups. Transitions showing clear interference have been deleted. Peptide quantification was performed label-free, where the summarized area under the curve of all transitions is used as a quantitative readout. Significance analysis has been performed using SRMstats [48]. This includes transforming intensities into $\log_2$-scale and subsequent intensity normalization based on equalizing medians. A linear mixed-effects model is successively used to test for abundance changes between different feeding conditions. A FDR cutoff of ≤0.05 in combination with a fold change of ≥2 were considered significant.

3 Results and Discussion

3.1 Ionization, Fragmentation, and Initial SRM Assay Development

For a working set of SRM assays it is crucial to know the most abundant precursor charge state and the most abundant fragment ions. For tryptic digests, this information is often directly extracted from shotgun experiments measured on mass spectrometers capable of producing fragment spectra highly similar to the ones acquired in a triple quadrupole in-

strument (i.e., utilizing beam-type collision-induced dissociation (CID) [45]). An alternative is the use of SRM-triggered MS/MS [42] on either complex samples or peptide libraries. As the discovery based study used here was performed using, in part, ETD [15, 16], a direct transfer into SRM assays was not possible. Therefore, for assay development we constructed a peptide library containing the synthetic counterparts of the neuropeptides of interest. Measuring full scan spectra enabled us to determine predominant precursor charge states (Figure 1A). As expected overall observed charge states are much higher than observed for peptides originating from tryptic digests (Supplementary Figure 1 and Supplementary Table S-6). Fragmentation spectra were subsequently obtained for each peptide on a Q Exactive or triple-quadrupole TSQ Vantage, using beam-type CID. As expected for long peptides with high charge states, CID is not the optimal fragmentation method providing less informative fragment spectra than, e.g., ETD. The fragmentation spectra for most peptides contain only a few peaks, which however could be annotated clearly to specific fragment ions. Much unlike tryptic peptides, those fragment ions are often multiply charged (Figure 1B). Combining the most intense precursor and fragment *m/z* gave rise to an initial set of transitions used as a base for optimization (Figure 1C).

3.2 Optimizing Collision Energy

Equations to predict the optimal collision energies for SRM assays exist for most instrument platforms capable of measuring SRM type experiments. However, they are mostly limited to predictive equations for precursor charges of 2+ and 3+ only. For transitions combining both highly charged precursor and fragment ions, those equations are of little use. By stepwise screening through different normalized collision energy percentages, we were able to optimize the signal for most of our transitions substantially. Signals increasing by up to 1700% for individual transitions were observed throughout our dataset (Figure 1D). These results clearly show that parameters for optimization of collision energy for larger multiple charged peptides cannot be extracted from parameters used for tryptic digests. Implementing the optimized settings into the final SRM assay considerably increased our sensitivity when measuring the endogenous neuropeptides. This however can drastically change the relative intensity patterns from the ones observed in the spectral library, as they were acquired using fixed collision energies depending only on the precursor *m/z*. Therefore, comparing relative intensities of our optimized SRM transitions and relative fragment ion intensities in the spectral library, which is often used as a quality control step to validate the SRM traces, is no longer possible. Instead, we verified relative intensities of the transitions measured in hypothalamus samples (Figure 1E) against SRM experiments with pure synthetic peptides.

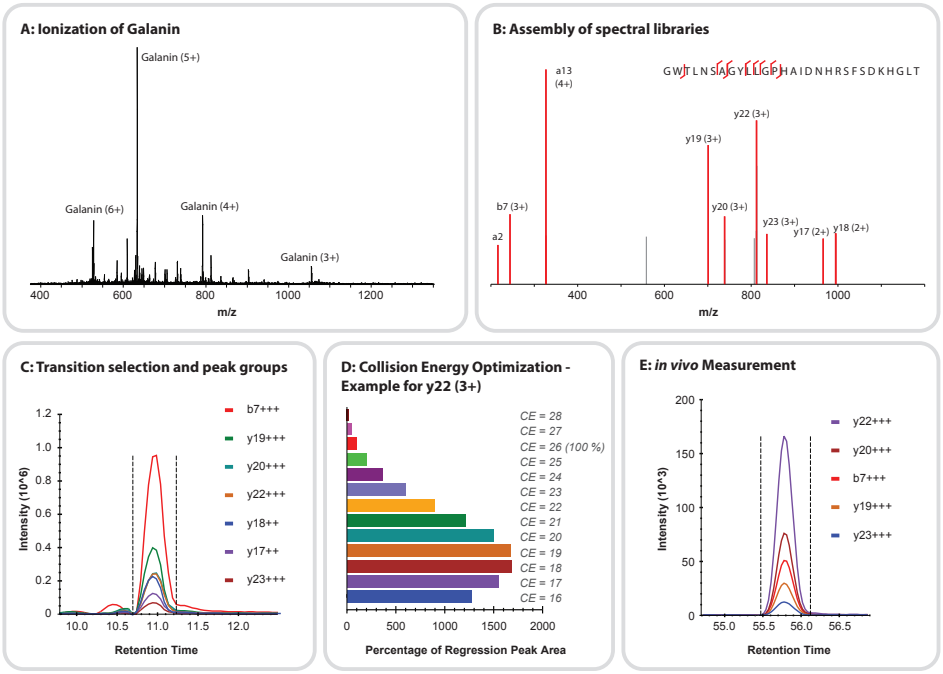


Figure 1: ***SRM assay development for highly charged neuropeptides using the example Galanin****. (A) The preferred charge state of the precursor ion was determined based on full scans acquired on a TSQ Vantage. (B) Thereof HCD fragmentation spectra were acquired on a Q Exactive. (C) The most intense fragment ions were used as transitions for an initial SRM assay. (D) Further assay refinement by optimizing collision energy for each transition. Response equations for transitions combining highly charged precursor and fragment ions are unknown, hence signal improvements of more than 1700% are possible as shown for the transition y22 (3+). (E) Optimized transitions were used to measure endogenous peptides in rat hypothalamus samples. SRM transitions showing interference were removed.*

3.3 In Vivo Neuropeptide Detection

Once the SRM assays were optimized, we applied them to measure endogenous neuropeptides from rat brain samples. Previously we investigated diet-induced changes in abundance of endogenous peptides in hypothalamus and striatum from rats from four differentially treated dietary groups including a high-fat/high-sucrose diet, mimicking diet-induced obesity. In that study the high variation observed put a strain on the quantification of the data, which we believe to be caused by the low number of animals per group (n = 3). The development of targeted SRM assays for several selected peptides from this initial study now allows us to increase the number of animals used, keeping the size of the experiment controlled while, as we believe, improving the reliability and significance of the quantification. The feeding conditions used here were regular diet (CTRL), high-fat/high-sucrose diet (HFHS), and restricted chow (RFS) [16, 49, 50], with the three dietary groups each consisting of eight animals. A schematic workflow is depicted in Figure 2.

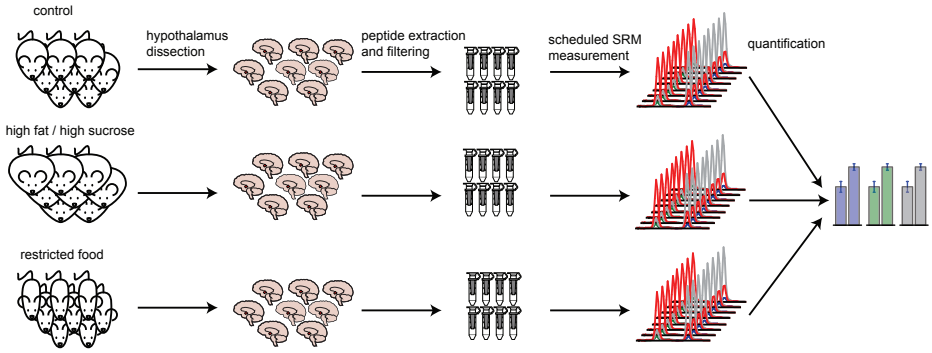


Figure 2: ***Workflow for measuring neuropeptide abundances in rat brain*** **ex vivo**. *In total, 24 rats were divided into three groups of 8 animals each and subjected to different feeding schedules including standard diet (control), standard diet plus ad libitum access to lard and sugar (high fat/high sucrose), and time restricted food access (restricted food). After 8 days of dietary exposure, hypothalami were dissected and subjected to peptide extraction. Neuropeptides of interest were measured in SRM mode on a triple quadrupole mass spectrometer using the area under the curve for quantification.*

The individual retention times of all peptides were determined using their synthetic standard counterpart, enabling us to measure the endogenous peptides in a scheduled manner and thus increasing the dwell time for each transition. All peptides were detectable *in vivo* except for Dynorphin A (1-13), which did not show any peak group across all 24 samples. The absence of this peptide can be explained as its endogenous processing and thus presence is unclear and often synthetic versions are described [51, 52]. All other peptides were clearly detectable, showing distinct peak groups in each sample with mostly little to no visible interference (section S2 in the Supporting Information).

Many of the 14 neuropeptides detected here exhibit features, so far not routinely implemented in SRM assays. For instance the peptides Neurotensin and Gonadoliberin-1 contain an N-terminal glutamate to pyroglutamate conversion (Table 1) while the transitions for the peptides Galanin (Figure 1E), AQEE-30, and Dynorphin A (1-17) are based on precursor charges of 5 in combination with fragment charges of up to 3. To the best of our knowledge, this is the first study showing the feasibility of a conventional triple quadrupole mass spectrometer to analyze peptides with very high charge states in SRM mode.

For each sample, all peptides were quantified individually using the area under the curve. At this point of the study, the peptide β-neoendorphine was excluded due to chromatographic peak broadening. The remaining 13 neuropeptides performed well under the optimized SRM conditions and could be confidently quantified in all dietary groups. Using the software package SRMstats [48], the samples were subjected to statistical significance analysis, testing for significant changes in peptide abundance across the three different feeding conditions. The analysis pipeline includes global signal normalization, quality control based on variations of the relative intensities of all transitions, and the use of a linear mixed-effect model to test for condition specific changes in peptide abundance. Figure 3A

depicts average fold changes between feeding conditions with error bars representing standard errors. No overlap between the error bars indicates statistical significance ($p \leq 0.05$). When comparing group specific changes in abundance of the peptides detected in our study to the changes reported by Frese *et al.*, we see similar trends for most of the investigated peptides. For instance, the increased abundances of Gonadoliberin-1, Neurotensin, Dynorphin A(1-8), α-neoendorphine, Dynorphin B, Neuropeptide EI, anthirin and AQEE-30 in high caloric feeding compared to the control group has been observed in both studies. The HFHS diet results in obesity [49], whereas RFS rats loose weight [53]. These positive and negative energy balance states are reflected by Galanin and by AQEE-30. In contrast most peptides shown in Figure 3 show an effect in the same direction during positive and negative energy balance. The latter results are in line with previous results in which upon exposure to the HFHS diet for 1 week, POMC mRNA levels decreased whereas NPY mRNA levels increased in the arcuate nucleus. This change in neuropeptide mRNA expression was in the same direction as in rats that were food restricted [54]. Exposure to the HFHS diet results in leptin resistance [55], in which high levels of plasma leptin fail to limit food intake and to suppress NPY and to increase POMC expression [49]. Taken together, although some neuropeptides reflect energy balance status (such as Galanin and AQEE-30), others (like Dynorphin) reflect leptin resistance in a state of positive energy balance.

The higher expression of the peptides Neurotensin, Dynorphin A(1-8), α-neoendorphine, Dynorphin B, Neuropeptide EI, and Anthirin under restricted food conditions compared to the control group could be confirmed with the targeted approach. In many of those cases these expression changes however showed no statistical significance in the shotgun approach. By using the approach of higher sample numbers in combination with SRM measurements, these trends could now be confirmed significantly for all of the above-mentioned peptides. Only in a minority of cases the quantification strategies used in the previous shotgun study and this study showed different outcomes. For instance, the increased abundance of Galanin and Dynorphin A(1-17) under high caloric diet and the increased abundance of Gonadoliberin-1 under food restriction as observed by us before could not be confirmed here, however the later was not significant in the shotgun study. Likewise the trends for the peptides Cerebellin and [des-Ser1]-Cerebellin vary between the two studies. Moreover, the observed diet-induced changes in peptide abundance are consistently less pronounced in the targeted SRM approach than in the shotgun experiments. We believe that this effect is likely due to the higher number of samples in combination with the more accurate quantification strategy, putting less weight on potential outliers. Especially in the above-mentioned cases of Galanin and Dynorphin A(1-17), the fold changes observed here were close to zero. The increased number of replicas used in this study is likely the cause of the different regulation patterns observed for the two forms of Cerebellin compared to Frese *et al*. It is however interesting to note that [des-

Ser1]-Cerebellin and Cerebellin originating from the same precursor protein show very similar diet related abundance changes. Likewise for the four peptides originating from Pdyn, three show very similar abundance patterns with only Dynorphin A(1-17) deviating from this trend. This could suggest a higher diet induced effect on precursor synthesis than actual peptide processing.

As label-free SRM assays are inherently less accurate than experiments using internal standards, most studies using label-free SRM experiments make use of a fold-change cutoff of two. Here a cutoff of 2-fold was surpassed in one occasion. The peptide AQEE-30 showed a higher than 2-fold, significant increase in rats subjected to a high fat/high sucrose diet compared to the control group (Figure 3A). AQEE-30 is one of several neuroendocrine peptides resulting from post-translational processing of the VGF protein [56] (Figure 3B). VGF is exclusively expressed in neuronal and neuroendocrine cells [57]. It has been linked to diverse biological processes throughout the last decades, including food intake and energy balance. Its mRNA levels have been shown to be affected by feeding conditions [58, 59]. Likewise, *vgf*-knockout mice are thin, small, and hyperactive and lose the ability to become obese [58].

So far most studies investigating the effect of VGF derived neuropeptides focus on another peptide product, namely, TLQP-21 [60]. This is so far the only VGF derived peptide for which potential cell surface receptors have been identified [61, 62]. While VGF itself has been related to the onset of obesity, interestingly TLQP-21 has the opposite effect [63, 64]. A chronic intracerebroventricular (ICV) infusion of TLQP-21 in mice fed with a normal laboratory diet did not induce obesity and resulted only in a small increase in resting expenditure [65]. Similar effects were observed in rats [66] and Siberian hamsters [57, 67]. These contradictory results regarding the role of VGF derived peptides in energy balance lead Lewis *et al*. [60] to suggest a biphasic role of VGF in energy regulation, requiring a further investigation into the role of other VGF-derived peptides.

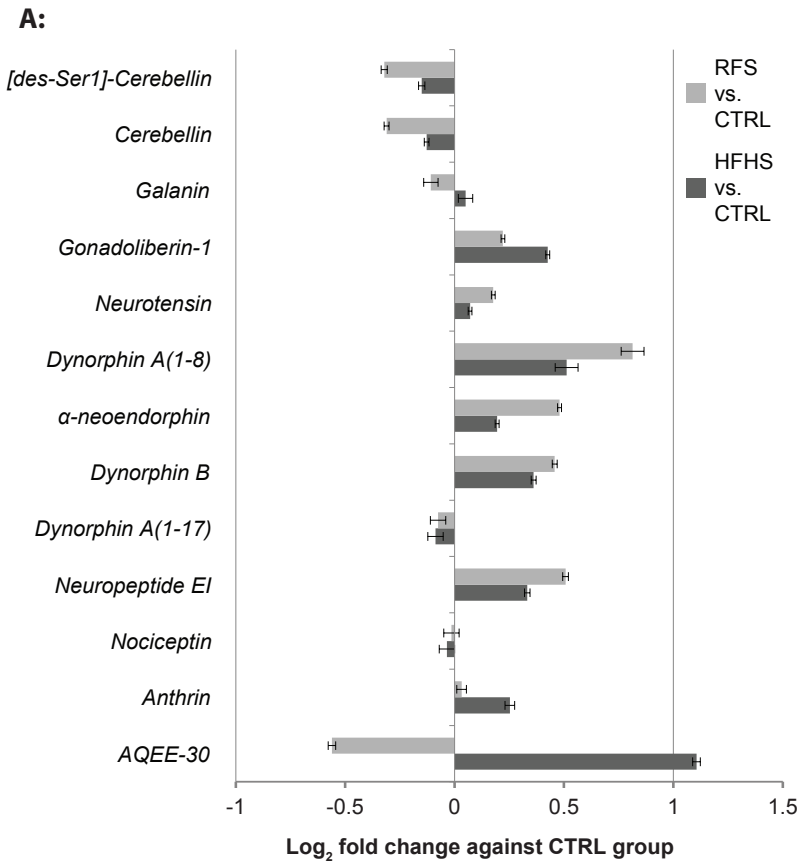


Figure 3: (A) Changes of neuropeptide abundance in brains of rats subjected to different diets. Bars indicate log2-fold change of each peptide upon restricted food access (light gray n = 8) and upon high caloric feeding (dark gray n = 8) against a control group (n = 8). By applying a significance cutoff of $p \leq 0.05$ and fold change cutoff of ≥2, the peptide AQEE-30 shows an elevated level upon high caloric feeding. (B) AQEE-30 is a product of multiple trimming steps guided by prohormone convertases. Its precursor polypeptide is VGF, a gene product often linked to food intake control and energy balance.

4 Conclusion

In the work presented here, we demonstrate the potential of SRM to specifically monitor endogenous neuropeptides of interest. Despite current common practice, we could show that SRM is not limited to tryptic peptides but can be expanded to measure long and highly basic (endogenous) peptides. Precursor charges of up to 5 and fragment charges of up to 3 prove not to limit the quality of the SRM peak groups and can be easily used for accurate quantification, although they require a dedicated individual optimization of analysis parameters, in particular the used collision energies for peptide fragmentation.

Finding a significant upregulation of the VGF derived peptide AQEE-30 upon high caloric diet in rats strengthens the reliability of our approach, as VGF has often been shown to have an involvement in energy balance and obesity. However, the direct role of AQEE-30 on a peptide level has so far only sparsely been investigated. In view of contradictory results regarding the role of other VGF-derived peptides in energy balance, AQEE-30 and its potential involvement in obesity might become an interesting target for further investigation.

Acknowledgment

A.F.M.A. is supported by The Netherlands Organization for Scientific Research (NWO) through a VIDI grant (Grant 723.012.102). This work is part of the project Proteins At Work, a program of The Netherlands Proteomics Centre financed by The Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-Scale Research Facilities of The Netherlands (Project Number 184.032.201).

Associated Contend

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.5b03334.

- Spectral libraries for peptides of interest; SRM traces of peptides libraries and hypothalamus samples; SRMStats parameters; SRMStats QC Plot; detailed mass spec and gradient information; SRM parameters for peptides of interest
- Quantitative SRM data of all peptide transitions
- Data on rat body weight and feeding schedule

References

[1] Altelaar, A. F., Munoz, J., Heck, A. J., Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 2013, *14*, 35-48.

[2] Picotti, P., Bodenmiller, B., Aebersold, R., Proteomics meets the scientific method. *Nat Methods* 2013, *10*, 24-27.

[3] Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., *et al.*, Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* 2009, *27*, 633-641.

[4] Gallien, S., Duriez, E., Domon, B., Selected reaction monitoring applied to proteomics. *J Mass Spectrom* 2011, *46*, 298-312.

[5] Picotti, P., Rinner, O., Stallmach, R., Dautel, F., *et al.*, High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods* 2010, *7*, 43-46.

[6] Giansanti, P., Aye, T. T., van den Toorn, H., Peng, M., *et al.*, An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas. *Cell Rep* 2015.

[7] Guo, X., Trudgian, D. C., Lemoff, A., Yadavalli, S., Mirzaei, H., Confetti: a multiprotease map of the HeLa proteome for comprehensive proteomics. *Mol Cell Proteomics* 2014, *13*, 1573-1584.

[8] Mohammed, S., Lorenzen, K., Kerkhoven, R., van Breukelen, B., *et al.*, Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. *Anal Chem* 2008, *80*, 3584-3592.

[9] Swaney, D. L., Wenger, C. D., Coon, J. J., Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* 2010, *9*, 1323-1329.

[10] Moradian, A., Kalli, A., Sweredoski, M. J., Hess, S., The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics* 2014, *14*, 489-497.

[11] Laskay, Ü., Lobas, A. A., Srzentić, K., Gorshkov, M. V., Tsybin, Y. O., Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments. *J Proteome Res* 2013, *12*, 5558-5569.

[12] Brunner, A. M., Lössl, P., Liu, F., Huguet, R., *et al.*, Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. *Anal Chem* 2015, *87*, 4152-4158.

[13] Gregorich, Z. R., Ge, Y., Top-down proteomics in health and disease: challenges and opportunities. *Proteomics* 2014, *14*, 1195-1210.

[14] Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E., *et al.*, Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 2011, *480*, 254-258.

[15] Altelaar, A. F., Mohammed, S., Brans, M. A., Adan, R. A., Heck, A. J., Improved identification of endogenous peptides from murine nervous tissue by multiplexed peptide extraction methods and multiplexed mass spectrometric analysis. *J Proteome Res* 2009, *8*, 870-876.

[16] Frese, C. K., Boender, A. J., Mohammed, S., Heck, A. J., *et al.*, Profiling of diet-induced neuropeptide changes in rat brain by quantitative mass spectrometry. *Anal Chem* 2013, *85*, 4594-4604.

[17] Li, L., Kelley, W. P., Billimoria, C. P., Christie, A. E., *et al.*, Mass spectrometric investigation of the neuropeptide complement and release in the pericardial organs of the crab, Cancer borealis. *J Neurochem* 2003, *87*, 642-656.

[18] Mommen, G. P., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., *et al.*, Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc Natl Acad Sci U S A* 2014, *111*, 4507-4512.

[19] Hökfelt, T., Broberger, C., Xu, Z. Q., Sergeyev, V.*, et al.*, Neuropeptides--an overview. *Neuropharmacology* 2000, *39*, 1337-1356.

[20] Steiner, D. F., The proprotein convertases. *Curr Opin Chem Biol* 1998, *2*, 31-39.

[21] Hummon, A. B., Richmond, T. A., Verleyen, P., Baggerman, G.*, et al.*, From the genome to the proteome: uncovering peptides in the Apis brain. *Science* 2006, *314*, 647-649.

[22] Bora, A., Annangudi, S. P., Millet, L. J., Rubakhin, S. S.*, et al.*, Neuropeptidomics of the supraoptic rat nucleus. *J Proteome Res* 2008, *7*, 4992-5003.

[23] Fricker, L. D., Neuropeptidomics to study peptide processing in animal models of obesity. *Endocrinology* 2007, *148*, 4185-4190.

[24] Neupert, S., Rubakhin, S. S., Sweedler, J. V., Targeted single-cell microchemical analysis: MS-based peptidomics of individual paraformaldehyde-fixed and immunolabeled neurons. *Chem Biol* 2012, *19*, 1010-1019.

[25] Nilsson, A., Stroth, N., Zhang, X., Qi, H.*, et al.*, Neuropeptidomics of mouse hypothalamus after imipramine treatment reveal somatostatin as a potential mediator of antidepressant effects. *Neuropharmacology* 2012, *62*, 347-357.

[26] Svensson, M., Sköld, K., Svenningsson, P., Andren, P. E., Peptidomics-based discovery of novel neuropeptides. *J Proteome Res* 2003, *2*, 213-219.

[27] Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 2004, *101*, 9528-9533.

[28] Buchberger, A., Yu, Q., Li, L., Advances in Mass Spectrometric Tools for Probing Neuropeptides. *Annu Rev Anal Chem (Palo Alto Calif)* 2015, *8*, 485-509.

[29] Kultima, K., Nilsson, A., Scholz, B., Rossbach, U. L.*, et al.*, Development and evaluation of normalization methods for label-free relative quantification of endogenous peptides. *Mol Cell Proteomics* 2009, *8*, 2285-2295.

[30] Lee, J. E., Atkins, N., Hatcher, N. G., Zamdborg, L.*, et al.*, Endogenous peptide discovery of the rat circadian clock: a focused study of the suprachiasmatic nucleus by ultrahigh performance tandem mass spectrometry. *Mol Cell Proteomics* 2010, *9*, 285-297.

[31] Southey, B. R., Lee, J. E., Zamdborg, L., Atkins, N., Jr.*, et al.*, Comparing label-free quantitative peptidomics approaches to characterize diurnal variation of peptides in the rat suprachiasmatic nucleus. *Anal Chem* 2014, *86*, 443-452.

[32] Che, F. Y., Fricker, L. D., Quantitative peptidomics of mouse pituitary: comparison of different stable isotopic tags. *J Mass Spectrom* 2005, *40*, 238-249.

[33] Chen, R., Xiao, M., Buchberger, A., Li, L., Quantitative neuropeptidomics study of the effects of temperature change in the crab Cancer borealis. *J Proteome Res* 2014, *13*, 5767-5776.

[34] Schmerberg, C. M., Liang, Z., Li, L., Data-independent MS/MS quantification of neuropeptides for determination of putative feeding-related neurohormones in microdialysate. *ACS Chem Neurosci* 2015, *6*, 174-180.

[35] Jäverfalk-Hoyes, E. M., Bondesson, U., Westerlund, D., Andrén, P. E., Simultaneous analysis of endogenous neurotransmitters and neuropeptides in brain tissue using capillary electrophoresis--microelectrospray-tandem mass spectrometry. *Electrophoresis* 1999, *20*, 1527-1532.

[36] Kosanam, H., Ramagiri, S., Dass, C., Quantification of endogenous alpha- and gamma-endorphins in rat brain by liquid chromatography-tandem mass spectrometry. *Anal Biochem* 2009, *392*, 83-89.

[37] Dass, C., Kusmierz, J. J., Desiderio, D. M., Mass spectrometric quantification of endogenous beta-endorphin. *Biol Mass Spectrom* 1991, *20*, 130-138.

[38] Marx, V., Targeted proteomics. *Nat Methods* 2013, *10*, 19-22.

[39] Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., *et al.*, High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* 2007, *6*, 1809-1817.

[40] Maclean, B., Tomazela, D. M., Abbatiello, S. E., Zhang, S., *et al.*, Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal Chem* 2010, *82*, 10116-10124.

[41] Lange, V., Picotti, P., Domon, B., Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* 2008, *4*, 222.

[42] Picotti, P., Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* 2012, *9*, 555-566.

[43] Ong, S. E., Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 2005, *1*, 252-262.

[44] Che, F. Y., Lim, J., Pan, H., Biswas, R., Fricker, L. D., Quantitative neuropeptidomics of microwave-irradiated mouse brain and pituitary. *Mol Cell Proteomics* 2005, *4*, 1391-1405.

[45] de Graaf, E. L., Altelaar, A. F., van Breukelen, B., Mohammed, S., Heck, A. J., Improving SRM assay development: a global comparison between triple quadrupole, ion trap, and higher energy CID peptide fragmentation spectra. *J Proteome Res* 2011, *10*, 4334-4341.

[46] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551-3567.

[47] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., *et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, *26*, 966-968.

[48] Chang, C. Y., Picotti, P., Hüttenhain, R., Heinzelmann-Schwarz, V., *et al.*, Protein significance analysis in selected reaction monitoring (SRM) measurements. *Mol Cell Proteomics* 2012, *11*, M111.014662.

[49] la Fleur, S. E., van Rozen, A. J., Luijendijk, M. C., Groeneweg, F., Adan, R. A., A free-choice high-fat high-sugar diet induces changes in arcuate neuropeptide expression that support hyperphagia. *Int J Obes (Lond)* 2010, *34*, 537-546.

[50] la Fleur, S. E., Luijendijk, M. C., van Rozen, A. J., Kalsbeek, A., Adan, R. A., A free-choice high-fat high-sugar diet induces glucose intolerance and insulin unresponsiveness to a glucose load not explained by obesity. *Int J Obes (Lond)* 2011, *35*, 595-604.

[51] Mamiya, T., Hasegawa, Y., Hiramatsu, M., Dynorphin a (1-13) alleviated stress-induced behavioral impairments in mice. *Biol Pharm Bull* 2014, *37*, 1269-1273.

[52] Schwarzer, C., 30 years of dynorphins--new insights on their functions in neuropsychiatric diseases. *Pharmacol Ther* 2009, *123*, 353-370.

[53] Merkestein, M., van Gestel, M. A., van der Zwaal, E. M., Brans, M. A., *et al.*, GHS-R1a signaling in the DMH and VMH contributes to food anticipatory activity. *Int J Obes (Lond)* 2014, *38*, 610-618.

[54] de Rijke, C. E., Hillebrand, J. J., Verhagen, L. A., Roeling, T. A., Adan, R. A., Hypothalamic neuropeptide expression following chronic food restriction in sedentary and wheel-running rats. *J Mol Endocrinol* 2005, *35*, 381-390.

[55] van den Heuvel, J. K., Eggels, L., van Rozen, A. J., Luijendijk, M. C., *et al.*, Neuropeptide Y and leptin sensitivity is dependent on diet composition. *J Neuroendocrinol* 2014, *26*, 377-385.

[56] Ferri, G. L., Noli, B., Brancia, C., D'Amato, F., Cocco, C., VGF: an inducible gene product, precursor of a diverse array of neuro-endocrine peptides and tissue-specific disease biomarkers. *J Chem Neuroanat* 2011, *42*, 249-261.

[57] Jethwa, P. H., Warner, A., Nilaweera, K. N., Brameld, J. M., *et al.*, VGF-derived peptide, TLQP-21, regulates food intake and body weight in Siberian hamsters. *Endocrinology* 2007, *148*, 4044-4055.

[58] Hahm, S., Fekete, C., Mizuno, T. M., Windsor, J.*, et al.*, VGF is required for obesity induced by diet, gold thioglucose treatment, and agouti and is differentially regulated in pro-opiomelanocortin- and neuropeptide Y-containing arcuate neurons in response to fasting. *J Neurosci* 2002, *22*, 6929-6938.

[59] Hahm, S., Mizuno, T. M., Wu, T. J., Wisor, J. P.*, et al.*, Targeted deletion of the Vgf gene indicates that the encoded secretory peptide precursor plays a novel role in the regulation of energy balance. *Neuron* 1999, *23*, 537-548.

[60] Lewis, J. E., Brameld, J. M., Jethwa, P. H., Neuroendocrine Role for VGF. *Front Endocrinol (Lausanne)* 2015, *6*, 3.

[61] Hannedouche, S., Beck, V., Leighton-Davies, J., Beibel, M.*, et al.*, Identification of the C3a receptor (C3AR1) as the target of the VGF-derived peptide TLQP-21 in rodent cells. *J Biol Chem* 2013, *288*, 27434-27443.

[62] Chen, Y. C., Pristerá, A., Ayub, M., Swanwick, R. S.*, et al.*, Identification of a receptor for neuropeptide VGF and its role in neuropathic pain. *J Biol Chem* 2013, *288*, 34638-34646.

[63] Cassina, V., Torsello, A., Tempestini, A., Salerno, D.*, et al.*, Biophysical characterization of a binding site for TLQP-21, a naturally occurring peptide which induces resistance to obesity. *Biochim Biophys Acta* 2013, *1828*, 455-460.

[64] Possenti, R., Muccioli, G., Petrocchi, P., Cero, C.*, et al.*, Characterization of a novel peripheral pro-lipolytic mechanism in mice: role of VGF-derived peptide TLQP-21. *Biochem J* 2012, *441*, 511-522.

[65] Bartolomucci, A., La Corte, G., Possenti, R., Locatelli, V.*, et al.*, TLQP-21, a VGF-derived peptide, increases energy expenditure and prevents the early phase of diet-induced obesity. *Proc Natl Acad Sci U S A* 2006, *103*, 14584-14589.

[66] Severini, C., La Corte, G., Improta, G., Broccardo, M.*, et al.*, In vitro and in vivo pharmacological role of TLQP-21, a VGF-derived peptide, in the regulation of rat gastric motor functions. *Br J Pharmacol* 2009, *157*, 984-993.

[67] Barrett, P., Ross, A. W., Balik, A., Littlewood, P. A.*, et al.*, Photoperiodic regulation of histamine H3 receptor and VGF messenger ribonucleic acid in the arcuate nucleus of the Siberian hamster. *Endocrinology* 2005, *146*, 1930-1939.

Chapter 3

Kinase Activity Profiling by MS-Based Quantification of t-Loop Phosphorylations

Thierry Schmidlin[1], Charlotte van Gelder[1], Stamatia Rontogianni[1], Bart L. van den Eshof[2], Albert J. R. Heck[1], Maartje van den Biggelaar[2], A. F. Maarten Altelaar[1]

Manuscript in preparation

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University and Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, The Netherlands

[2] Department of Plasma Proteins, Sanquin Research , 1066 CX Amsterdam, The Netherlands

Introductory paragraph

Protein kinases play a pivotal role in cellular communication and aberrant kinase activity has been linked to a variety of disorders ranging from cancer [1] to diabetes [2, 3] and cardiovascular disease [4]. Therefore, there is an urgent need for direct monitoring of kinase activity, which until now has mainly been extracted from either kinase protein expression levels or data on substrate phosphorylation. Here, we describe a robust strategy to determine kinase activation through targeted quantification of t-loop phosphorylation. By combining selective phosphopeptide enrichment with robust targeted mass spectrometry (MS), we provide highly specific assays for monitoring kinase activation at a so far unachieved level of sensitivity. To showcase our novel technology, we developed assays for 248 peptides, covering 221 phosphorylation sites in the t-loop region of 178 human kinases (33 % kinome coverage). Using these assays, we were able to monitor the activation of 54 kinases through 66 t-loop phosphosites across different cell types, patient-derived melanoma cells and primary blood platelets. The extreme sensitivity of our assays is highlighted by the reproducible detection of RIPK1 activating phosphorylation at serine 161 after TNFα stimulation, which has thus far only been reported using a protein expression system in combination with an *in vitro* kinase assay.

Main

Kinases are key regulators of inter- and intracellular communication and their inhibitors play central roles in targeted therapy and precision medicine [5]. Therefore, the investigation of global kinase dynamics is fundamental to our understanding of cellular function and to advance rational drug development. This clear relevance of kinases has sparked numerous endeavors to assess kinome activity states. Most of them however suffer from substantial drawbacks. In its most naïve application, activity of kinases is extracted from their abundance obtained in shotgun proteomics experiments. To account for the low abundance of most protein kinases, kinase specific enrichment protocols have been used, mainly consisting of immobilized broad-specificity kinase inhibitors and/or ATP-mimetics [6-9]. This kind of approach yields substantial kinase enrichment, however at the cost of sensitivity, requiring substantial input material and thereby limiting the approach to cell line-based model systems. Importantly, assessment of kinase abundance does not *per se* reflect kinase activity, which thus far has mainly been extracted from phosphoproteomics data. Here, the alignment of over- or underrepresented sequence motifs in the phosphoproteomics dataset are linked to known kinase consensus sequences [10]. This method however, suffers from a large gap of knowledge, as most of the detected phosphosites have not been characterized in terms of functionality [11] and substrate specificity for many kinases are redundant or simply unknown [12].

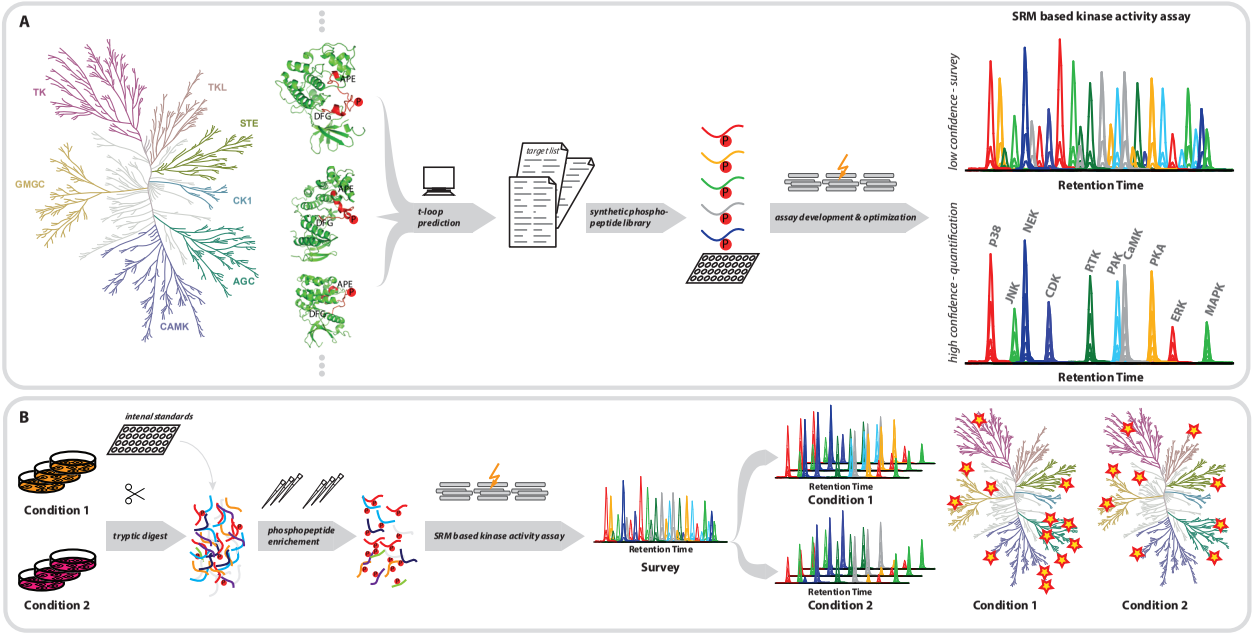


Figure 1: ***Kinase Activity Profiling.*** *(A) Kinase activity inducing sites located in the t-loop region were determined for numerous kinases by a combination of literature mining and sequence homology alignments. From the resulting target list a synthetic heavy isotope labeled peptide library was generated assuming complete tryptic digests. SRM assays were built and optimized covering t-loop phosphorylation of 33 % of the human kinome. Assays for each peptide can be used in 'survey mode' and 'quantification mode'. Survey mode operates with limited amount of transitions per peptide, short dwell times and a long cycle time and can thus be highly multiplexed enabling system wide kinase activity screening within a single LC-MS run per sample and condition, however at the cost of lower confidence and quantitative accuracy. Quantification mode operates with a higher number of transitions per peptide (including phosphosite localization specific transitions), long dwell times and a short cycle time enabling accurate quantification. (B) Samples subjected to kinase activity profiling are lysed and digested individually and supplemented with equal amounts of heavy isotope labeled peptides used as internal standards. Subsequently samples were enriched for phosphopeptides by Fe(III)-IMAC and subjected to SRM analysis. One sample per condition was subjected to an analysis in survey mode, providing an initial indication of the detectable targets. Potential hits are then measured in quantification mode for each sample, thus enabling an accurate quantification of kinase activations on a system wide level across all measured conditions.*

To alleviate these shortcomings, we demonstrate an alternative method to examine system wide kinase activation states. Our approach is based on the concept of t-loop phosphorylations, which are critical switches initiating activity in most kinases [13-15]. Notably, the functionality of such sites can be directly transferred to numerous understudied kinases based on their high sequence similarities [16-18]. This enables a direct link between phosphosite abundance and kinase activaton. One of the major challenges in this approach is that t-loop phosphorylations are notoriously underrepresented in shotgun proteome analyses due to their low abundance, unfavorable LC-MS characteristics and high prevalence of tyrosine phosphorylations. In this study we exploited the high specificity of Fe(III)-IMAC phosphopeptide enrichment [19] in combination with the unparalleled sensitivity and selectivity of nano-LC-MS/MS operated in SRM mode [20] to monitor system wide t-loop phosphorylations as a probe for kinase activity (Figure 1).

Based on stable isotope labelled synthetic peptide standards, we developed SRM assays for 248 peptides covering 221 phosphorylation sites in the t-loop region of 178 kinases (33 % of the human kinome [21]). The targeted kinases included numerous clinically relevant kinases with FDA approved inhibitors such as MET, ABL, SRC, BTK, JAK3 and KIT (Figure 2 and Table S1) [22].

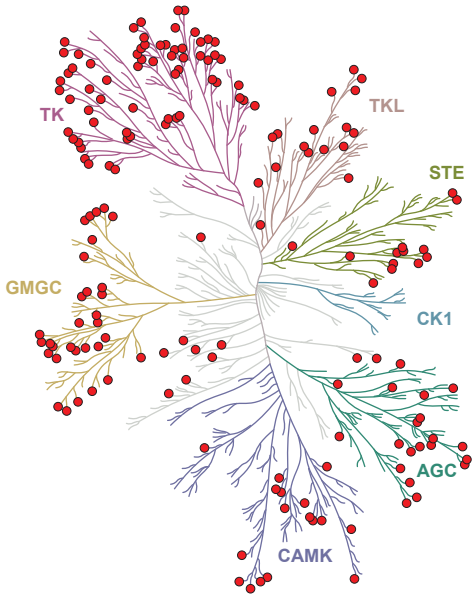


Figure 2: ***Human kinome tree highlighting kinases included in our kinase activity profiling approach.*** *Assays were built for a total of 248 peptides covering 221 phosphorylation sites in the t-loop region of 178 human kinases (covering 33 % of the human kinome).*

The targeted assay construction entailed several analytical challenges, most notable being the sheer number of target peptides, which required a tight control of SRM multiplexing

parameters such as cycle time, dwell time, RT scheduling, and gradient length. Common strategies to account for a high number of target peptides, i.e. limiting the number of transitions per peptide or distributing assays over multiple methods, can compromise assay specificity and sensitivity or increase the demand on input material. However, when developing kinase activation assays it is imperative to achieve highest sensitivity to handle the low abundance of the target peptides in combination with sufficient selectivity to accurately determine phosphosite localization. Additionally, prediction of interferences is impracticable when applying assays to various sample types. To account for these challenges, we developed a two-phase strategy comprising a survey and a quantification phase (Figure 1A and Table S2). The survey phase encompasses a 'low-confidence' method consisting of assays for all 248 peptides of interest, not stringent enough for reliable quantification but providing valuable information about which t-loop phosphorylations are detectable in the sample at hand. The results obtained from the survey phase then determined the method characteristics of the subsequent 'high confidence' quantification phase. These quantitative assays are highly sample specific as they encompass fully optimized assays for all detected t-loop phosphorylations in the survey phase, including a high number of transitions per peptide and low cycle and high dwell times (Figure 1B and Table S2).

To test this two-phase strategy, we analyzed the baseline kinome activity state of three different human cell lines without any form of stimulation. The analyzed cell lines were (1) Jurkat cells, an immortalized line of T lymphocytes (2) PC9 cells, a non-small cell lung cancer cell line and (3) Hek239 cells, a cell line derived from human embryonic kidney cells. Using an input amount of 250 µg of digested proteins per phosphopeptide enrichment, this resulted cumulatively in the detection of 52 t-loop phosphorylation sites for the three cell lines (Figure 3 and Table S3). Due to the highly conserved nature of the kinases' t-loop sequence the representative tryptic peptides are not always unique. This results in a certain amount of ambiguity, for instance for the kinase family members Mark1, Mark2, Mark3 and Mark4, or closely related kinases that often exhibit redundant functions such as the tyrosine kinases FYN, YES1 and SRC [23]. To deal with this ambiguity we followed the principle of protein grouping [24] and refer to these instances as kinase groups throughout this study. For the 52 phosphorylation sites observed, this resulted in 48 kinase groups. Moreover, while the t-loop is clearly defined through the flanking DFG and (A)PE motifs [16], phosphorylation can still occur at various or even multiple residues. Whereas for numerous kinases the activation residue(s) are clearly established, for others these sites within the t-loop are not known. Therefore, for the unknown cases we took into account multiple possibilities resulting in the development of SRM assays for various phosphosite isomers (Supporting Information Figure S1).

A large part of the 48 detected kinase groups represented typical housekeeping kinases crucial for growing cells in typical culture conditions, such as CDKs and MAPKs as well as the two abundant kinases PDK1 and GSK3. Additionally, several kinases involved in anti-

apoptotic processes were detected in an active state such as HIPK3 [25]. Both Jurkat and PC9 cells showed an increased activity of Ca^{2+}/DAG dependent signaling compared to Hek cells, with several kinases from the CaMK group and the PKC family being detected in their active state. These included CaMK1δ, PKCθ, and the kinase group PKCα/PKCβ/PKCγ. CaMKIV activity on the other hand was detected exclusively in Jurkat cells. Interestingly many kinases showed cell line dependent activity profiles. Some of them are known to be expressed in a tissue specific manner, such as the tyrosine kinase ZAP70 which is exclusively expressed in certain cell types associated with the immune system, including T-cells [26]. Accordingly, t-loop phosphorylation of ZAP70 was exclusively detected in Jurkat cells. Other tyrosine kinases such as FAK, MET and the two kinase groups EPHA3/4/5 and HCK/LYN could not be detected in Jurkat cells whereas they showed high activity especially in PC9 cells. This is an indication for elevated tyrosine kinase activity in PC9, likely due to activated EGFR signaling, which is a common alteration in non-small cell lung cancer [27]. Notably, these activating phosphorylations occur on tyrosine residues within the t-loop sequence, which control the primary activation for a substantial number of kinases. Especially, tyrosine kinases and MAP kinases require tyrosine phosphorylation in their t-loop

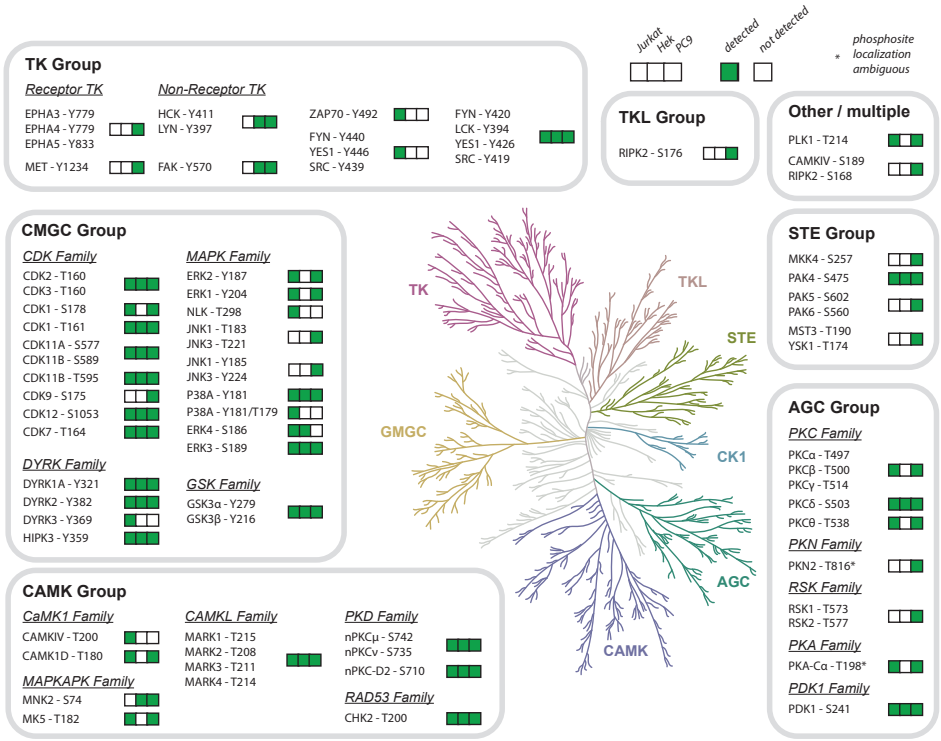


Figure 3: ***Detected baseline kinase activity in three cell lines.*** *Compendium of t-loop phosphorylations detected in Jurkat, Hek and PC9 cell lines cultured under standard conditions without any form of stimulation.*

for full activation, while the primary activation site for most other kinases is a threonine residue. Tyrosine phosphorylations are naturally underrepresented in phosphoproteome analyses unless specific phosphotyrosine enrichment is performed upfront, not compatible with our current approach due to the high sample amount constraint. Nonetheless, the Fe(III)-IMAC phosphopeptide enrichment in combination with sensitive SRM analysis in our technology resulted in a substantial recovery of tyrosine t-loop phosphorylations.

After the successful detection of several t-loop phosphorylations in unstimulated cells, we reasoned that since a large part of the kinome will be present in an inactive (unphosphorylated) state, our technique should be able to reveal activation of specific kinases from the steady-state background upon selected stimuli. To demonstrate this capacity, we treated Jurkat cells with TNFα for 8 h, which resulted in increased cell death (Figure 4A). Upon TNFα stimulation, the receptor-interacting protein serine-threonine kinase (RIPK) is recruited to the TNF receptor complex and mediates apoptosis (Figure 4B) [28]. Indeed, we were able to reproducibly detect RIPK1 phosphorylation at serine S161 already upon TNFα treatment for 8 h, a phosphorylation not detectable in untreated Jurkat cells (Figure 4C). Remarkably, despite the well characterized role of RIPK1 in cell death, to the best of our knowledge this presents the first direct detection by MS of RIPK1 t-loop phosphorylation from cell lysates. Thus far, the only report of successful MS detection of phosphorylated S161 in RIPK1 was described by Degterev *et al.* [29] using a protein expression system in combination with IP, *in vitro* kinase assay and phosphopeptide enrichment. At the same time, their mutation analyses also conclusively revealed the essential role of S161 phosphorylation in RIPK1 kinase activation. Hence, our RIPK1 assay provides a new and robust

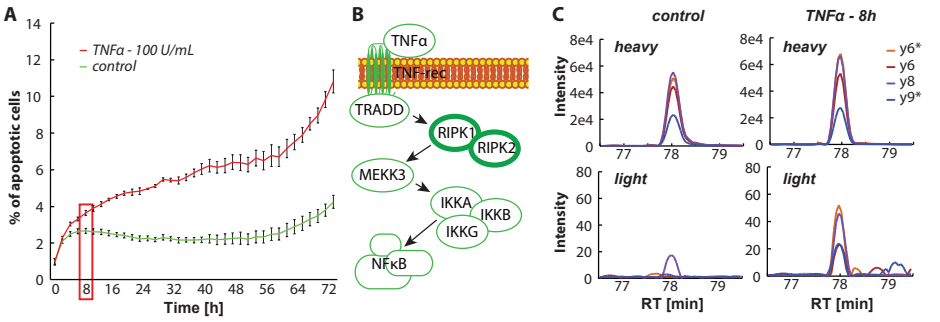


Figure 4: ***Detection of RIPK1 S161 phosphorylation upon TNFα treatment.*** *(A) Percentage of apoptotic cells in Jurkat cultures treated with 100 U/mL TNFα determined by Caspase-3/7 green apoptosis reagent (red line). Untreated cells were used as a control (green line, n = 4 for both groups, error bars depicting standard deviation). (B) Schematic depiction of TNFα signaling. Upon TNF receptor stimulation the receptor-interacting serine/threonine-protein kinases 1 and 2 (RIPK1 and RIPK2) are recruited to the TNF receptor complex, mediating apoptosis. (C) Representative SRM traces of RIPK1 t-loop phosphorylation at serine 161. Upper panels represent internal heavy labeled standard peptides and lower panels signals from endogenous peptides (* neutral loss of phosphoric acid on the fragment ions). Phosphorylation of RIPK1 at serine 161 is exclusively detectable upon TNFα treatment already at early time points (e. g. 8 h, indicated as red rectangle in panel A).*

readout to monitor the complex regulation of cell death, while it also demonstrates the sensitivity of our technology.

We next wanted to exploit the sensitivity of our kinase activity profiling approach, by performing in-depth analyses of kinome dynamics in primary cells. We applied our technique to study the mechanism of platelet activation. For this we analyzed blood platelets activated by the SFLLRN hexapeptide, mimicking thrombin-induced activation of protease-activated receptor 1 (PAR1). We performed PAR1 activation for 1 min and 5 min and compared the platelets' kinase activity profile to non-activated blood platelets [30]. Using 300 µg of input material per sample, we were able to detect and quantify 31 t-loop phosphorylations in 25 kinase groups (Table S3). The comparative quantitative analysis between activated and non-activated platelets revealed drastic changes of kinome activity levels upon PAR1 activation. Figure 5A depicts the quantitative differences as volcano plots for 1 min and 5 min activation, respectively. Platelet activation involves various intracellular signaling events, the majority of which converge into the common pathway depicted in Figure 5B [31, 32]. Noteworthy here is the detection of t-loop phosphorylations in the two TEC family tyrosine kinases BTK and TEC, which have both been associated with platelet activation. Especially the role of BTK as a major activator of PLCγ2 is well established [33], while TEC has been connected to a more compensatory role when BTK is absent or malfunctioning [34], e.g. in case of X-linked agammaglobulineamia [35]. The kinase activity profiles presented here corroborate these essential roles for BTK and TEC in platelet activation. Interestingly, the magnitude and kinetics of kinase activation differed between the two TEC family kinases. Where BTK activity increased almost 4-fold upon 1 min PAR1 activation and kept increasing at 5 min, TEC showed much lower activation and lacked behind in time, corroborating the leading role of BTK (Figure 5B).

Downstream, a central converging point after PAR1 stimulation is the activation of phospholipase C (PLC) leading to an increase in intracellular Ca^{2+} and thereby increased activity of calcium/calmodulin-dependent kinases (CaMK). With our assay we could monitor activation states of CaMK2 through quantification of two adjacent phosphosites in the t-loop (T305 and T306, Figure 5B). Both phosphosites showed a drastic spike upon 1 min PAR1 activation, which dropped substantially after 5 min, demonstrating the short timescales involved in kinase activation. Interestingly, the specific kinetics of both phosphosites differed slightly. Whereas the T306 phosphorylation increased more than 10-fold after 1 min and returned to almost baseline after 5 min, the T305 spike was much less intense but showed a slower attenuation, suggesting dynamic regulation of the two phosphorylation events occurring in the CaMK2 t-loop.

Another target of PLC activity is PKC, which in turn activates p38 [32]. With our technique we could quantify several t-loop phosphorylations of PKC family members, including, PKCδ, PKCθ and the redundant t-loop sequence of PKCα, PKCβ and PKCγ. Interestingly, none of them showed any significant difference, while their downstream target p38A

changed quite drastically. P38A requires double phosphorylation at T179 and Y181 for full kinase activity [36], a state we observed especially pronounced after 1 min PAR1 activation and which drastically decreased after 5 min (Figure 5B). In parallel, singly phosphorylated p38A at Y181 remained at baseline levels after 1 min and only slightly increased after 5 min. This suggests a rapid double phosphorylation of p38A followed by a slower, partial, dephosphorylation resulting in the observed upregulation of single Y181 phosphorylation after 5 min, in line with previous studies [37-40]. Combining the observed PKC and p38 t-loop phosphorylation implies either an extremely fast spike in PKC activity, already disappearing in less than 1 min, but able to activate p38 further downstream, or a massif signal amplification of p38 activation compared to PKC activity levels.

Another well studied effect of increasing intracellular Ca^{2+} levels is activation of RAS, via its translocation to the plasma membrane, and the subsequent activation of the MAPK cascade [41]. The role of the MAPK cascade in platelets has not yet been fully elucidated, since platelets are anucleate cells with no potential to grow, differentiate or proliferate, however it seems to be a relevant factor in maintaining elevated intracellular Ca^{2+} levels [42]. Our experiments indeed showed an increase in kinase activity for individual members of the MAPK signaling cascade. A slight increase in RAF activity was observed upon PAR1 activation for 1 min, leading to strong activation ERK1 and ERK2 (Figure 5B). Finally, we also observed several t-loop phosphorylations for kinases functioning in alternative platelet activation routes, e.g. Fyn/ Yes, Lyn, MKK4, JNK2 and FAK2 [43], however no substantial differences in phosphorylation were observed suggesting the presence of baseline activity (Figure 5C).

Following the effective analysis of kinase activation in primary cells, we wanted to explore the usefulness of our technology to study unbalanced activity of kinases in disease. Kinases have become a major class of drug targets, especially in cancer where 25 kinase-targeting drugs have been approved and numerous candidates are under clinical evaluation [44]. However, in the identification of these candidates, through synthetic lethality screens, the (long-term) effect of inhibition of one kinase on the rest of the kinome is often neglected. This consistently leads to treatment resistance to targeted kinase inhibition due to adaptation of signaling networks.

To demonstrate the potential of our technology we studied acquired drug resistance in melanoma. A majority of melanoma is driven by a $BRAF^{V600E}$ mutation resulting in constitutive activity of BRAF [45]. Despite initial success, treatment of patients with BRAF inhibitors (BRAFi) usually results in rapid acquisition of acquired drug resistance [46]. Here, we studied t-loop phosphorylations in matched patient-derived melanoma cell lines from treatment naïve, treatment sensitive and resistant tumor states established from patient-derived xenografts (Figure 6A), where the acquired drug resistance is based on an $NRAS^{Q61K}$ mutation [47].

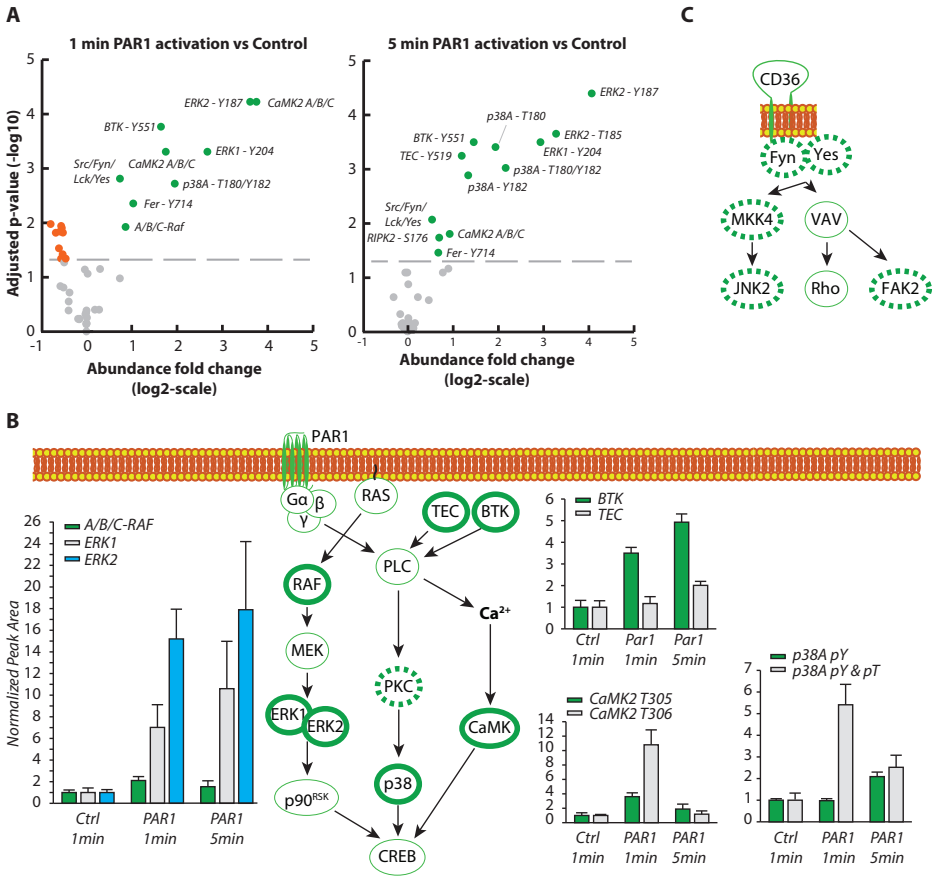


Figure 5: ***Kinase activity profiling in PAR1 activated human blood platelets.*** *(A) Observed changes in kinase activity upon PAR1 activation of platelets for 1 min and 5 min respectively, depicted as volcano plot (significance cutoff p < 0.05). (B) Molecular signaling network involved in platelet activation. Kinases outlined in bold green showed changes in activity upon PAR1 activation, kinases outlined in dashed bold green showed no change in activity upon PAR1 activation. Bar graphs show representative dynamic regulation of kinase activity for key players in the pathway such as RAF/ERK, BTK/TEC, CaMK2 and p38A. (Quantification based on Heavy-to-Light Ratio, normalized on Ctrl sample, error bars depict standard deviation) (C) Schematic representation of alternative signaling pathways known to be involved in platelet activation. Activation of Fyn/Yes, MKK4, JNK2 and FAK2 were observed, however no quantitative change could be observed upon PAR1 activation suggesting baseline activity.*

Through differential comparison of kinome activity in all three states we were able to detect and quantify 39 phosphosites representing t-loop phosphorylations of 36 kinase groups (Figure 6B-D and Table S3). Several of the quantified kinases showed increased activity in the resistant cell line compared to the treatment naïve and sensitive cells. These kinases included CaMKIV, several members of the PKC kinase family such as PKCδ, PKCμ/v, several members associated to the MAPK cascade such as p38A, ERK2, ERK4 and MKK4 as well as the MAPK effector kinase NLK and cell cycle related kinases CDK2/3 and Chk2.

Surprisingly, several kinases specifically activated in the drug resistant cell line have mainly been linked to tumor suppressing activities such as Chk2 [48], p38A [49] and ERK4.

The strongest activation we observed was that of the kinase ERK4, which, together with the also observed ERK3, belongs to the atypical MAPK family due to their lack of a tyrosine domain in the t-loop. Thus far, MK5 is the only known substrate of both ERK3 and ERK4 [50, 51]. MK5 is mainly known for tumor suppressing functionality such as activation of p53 [52] and FOXO3 [53]. Recent studies however also showed oncogenic potential for the ERK3/ERK4/MK5 module, for instance through inhibition of JNK activity [54] and support of angiogenesis [55]. Additionally, increased mRNA expression of MK5 was linked to increased probability of the development of metastasis [53].

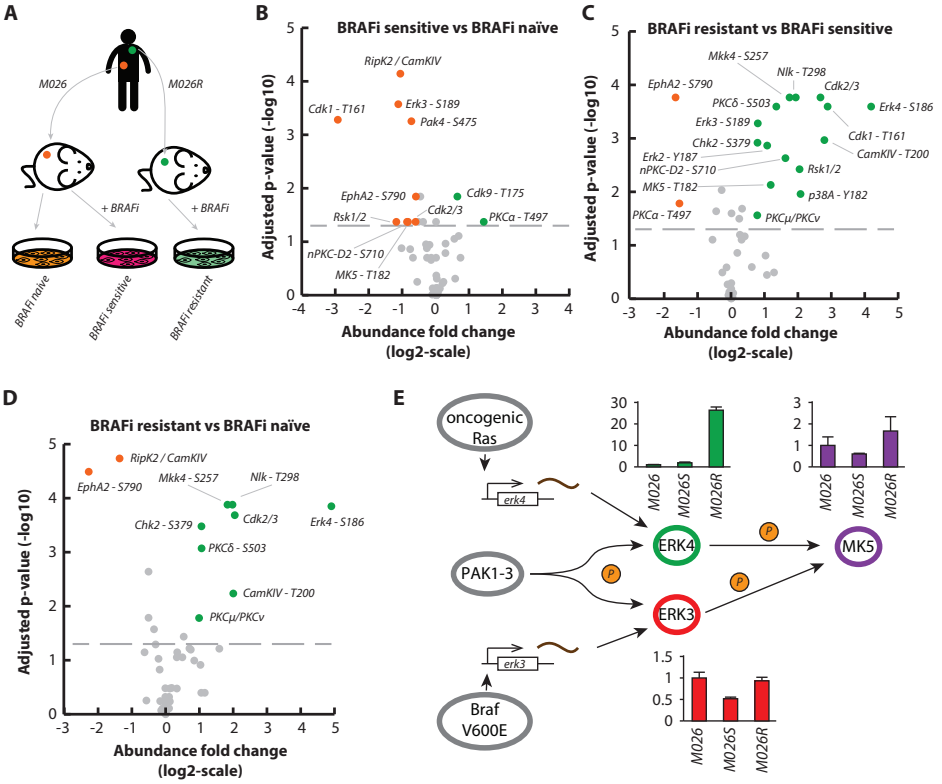


Figure 6: ***Reorganization of kinase activities upon acquired BRAF inhibitor resistance.*** *(A) Matched melanoma cell lines were established from PDXs from the same patient before and after acquired resistance to the BRAF inhibitor PLX-4720, giving rise to a model system comprising treatment naïve, treatment sensitive and treatment resistant cell lines. (B-D). Pairwise comparison of kinase activity between the 3 conditions depicted as volcano plot using arbitrary significance cutoff ($p < 0.05$) and fold change cutoff (1.5). (E) Molecular interaction of the ERK3/ERK4/MK5 system and the detected abundance of t-loop phosphorylation (normalized to values observed in naïve cells, error bars depicting standard deviation).*

Interestingly, the dynamics of ERK3/ERK4 expression has been linked to increased expression of both $BRAF^{V600E}$ and oncogenic RAS, leading to increased expression of ERK3 and ERK4, respectively [56, 57]. Our data reveals that inhibition of $BRAF^{V600E}$ goes in parallel with a decrease in t-loop phosphorylation of ERK3, which is accompanied with a decrease in MK5 (Figure 6E). ERK3 levels are restored upon acquired drug resistance and strikingly, oncogenic RAS drastically elevates ERK4 t-loop phosphorylation, over 20-fold compared to the naïve and sensitive cells. Globally this leads to a higher MK5 kinase activity in resistant cells when compared to treatment naïve cells. Our results indicate a possible interesting role of the ERK3/ERK4/MK5 system in NRAS driven BRAFi resistance in melanoma and highlight the potential of our technology to detect altered kinome activity upon targeted oncotherapy.

In conclusion, we have developed a robust approach to quantify system wide t-loop phosphorylations on human protein kinases. We provide carefully optimized assays for t-loop phosphorylation on 178 protein kinases, accounting for roughly 33 % of the global human kinome. The strength of the technology lies in the combination of sensitivity, enabling substantial kinome coverage even from limited starting material (around 250 µg protein per sample), and throughput, using an automated platform with parallel phosphopeptide enrichment of up to 96 samples. Applying this approach, we were able to detect 66 t-loop phosphorylations for 54 kinases across numerous cell types (primary and patient-derived, Table S3), which to the best of our knowledge presents the largest compendium of kinase activation sites reported to date. The use of sample fractionation or the enrichment for specific cell compartments such as nuclei or cell membrane will further increase kinome coverage. We demonstrate the utility of our technology for exclusive detection of activated kinases upon selected stimuli or specific treatment regimes. Our technology reveals, besides known signaling pathway activation, t-loop phosphorylations of the elusive RIPK1 S161, dynamic activation of primary blood platelets and potential rewiring of signal transduction in melanoma upon acquired drug resistance.

Acknowledgments

A.F.M.A. is supported by the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (723.012.102). This work is part of the project Proteins at Work, a program of the Netherlands Proteomics Centre financed by the Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-scale Research Facilities of the Netherlands (project number 184.032.201). We like to thank Dr. Wei Wu and T. Celine Mulder for providing cell pellets used in the proof of principle analysis. Dr. Kelly Stecker is acknowledged for initial support in performing phosphopeptide enrichment.

References

[1] Blume-Jensen, P., Hunter, T., Oncogenic kinase signalling. *Nature* 2001, *411*, 355-365.

[2] Kizub, I. V., Klymenko, K. I., Soloviev, A. I., Protein kinase C in enhanced vascular tone in diabetes mellitus. *Int J Cardiol* 2014, *174*, 230-242.

[3] Malek, R., Davis, S. N., Tyrosine kinase inhibitors under investigation for the treatment of type II diabetes. *Expert Opin Investig Drugs* 2016, *25*, 287-296.

[4] Shahin, R., Shaheen, O., El-Dahiyat, F., Habash, M., Saffour, S., Research advances in kinase enzymes and inhibitors for cardiovascular disease treatment. *Future Sci OA* 2017, *3*, FSO204.

[5] Garay, J. P., Gray, J. W., Omics and therapy - a basis for precision medicine. *Mol Oncol* 2012, *6*, 128-139.

[6] Daub, H., Olsen, J. V., Bairlein, M., Gnad, F*., et al.*, Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol Cell* 2008, *31*, 438-448.

[7] Ranjitkar, P., Perera, B. G., Swaney, D. L., Hari, S. B*., et al.*, Affinity-based probes based on type II kinase inhibitors. *J Am Chem Soc* 2012, *134*, 19017-19025.

[8] Médard, G., Pachl, F., Ruprecht, B., Klaeger, S*., et al.*, Optimized chemical proteomics assay for kinase inhibitor profiling. *J Proteome Res* 2015, *14*, 1574-1586.

[9] Ruprecht, B., Zecha, J., Heinzlmeir, S., Médard, G*., et al.*, Evaluation of Kinase Activity Profiling Using Chemical Proteomics. *ACS Chem Biol* 2015, *10*, 2743-2752.

[10] Ding, V. M., Boersema, P. J., Foong, L. Y., Preisinger, C*., et al.*, Tyrosine phosphorylation profiling in FGF-2 stimulated human embryonic stem cells. *PLoS One* 2011, *6*, e17538.

[11] Lemeer, S., Heck, A. J., The phosphoproteomics data explosion. *Curr Opin Chem Biol* 2009, *13*, 414-420.

[12] Fujii, K., Zhu, G., Liu, Y., Hallam, J*., et al.*, Kinase peptide specificity: improved determination and relevance to protein phosphorylation. *Proc Natl Acad Sci U S A* 2004, *101*, 13744-13749.

[13] Newton, A. C., Regulation of the ABC kinases by phosphorylation: protein kinase C as a paradigm. *Biochem J* 2003, *370*, 361-371.

[14] Wiechmann, S., Czajkowska, H., de Graaf, K., Grötzinger, J*., et al.*, Unusual function of the activation loop in the protein kinase DYRK1A. *Biochem Biophys Res Commun* 2003, *302*, 403-408.

[15] Meyer, R. D., Mohammadi, M., Rahimi, N., A single amino acid substitution in the activation loop defines the decoy characteristic of VEGFR-1/FLT-1. *J Biol Chem* 2006, *281*, 867-875.

[16] Nolen, B., Taylor, S., Ghosh, G., Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell* 2004, *15*, 661-675.

[17] Adams, J. A., Activation loop phosphorylation and catalysis in protein kinases: is there functional evidence for the autoinhibitor model? *Biochemistry* 2003, *42*, 601-607.

[18] Mendrola, J. M., Shi, F., Park, J. H., Lemmon, M. A., Receptor tyrosine kinases with intracellular pseudokinase domains. *Biochem Soc Trans* 2013, *41*, 1029-1036.

[19] Post, H., Penning, R., Fitzpatrick, M. A., Garrigues, L. B*., et al.*, Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* 2017, *16*, 728-737.

[20] de Graaf, E. L., Kaplon, J., Mohammed, S., Vereijken, L. A*., et al.*, Signal Transduction Reaction Monitoring Deciphers Site-Specific PI3K-mTOR/MAPK Pathway Dynamics in Oncogene-Induced Senescence. *J Proteome Res* 2015, *14*, 2906-2914.

[21] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., Sudarsanam, S., The protein kinase complement of the human genome. *Science* 2002, *298*, 1912-1934.

[22] Fabbro, D., Cowan-Jacob, S. W., Moebitz, H., Ten things you should know about protein kinases: IUPHAR Review 14. *Br J Pharmacol* 2015, *172*, 2675-2700.

[23] Lin, Y. C., Huang, J., Kan, H., Frisbee, J. C., Yu, H. G., Rescue of a trafficking defective human pacemaker channel via a novel mechanism: roles of Src, Fyn, and Yes tyrosine kinases. *J Biol Chem* 2009, *284*, 30433-30440.

[24] Nesvizhskii, A. I., Aebersold, R., Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* 2005, *4*, 1419-1440.

[25] Curtin, J. F., Cotter, T. G., JNK regulates HIPK3 expression and promotes resistance to Fas-mediated apoptosis in DU 145 prostate carcinoma cells. *J Biol Chem* 2004, *279*, 17090-17100.

[26] Chan, A. C., Iwashima, M., Turck, C. W., Weiss, A., ZAP-70: a 70 kd protein-tyrosine kinase that associates with the TCR zeta chain. *Cell* 1992, *71*, 649-662.

[27] Sharifnia, T., Rusu, V., Piccioni, F., Bagul, M.*, et al.*, Genetic modifiers of EGFR dependence in non-small cell lung cancer. *Proc Natl Acad Sci U S A* 2014, *111*, 18661-18666.

[28] Holler, N., Zaru, R., Micheau, O., Thome, M.*, et al.*, Fas triggers an alternative, caspase-8-independent cell death pathway using the kinase RIP as effector molecule. *Nat Immunol* 2000, *1*, 489-495.

[29] Degterev, A., Hitomi, J., Germscheid, M., Ch'en, I. L.*, et al.*, Identification of RIP1 kinase as a specific cellular target of necrostatins. *Nat Chem Biol* 2008, *4*, 313-321.

[30] Coughlin, S. R., Thrombin signalling and protease-activated receptors. *Nature* 2000, *407*, 258-264.

[31] Estevez, B., Du, X., New Concepts and Mechanisms of Platelet Activation Signaling. *Physiology (Bethesda)* 2017, *32*, 162-177.

[32] Martorell, L., Gentile, M., Rius, J., Rodríguez, C.*, et al.*, The hypoxia-inducible factor 1/NOR-1 axis regulates the survival response of endothelial cells to hypoxia. *Mol Cell Biol* 2009, *29*, 5828-5842.

[33] Laffargue, M., Ragab-Thomas, J. M., Ragab, A., Tuech, J.*, et al.*, Phosphoinositide 3-kinase and integrin signalling are involved in activation of Bruton tyrosine kinase in thrombin-stimulated platelets. *FEBS Lett* 1999, *443*, 66-70.

[34] Atkinson, B. T., Ellmeier, W., Watson, S. P., Tec regulates platelet activation by GPVI in the absence of Btk. *Blood* 2003, *102*, 3592-3599.

[35] Tsukada, S., Saffran, D. C., Rawlings, D. J., Parolini, O.*, et al.*, Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell* 1993, *72*, 279-290.

[36] Cargnello, M., Roux, P. P., Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiol Mol Biol Rev* 2011, *75*, 50-83.

[37] Li, Z., Xi, X., Du, X., A mitogen-activated protein kinase-dependent signaling pathway in the activation of platelet integrin alpha IIbbeta3. *J Biol Chem* 2001, *276*, 42226-42232.

[38] Li, Z., Zhang, G., Feil, R., Han, J., Du, X., Sequential activation of p38 and ERK pathways by cGMP-dependent protein kinase leading to activation of the platelet integrin alphaIIb beta3. *Blood* 2006, *107*, 965-972.

[39] Saklatvala, J., Rawlinson, L., Waller, R. J., Sarsfield, S.*, et al.*, Role for p38 mitogen-activated protein kinase in platelet aggregation caused by collagen or a thromboxane analogue. *J Biol Chem* 1996, *271*, 6586-6589.

[40] Papkoff, J., Chen, R. H., Blenis, J., Forsman, J., p42 mitogen-activated protein kinase and p90 ribosomal S6 kinase are selectively phosphorylated and activated during thrombin-induced platelet activation and aggregation. *Mol Cell Biol* 1994, *14*, 463-472.

[41] Rosado, J. A., Sage, S. O., Farnesylcysteine analogues inhibit store-regulated Ca2+ entry in human platelets: evidence for involvement of small GTP-binding proteins and actin cytoskeleton. *Biochem J* 2000, *347 Pt 1*, 183-192.

[42] Rosado, J. A., Sage, S. O., Role of the ERK pathway in the activation of store-mediated calcium entry in human platelets. *J Biol Chem* 2001, *276*, 15659-15665.

[43] Zimman, A., Podrez, E. A., Regulation of platelet function by class B scavenger receptors in hyperlipidemia. *Arterioscler Thromb Vasc Biol* 2010, *30*, 2350-2356.

[44] Gross, S., Rahal, R., Stransky, N., Lengauer, C., Hoeflich, K. P., Targeting cancer with kinase inhibitors. *J Clin Invest* 2015, *125*, 1780-1789.

[45] Davies, H., Bignell, G. R., Cox, C., Stephens, P*., et al.*, Mutations of the BRAF gene in human cancer. *Nature* 2002, *417*, 949-954.

[46] Wagle, N., Emery, C., Berger, M. F., Davis, M. J*., et al.*, Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol* 2011, *29*, 3085-3096.

[47] Kemper, K., Krijgsman, O., Cornelissen-Steijger, P., Shahrabi, A*., et al.*, Intra- and inter-tumor heterogeneity in a vemurafenib-resistant melanoma patient and derived xenografts. *EMBO Mol Med* 2015, *7*, 1104-1118.

[48] Miyasaka, Y., Nagai, E., Yamaguchi, H., Fujii, K*., et al.*, The role of the DNA damage checkpoint pathway in intraductal papillary mucinous neoplasms of the pancreas. *Clin Cancer Res* 2007, *13*, 4371-4377.

[49] Wagner, E. F., Nebreda, A. R., Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat Rev Cancer* 2009, *9*, 537-549.

[50] Perander, M., Keyse, S. M., Seternes, O. M., Does MK5 reconcile classical and atypical MAP kinases? *Front Biosci* 2008, *13*, 4617-4624.

[51] Perander, M., Aberg, E., Johansen, B., Dreyer, B*., et al.*, The Ser(186) phospho-acceptor site within ERK4 is essential for its ability to interact with and activate PRAK/MK5. *Biochem J* 2008, *411*, 613-622.

[52] Sun, P., Yoshizuka, N., New, L., Moser, B. A*., et al.*, PRAK is essential for ras-induced senescence and tumor suppression. *Cell* 2007, *128*, 295-308.

[53] Kress, T. R., Cannell, I. G., Brenkman, A. B., Samans, B*., et al.*, The MK5/PRAK kinase and Myc form a negative feedback loop that is disrupted during colorectal tumorigenesis. *Mol Cell* 2011, *41*, 445-457.

[54] Chen, G., Hitomi, M., Han, J., Stacey, D. W., The p38 pathway provides negative feedback for Ras proliferative signaling. *J Biol Chem* 2000, *275*, 38973-38980.

[55] Yoshizuka, N., Chen, R. M., Xu, Z., Liao, R*., et al.*, A novel function of p38-regulated/activated kinase in endothelial cell migration and tumor angiogenesis. *Mol Cell Biol* 2012, *32*, 606-618.

[56] Hoeflich, K. P., Eby, M. T., Forrest, W. F., Gray, D. C*., et al.*, Regulation of ERK3/MAPK6 expression by BRAF. *Int J Oncol* 2006, *29*, 839-849.

[57] Lee, S., Kang, J., Cho, M., Seo, E*., et al.*, Profiling of transcripts and proteins modulated by K-ras oncogene in the lung tissues of K-ras transgenic mice by omics approaches. *Int J Oncol* 2009, *34*, 161-172.

Supporting Information

Materials and Methods

Materials

Heavy labeled phosphopeptide representing t-loop phosphorylations were purchased from JPT (Berlin, Germany). HRM Calibration KIT was purchased from Biognosys (Schlieren, Switzerland) and used as internal retention time and injection amount control at concentrations of 0.1x/µL.

Cell culture

PC9 NSCLC cells were purchased from Sigma-Aldrich and were cultured in standard Roswell ParkMemorial Institute medium 1640 medium (Lonza), containing 10% FBS (Thermo), 2 mM L-glutamine and 1% penicillin/streptomycin (Lonza), at 37 °C in a humidified atmosphere containing 5% CO2. Cells were detached from the culture surface using trypsin (Lonza), and washed three times with PBS before lysis. HEK293T cells were seeded at 15% density in 15-cm plates, allowed to adhere in full DMEM (Lonza) containing 10% heat-inactivated fetal bovine serum (Gibco), 2mM L-glutamine (Lonza) and 20mM HEPES(Sigma-Aldrich), and cultured to ~90% confluence over 2.5 days. Twelve hours prior to harvesting, growth medium was replaced with fresh pre-warmed full DMEM. At harvesting, no dead or floating cells were visible by microscopic examination. Cells were washed twice with ice-cold PBS on-plate, detached by trypsin (Lonza), and collected by low-speed centrifugation at 200*g* for 5min. Jurkat cells were cultured in RPMI-1640 media complemented with l-glutamine, 10% fetal bovine serum, and PenStrep at a density between 5 x105 and 5 x106 per ml. One day prior stimulation, cells were seeded at a concentration of 1 x106 per ml in fresh media. TNFα (PeproTech, USA) was added at a concentration of 100 U/ml for eight hours, after which the cells were washed with ice cold PBS and snap frozen.

PDX melanoma cell lines (M026, M026R) were cultured in RPMI 1640 with HEPES and L-Glutamine supplemented with 10 % fetal bovine serum and Penicillin-Streptomycin. The resistant cell lines were grown in presence of 1 µM PLX-4720. When specified, the cell lines were treated with an IC50 dose (1 µM) of the inhibitors PLX-4720 (Selleckchem). All cell pellets were stored at -80 °C prior to cell lysis.

Pooled platelet concentrates in plasma (platelet concentrates from 5 donors in ABO blood-group-matched plasma) were obtained from the Dutch Bloodbank (Sanquin). Multiple platelet concentrates were pooled in three separate pools and processed independently. Platelets were isolated as described before [1]. Briefly, platelet concentrates were centrifuged for 20 minutes at 120g, to remove remaining red and white blood cells. Next, platelets were washed twice with isolation buffer (36 mM citric acid, 103 mM NaCl, 5 mM KCl, 5 mM ethylenediaminetetraacetic acid (EDTA), 5.6 mM D-glucose, pH 6.5, containing 0.35% [w/v] bovine serum albumin (BSA)) by centrifugation for 10 minutes at 2,000g. Platelets were washed once with Tyrode's Solution (Sigma), centrifuged for 10 minutes at

2,000g and resuspended to a final concentration of 2 × 108 platelets/mL in Tyrode's Solution supplemented with tirofiban (1:1000) (Iroko cardio, UK). Platelet were activated with 50 µM PAR1 peptide (SFLLRN-NH2, Peptides international) or negative control (Tyrode's solution) for 2 or 5 minutes at 37 °C in a thermomixer (Eppendorf). Activation was stopped by lysis of the platelets by addition of 2 times (% v/v) 1.5x SDC lysis buffer (1.5% SDC, 15 mM TCEP, 60 mM Chloroactamide, 150 mM tris(hydroxymethyl)aminomethane (Tris) pH 8.5, 1.5% phosphatase and protease inhibitor cocktail (Thermo Scientific)). Platelet lysates were snap frozen and stored at -80 °C.

Apoptosis assay

Jurkat cells were plated at a density of 30,000 cells in a 96 well plate coated with 0.01% poly-l-ornithine. Cells were stimulated with TNFα at a concentration of 100U/ml in the presence of 2.5 µM Caspase-3/7 green apoptosis reagent (Essen Bioscience). Apoptosis was monitored using the IncuCyte Zoom system (Essen Bioscience) every hour for 24 hours. Each well was divided into four views, and the number of green fluorescent cells was counted using the IncuCyte Zoom software (Essen Bioscience).

Sample Preparation and Phosphopeptide enrichment

Frozen cell pellets were lysed, reduced and alkylated in lysis buffer (1% sodium deoxycholate (SDC), 10 mM tris(2-carboxyethyl)phosphinehydrochloride (TCEP)), 40 mM chloroacetamide (CAA), and 100 mM TRIS, pH 8.0 supplemented with phosphotase inhibitor (PhosSTOP, Roche) and protease inhibitor (cOmplete mini EDTA-free, Roche). Cells were heated at 95 °C and sonicated with a Bioruptor Plus (Diagenode) for 15 cycles of 30 s. Bradford protein assay was used to determine protein amount. To avoid digestion bias, samples were split into aliquots containing equal amount of protein. Proteins were digested overnight at 37 °C with trypsin (Sigma-Aldrich) with an enzyme/substrate ratio of 1:50 and lysyl endopeptidase (Wako) with an enzyme/substrate ratio of 1:75. SDC was precipitated with 2% formic acid (FA) and samples were desalted using Sep-Pak C18 cartridges (Waters). Subsequently samples were dried in vacuo and stored at -80 °C until further use.

Phosphorylated peptides were enriched using Fe(III)-NTA 5 µL (Agilent technologies) in an automated fashion using the AssayMAP Bravo Platform (Agilent Technologies) as previously described by Post *et al*. [2]. For all samples, 250 µg of peptides were used as input for one cartridge. In brief, Fe(III)-NTA cartridges were primed with 200 µL of 0.1% TFA in ACN and equilibrated with 250 µL of loading buffer (80% ACN/0.1% TFA). Samples were dissolved in 200 µL of loading buffer containing 100 fmol synthetic isotope labelled t-loop standard peptides and loaded onto the cartridge at a loading speed of 5 µL/min. Subsequently columns were washed with 250 µL loading buffer and eluted with 35 µL of 10 % ammonia directly into 35 µL of 10% formic acid. Samples were dried down and stored at −80 °C until LC–MS analysis.

LC-MS/MS setup

Spectral libraries were partly acquired on a TripleTOF 5600 (Sciex) coupled to an Agilent 1290 Infinity System (Agilent Technologies) adapted to nanoflow conditions by using a split flow setup as described in [3]. The system was operated with in-house packed trap column (Dr. Maisch Reprosil C18, 3 µm, 2 cm × 100 µm) and analytical column (Agilent Poroshell 120 EC-C18, 2.7 µm, 50 cm × 75 µm). The split flow was adapted to achieve 300 nl/min flow at the front end of the column upon applying a flow rate of 0.2 mL/min. 0.6 % acetic acid in water (Milli-Q, Millipore) was used as buffer A an 0.6 % acetic acid, 80 % ACN was used as buffer B. Upon injection, peptides were trapped at 5 µL/min during 5 min with 100 % solvent A (0.1 % FA in water) before being separated on the analytical column.

All remaining measurements were executed on a TSQ-Vantage (Thermo Fisher) coupled to an Easy-nLC 1000 (Proxeon, Odense, DK). LC configuration was in one-column setup (25 cm, 75 µm ID PepMap RLSC, C18, 100 Å, 2 µm particle size column (Thermo Scientific, Odense, DK)). Formic acid (0.1%, Merck, Darmstadt, Germany) in deionized water (Biosolve, Valkenswaard, NL, ULC/MS grade) was used as solvent A, 0.1% formic acid in acetonitrile (Biosolve, Valkenswaard, NL, ULC/MS grade) as solvent B. All measurements were performed at 200 nl/min flow rate and all samples were analyzed with injection volumes of 2 µL containing 10 % FA.

Spectral library generation

SRM assay development was guided by a spectral library generated in house. This provided confirmation of the ordered peptide sequence and essential information about LC and MS characteristics of each peptide. For this spectral library generation, the heavy labeled phosphopeptides were mixed with 1 x HRM Calibration KIT and analyzed in data-dependent mode on two different LC-MS setups. (1) Crude peptides were analyzed on a 5600 TripleTOF (Sciex). LC-MS setup and data acquisition methods was used as previously described [4]. In brief, peptides were separated on a 2h gradient analyzed in TOP20 mode (selection criteria: intensity >50 cps, charge state ≥ 2+, dynamic exclusion 15 s). (2) Crude peptides were analyzed on a TSQ Vantage (Thermo) in data-dependent acquisition mode operated at TOP2. Survey scans were acquired in Q3MS mode (1.5 s, 0.4 Da fwhm) spanning the 375-1350 m/z range. The TOP2 most abundant ions were analyzed in MS/MS mode (selection criteria: intensity > 1000 counts, 5 repeats at 3 s, 2 min dynamic exclusion). Results of both acquisition types were subjected to database search using Mascot accessed by Proteome Discoverer (version 1.4). Parameters were set to tryptic digest, allowing for up to three missed cleavages, using carbamidomethyl cysteine as fixed modification and allowing for serine/threonine/tyrosine phosphorylation methionine oxidation and C-terminal isotope labels. Precursor mass tolerance and MS/MS tolerance were set to 50 ppm and 0.15 Da respectively for TripleTOF files and to 0.9 Da for TSQ files.

Results were filtered using Percolator [5] to an FDR below 1 %. Spectral libraries were built in Skyline from both searches.

SRM assay development

All assays were developed and optimized on a TSQ Vantage as previously described [6, 7]. In brief, the most intense fragment ions found in the spectral libraries were directly used as initial transitions, multiplexing up to 3-10 transitions per precursor. Those initial SRM assays were applied to the synthetic peptide library enabling subsequent optimization of multiple parameters such as collision energy optimization and RT scheduling. Initial assays for a few peptides not identified in either of the spectral library were constructed from theoretically possible y- and b-ions in combination with most likely precursor charge states. Extensive manual validation of phosphosites localization isomers was performed, including the use of site determining ions as transitions. Collision energies were optimized for each transition individually in an empirical way assisted by Skyline [8]. Instrument specific CE parameters were used as a starting point (CE = 0.03m/z + 2.905 for doubly charged precursors and CE = 0.038m/z + 2.281 for precursor charges of three and higher) to scan through different normalized collision energy values using a step size of one.

t-loop detection in various sample types

Dried samples were reconstituted in 3 µL of 10 % formic acid, containing 0.1x/µL iRT peptides. Injection volumes for all analyses were kept at 2 µL. LC-MS analysis contained the following steps: Analytical column equilibration (3µL 100% Buffer A at 600 bars) followed by sample loading onto the column (loading volume 6 µL at 600 bars). Phosphopeptides were separated on a gradient from 2 % to 25 % B in 105 minutes, followed by a column washing step ramping up from 25% to 100% B in 5 minutes followed by 100% B for 15 minutes. To avoid carryover and monitor LC performance at least 1 BSA run was scheduled after each analysis. Samples within one experiment were analyzed in randomized order. Retention time scheduling was dynamically adapted to reflect prior to sample analysis. Low confidence survey runs applied low transition numbers per peptide (~3) in short scheduling windows (4 min) applying long cycle times (4 sec). High confidence quantification runs applied less target peptides with a higher number of transitions per peptide (up to 7, including phosphosites localization specific ions), longer scheduling windows (6-10 min) and shorter cycle times (2.5-3 sec).

Data assessment, quantification and significance analysis

All SRM experiments were analyzed using Skyline [9]. Signal quality was assessed visually, primarily relying on sequence similarity between heavy labeled standard peptides and endogenous peptides. Key points used were perfect co-elution of both peptide forms in terms of retention time and peak shape in combination with a high similarity of the relative

intensities of transitions found in the heavy and the endogenous peptides (rdotp > 0.9). Ratios between analyte and internal standard were used as quantitative readout. Unless otherwise specified significance analysis was performed with MSStats [10]. The analysis entailed log_2-transformation of the intensity values, followed by testing for abundance differences between different conditions using a linear-mixed effects model. An FDR cutoff of ≤ 5% was considered significant.

Supporting Table S1: Compendium of tryptic peptides representing t-loop phosphorylation included in our assay.

Peptide Sequence	Kinase (Phosphorylation Site)
APEMVNLYS*GK	AAK1 (S235)
APEMVNLY*SGK	AAK1 (Y234)
APEMVNLY*S*GK	AAK1 (Y234/S235)
LMTGDT*YTAHAGAK	Abl (T392), Abl2/Arg (T438)
LMTGDTYT*AHAGAK	Abl (T394) Abl2/Arg (T440)
LMTGDTY*TAHAGAK	Abl (Y393) Abl2/Arg (Y439)
IGDFGLAT*VK	Araf (T454) Braf (T599) C-Raf/Raf1 (T491)
IIDSEY*TAQEGAK	BLK (Y389)
PGEEDNAAISEVGT*IR	BMPR2 (T379)
YVLDDQY*VSSVGTK	BMX (Y566)
YVLDDEY*TSSVGSK	BTK (Y551)
IADFGLS*K	CaMK4 (S189) RIPK2 (S168)
T*VCGTPGYCAPEILR	CaMK4 (T200)
QET*VECLK	CaMKIIβ (T287)
SPEVLLGS*AR	CDK1 (S178)
VYT*HEVVTLWYR	CDK1 (T161)
EYGS*PLK	CDK11A (S577) CDK11B (S589)
AY*TPVVVTQWYR	CDK11A (Y582)
AY*T*PVVVTQWYR	CDK11A (Y582/T583)
AYT*PVVVTLWYR	CDK11B (T595)
AY*TPVVVTLWYR	CDK11B (Y594)
AY*T*PVVVTLWYR	CDK11B (Y594/T595)
QS*GVVVEEPPPSK	CDK12 (S1053)
NSS*PAPPQPAPGK	CDK12 (S1083)
TYT*HEVVTLWYR	CDK2 (T160) CDK3 (T160)
SFGS*PNR	CDK7 (S164)
AYT*HQVVTR	CDK7 (T170)
AFS*LAK	CDK9 (S175)
ILGETS*LMR	Chk2/Rad53 (S379)

Supporting Table S1: Compendium of tryptic peptides representing t-loop phosphorylation included in our assay. (continued)

Peptide Sequence	Kinase (Phosphorylation Site)
VVDFGSATFDHEHHSTIVST*R	CLK2 (T344)
DIMNDSNY*IVK	CSF1R (Y809)
NLY*AGDYYR	DDR1 (Y792)
NLYAGDY*YR	DDR1 (Y796)
NLYAGDYY*R	DDR1 (Y797)
NLY*SGDYYR	DDR2 (Y736)
NLYSGDY*YR	DDR2 (Y740)
NLYSGDYY*R	DDR2 (Y741)
LADFGS*CLK	DMPK (S216) MRCKA (S222) MRCKB (S221)
IVDFGS*SCQLGQR	Dyr1A (S310) Dyr1B (S262)
IY*QY*IQSR	Dyr1A (Y319/Y321) Dyr1B (Y271/Y273)
IYQY*IQSR	Dyr1A (Y321) Dyr1B Y273)
VYT*YIQSR	DYRK2 (T381) DYRK4 (T263)
VYT*Y*IQSR	DYRK2 (T381/Y382) DYRK4 (T263/Y264)
VYTY*IQSR	DYRK2 (Y382) DYRK4 (Y264)
VYT*YIQSR	DYRK2 (Y382) DYRK4 (Y264)
LYTY*IQSR	Dyrk3 (Y369)
WTAPEAIS*YR	EphA2 (S790)
WTAPEAIS*Y*R	EphA2 (S790/Y791)
VLEDDPEAT*Y*TTSGGK	EphA2 (T771/Y772)
WTAPEAISY*R	EphA2 (Y791)
VLEDDPEAAY*TTR	EphA3 (Y779) EphA4 (Y779) EphA5 (Y833)
VLEDDPEAAY*TTTGGK	EphA6 (Y830)
VIEDDPEAVY*TTTGGK	EphA7 (Y791)
IADPEHDHT*GFLTEYVATR	ERK1 (T198)
IADPEHDHTGFLT*EYVATR	ERK1 (T202)
IADPEHDHTGFLTEY*VATR	ERK1 (Y204)

Supporting Table S1: Compendium of tryptic peptides representing t-loop phosphorylation included in our assay. (continued)

Peptide Sequence	Kinase (Phosphorylation Site)
VADPDHDHTGFLT*EYVATR	ERK2 (T185)
VADPDHDHTGFLTEYVAT*R	ERK2 (T190)
VADPDHDHTGFLTEY*VATR	ERK2 (Y187)
Y*MEDSTYYK	FAK (Y570)
YMEDSTY*YK	FAK (Y576)
YMEDSTYY*K	FAK (Y577)
QEDGGVY*SSSGLK	Fer (Y714)
EEADGVY*AASGGLR	FES (Y713)
DIHHIDY*YK	FGFR1 (Y653)
DIHHIDYY*K	FGFR1 (Y654)
DINNIDY*YK	FGFR2 (Y656)
DINNIDY*Y*K	FGFR2 (Y656/Y657)
DINNIDYY*K	FGFR2 (Y657)
DVHNLDY*YK	FGFR3 (Y647)
DVHNLDY*Y*K	FGFR3 (Y647/Y648)
GVHHIDY*YK	FGFR4 (Y642)
GVHHIDY*Y*K	FGFR4 (Y642/Y643)
GVHHIDYY*K	FGFR4 (Y643)
DDEY*NPCQGSK	Fgr (412)
DIMSDSNY*VVR	FLT3 (Y842)
VDNEDIY*ESR	FRK (Y387)
LIEDNEY*TAR	Fyn (Y420) Lck (Y394) Yes (Y426) Src (Y419)
WTAPEAALY*GR	Fyn (Y440) Yes (446) Src (Y439)
VS*ENDFEDLLS*NQGFSSR	GAK (S1176&S1185)
VS*ENDFEDLLSNQGFSSR	GAK (S1176)
HPGHYAVYNLS*PR	GAK (S456)
DESEVS*DEGGSPISSEGQEPR	GAK (S829)
DESEVSDEGGS*PISSEGQEPR	GAK (S834)
SDPSGHLT*GMVGTALYVSPEVQGSTK	GCN2 (T899)
SDPSGHLTGMVGT*ALYVSPEVQGSTK	GCN2 (T904)
GEPNVSY*ICSR	GSK3α (Y279) GSK3β (Y216)

Supporting Table S1: Compendium of tryptic peptides representing t-loop phosphorylation included in our assay. (continued)

Peptide Sequence	Kinase (Phosphorylation Site)
VIEDNEY*TAR	HCK (Y411) Lyn (Y397)
SEIGHSPPPAY*TPMSGNQFVYR	Her4 (Y1056)
TVCSTY*LQSR	HIPK3 (Y359)
PPYT*DYVSTR	ICK (T157) MAK (T157)
PPYT*DY*VSTR	ICK (T157/Y159) MAK (T157/Y159)
PPYTDY*VSTR	ICK (Y159) MAK (Y159)
DIY*ETDYYR	IGF1R (Y1161) InsR (Y1185)
DIYETDY*YR	IGF1R (Y1165) InsR (Y1189)
DIYETDYY*R	IGF1R (Y1166) InsR (Y1190)
FVS*VYGTEEYLHPDMYER	IKKE (S172)
GT*LAYLPEEYIK	IRAK1 (T387)
FAQTVMTS*R	IRAK4 (S346)
FAQTVMT*SR	IRAK4 (T345)
DVYETDY*YR	IRR (Y1145)
DVYETDY*Y*R	IRR (Y1145/Y1146)
DVYETDYY*R	IRR (Y1146)
FVLDDQY*TSSTGTK	ITK (Y512)
EY*YTVK	JAK1 (Y1034)
EY*Y*TVK	JAK1 (Y1034/Y1035)
EYY*TVK	JAK1 (Y1035)
DY*YVVR	JAK3 (Y980) PKR2 (Y113) PKR1 (Y122)
DY*Y*VVR	JAK3 (Y980/Y981) PKR2 (Y113/Y114) PKR1 (Y122/Y123)
DYY*VVR	JAK3 (Y981) PKR2 (Y114) PKR1 (Y123)
TAGTSFMMT*PYVVTR	JNK1 (T183) JNK3 (T221)

Supporting Table S1: Compendium of tryptic peptides representing t-loop phosphorylation included in our assay. (continued)

Peptide Sequence	Kinase (Phosphorylation Site)
TAGTSFMMT*PY*VVTR	JNK1 (T183/Y185) JNK3 (T221/Y223)
TAGTSFMMTPY*VVTR	JNK1 (Y185) JNK3 (Y223)
TACTNFMMT*PYVVTR	JNK2 (T183)
TACTNFMMT*PY*VVTR	JNK2 (T183/Y185)
TACTNFMMTPY*VVTR	JNK2 (Y185)
GDVMST*ACGTPGYVAPEVLAQK	CaMK1δ (T180)
GAILT*TMLATR	CaMK2α (T305) CaMK2β (T306) CaMK2δ (T306)
GAILT*T*MLATR	CaMK2α (T305/T306) CaMK2β (T306/T307) CaMK2δ (T306/T307)
GAILTT*MLATR	CaMK2α (T306) CaMK2β (T307) CaMK2δ T307)
NDS*NYVVK	KIT (S821)
NDS*NY*VVK	KIT (S821/Y823)
NDSNY*VVK	KIT (Y823)
ENIFGES*R	PKCδ (S503)
AENGLLMT*PCYTANFVAPEVLK	RSK1 (T573) RSK2 (T577)
MS*AAGTYAWMAPEVIR	MLK2 (S262) MLK1 (S308)
MSAAGT*YAWMAPEVIR	MLK2 (T266) MLK1 (T312)
GS*AAWMAPEVFEGSNYSEK	TAK1 (S192)
GT*EIYMSPEVILCR	COT (T290)
MS*TAGTYAWMAPEVIK	MLK4 (S303)
IDQGDLMT*PQFTPYYVAPQVLEAQR	MK5 (T182)
LDTFCGS*PPYAAPELFQGK	MARK1 (S219) MARK2 (S212) MARK3 (S215) MARK4 (S218)
IADFGFSNEFT*VGNK	MARK1 (T208)

Supporting Table S1: Compendium of tryptic peptides representing t-loop phosphorylation included in our assay. (continued)

Peptide Sequence	Kinase (Phosphorylation Site)
LDT*FCGSPPYAAPELFQGK	MARK1 (T215) MARK2 (T208) MARK3 (T211) MARK4 (T214)
LDT*FCGS*PPYAAPELFQGK	MARK1 (T215/S219) MARK2 (T208/S212) MARK3 (T211/S215) MARK4 (T214/S218)
CLTSNLLQS*R	MASTL (S293)
DYLSSS*FLCSDDDR	MASTL (S552)
DYLSSS*FLCS*DDDR	MASTL (S552/S556)
GVENPAVQES*NQK	MASTL (S631)
S*FNSHINASNNSEPSR	MASTL (S657)
S*FNSHINASNNS*EPSR	MASTL (S657/S668)
SFNSHINASNNS*EPSR	MASTL (S668)
ILGT*PDYLAPELLLGR	MASTL (T741)
LCDFGVSGQLIDS*MANSFVGTR	MEK1 (S218) MEK2 (S222)
T*YVGTNAYMAPER	MEK5 (T315)
IY*SGDYYR	Mer (Y749) Tyro3/Sky (Y681)
IYSGDY*YR	Mer (Y753) Tyro3/Sky (Y685)
IYSGDYY*R	Mer (Y754) Tyro3/Sky (Y686)
EY*YSVHNK	Met (Y1234)
EY*Y*SVHNK	Met (Y1234/Y1235)
EYY*SVHNK	Met (Y1235)
GYLS*EGLVTK	ERK4 (S186)
GHLS*EGLVTK	ERK3 (S189)
HTDDEMT*GYVATR	p38α (T180)
HTDDEMT*GY*VATR	p38α (T180/Y182)
HTDDEMTGY*VATR	p38α (Y182)
LCDFGISGQLVDS*IAK	MKK4 (S2570
T*TQMSAAGTYAWMAPEVIK	MLK3 (T277)
TQTS*MSLGTTR	MLKL (S358)
TQTSMS*LGTTR	MLKL (S360)
ICDFGAS*R	MLTK (S155)

Supporting Table S1: Compendium of tryptic peptides representing t-loop phosphorylation included in our assay. (continued)

Peptide Sequence	Kinase (Phosphorylation Site)
ATDS*FSGR	MNK2 (S74)
CTIS*YR	MPSK1 (S197)
CTIS*Y*R	MPSK1 (S197/Y198)
QALT*LQDWAAQR	MPSK1 (T185)
CTISY*R	MPSK1 (Y198)
LADFGS*CLR	MRCKG (S216)
AYS*FCGTIEYMAPDIVR	MSK1 (S212)
NT*FVGTPFWMAPEVIK	MST3 (T190) YSK1 (T174)
NIYSADY*YK	MuSK (Y755)
ILNHDTS*FAK	NEK2 (S171)
ILNHDT*SFAK	NEK2 (T170)
ILNHDT*S*FAK	NEK2 (T170/S171)
T*FVGTPYYMSPEQMNR	NEK2 (T175)
TFVGT*PYYMSPEQMNR	NEK2 (T179)
FFSSETTAAHS*LVGTPYYMSPER	NEK6 (S206)
FFSSETT*AAHSLVGTPYYMSPER	NEK6 (T202)
HMT*QEVVTQYYR	NLK (T298)
QADEEMT*GYVATR	p38β (T180)
QADEEMT*GY*VATR	p38β (T180/Y182)
QADEEMTGY*VATR	p38β (Y182)
QADSEMT*GYVVTR	p38γ (T183)
QADSEMT*GY*VVTR	p38γ (T183/Y185)
QADSEMTGY*VVTR	p38γ (Y185)
HADAEMT*GYVVTR	p38δ (T180)
HADAEMT*GY*VVTR	p38δ (T180/Y182)
HADAEMTGY*VVTR	p38δ (Y182)
STMVGT*PYWMAPEVVTR	PAK1 (T427) PAK2 (T406) PAK3 (T440)
S*LVGTPYWMAPELISR	PAK4 (S475)
S*LVGTPYWMAPEVISR	PAK6 (S560) PAK7 (S602)
DSNY*ISK	PDGFRβ (Y857)
ANS*FVGTAQYVSPELLTEK	PDK1 (S241) PDK2 (S241)
DIMHDSNY*VSK	PGFRA (Y849)

Supporting Table S1: Compendium of tryptic peptides representing t-loop phosphorylation included in our assay. (continued)

Peptide Sequence	Kinase (Phosphorylation Site)
TWT*LCGTPEYLAPEIILSK	PKAα (T198) PKAβ (T198) PKAγ (T198)
TWTLCGT*PEYLAPEIILSK	PKAα (T202) PKAβ (T202) PKAγ (T202)
T*FCGTPDYIAPEIIAYQPYGK	PKCα (T497) PKCβ (T500) PKCγ (T514)
TFCGT*PDYIAPEIIAYQPYGK	PKCα (T501) PKCβ (T504) PKCγ (T518)
TNT*FCGTPDYIAPEILLGQK	PKCθ (T538)
S*VVGTPAYLAPEVLR	PKD1/PKCμ (S742) PKD3 (S735)
TST*FCGTPEFLAPEVLTETSYTR	PKN2 (T816)
TSTFCGT*PEFLAPEVLTETSYTR	PKN2 (T820)
TST*FCGTPEFLAPEVLTQEAYTR	PKN3 (T718)
TSTFCGT*PEFLAPEVLTQEAYTR	PKN3 (T722)
T*LCGTPNYIAPEVLSK	PLK1 (T210)
T*LCGT*PNYIAPEVLSK	PLK1 (T210/T214)
TLCGT*PNYIAPEVLSK	PLK1 (T214)
T*ICGTPNYLSPEVLNK	PLK2 (T239)
S*VVGTPAYLAPEVLLNQGYNR	PRKD2 (S710)
SVVGTPAY*LAPEVLLNQGYNR	PRKD2 (Y717)
EDVY*LSHDHNIPYK	PTK6 (Y342)
EDVY*LSHDHNIPY*K	PTK6 (Y342/Y351)
EDVYLSHDHNIPY*K	PTK6 (Y351)
YIEDEDY*YK	PYK2/FAK2 (Y579)
YIEDEDY*Y*K	PYK2/FAK2 (Y579/Y580)
YIEDEDYY*K	PYK2/FAK2 (Y580)
DVY*EEDSYVK	RET (Y900)
DVY*EEDSY*VK	RET (Y900/Y905)
DVYEEDSY*VK	RET (Y905)
IADLGLAS*FK	RIPK1 (S161)
MMSLS*QSR	RIPK2 (S176)
GS*FDGSSSQPSR	RK (S21)

Supporting Table S1: Compendium of tryptic peptides representing t-loop phosphorylation included in our assay. (continued)

Peptide Sequence	Kinase (Phosphorylation Site)
EY*YSVQQHR	RON (Y1238)
EY*Y*SVQQHR	RON (Y1238/Y1239)
EYY*SVQQHR	RON (Y1239)
EIYSADY*YR	ROR1 (Y645)
EVYAADYY*K	ROR2 (Y646)
SGEPLST*WCGSPPYAAPEVFEGK	SIK (T182)
DDIY*SPSSSSK	SRMS (Y849)
ADENY*YK	Syk (Y525)
ADENYY*K	Syk (Y526)
WYAPECINY*YK	Syk (Y546)
YVLDDQY*TSSSGAK	TEC (Y519)
EPLAVVGS*PYWMAPEVLR	TESK1 (S220)
LAVVGS*PFWMAPEVLR	TESK2 (S219)
GEEVY*VK	Tie1 (Y1007) GCN2 (Y70)
GQEVY*VK	TIE2 (Y992)
DIY*STDYYR	TrkA (Y676)
DIYSTDY*YR	TrkA (Y680)
DIYSTDYY*R	TrkA (Y681)
DVY*STDYYR	TrkB (Y702) TrkC (Y705)
DVYSTDY*YR	TrkB (Y706) TrkC (Y709)
DVYSTDYY*R	TrkB (Y707) TrkC (Y710)
VAGS*QPITVAWYK	TTN (S9203)
VAGS*QPIT*VAWYK	TTN (S9203/T9207)
VAGSQPIT*VAWYK	TTN (T9207)
Y*TCQIK	TTN (Y8490) [TBRG1 (Y234)]
YVLDDEY*VSSFGAK	TXK (Y420)
AVPEGHEY*YR	TYK2 (1054)
IYNGDYY*R	UFO (Y703)
GS*PLYMAPEMVCQR	ULK3 (S176)
DPDY*VR	VGFR2 (Y1059) VGFR3 (Y1068)
ALGADDSY*YTAR	ZAP70 (Y492)

Supporting Table S2: SRM parameters as used for survey and quantification mode. Parameters for survey mode were kept constant throughout all experiments, parameters for quantification mode were adapted according to the sample at hand. Examples shown here refer to the proof of principle analysis on Jurkat cells. For all methods however, cycle times were kept at 2.5-3 sec whilst keeping dwell time at ≥ 25 ms.

	Survey mode	Quantification mode
Peptides	*248*	*57*
total Peptides[A]	*308*	*76*
Transitions / peptide	*3-4*	*3-7*
Total Transitions[B]	*1994*	*676*
Scheduling window width	*4 min*	*6 min*
max. number of concurrent transitions	*267*	*106*
Cycle time	*4 sec*	*3 sec*
min. dwell time	*~15 ms*	*~28 ms*

[A] *includes various Met oxidation forms and iRT standards*

[B] *includes heavy isotope labelled and endogenous version*

Supporting Table S3: List of the detected t-loop phosphorylations across various cell lines.

Peptide Sequence measured	Kinase - t-loop phosphorylation site	Activity detected in:
IGDFGLAT*VK	ARAF - T454 / BRAF - T599 / RAF1 - T491	Platelets
YVLDDEY*TSSVGSK	BTK - Y551	Platelets
GDVMST*ACGTPGYVAPEVLAQK	CaMK1D - T180	Jurkat, PC9
GAILT*TMLATR	CaMK2α - T305 / CaMK2β - T306 / CaMK2δ - T306	Platelets
GAILTT*MLATR	CaMK2α - T306 / CaMK2β - T307 / CaMK2δ - T307	Platelets
IADFGLS*K	CAMKIV - S189 / RIPK2 - S168	PC9, M026, Platelets
T*VCGTPGYCAPEILR	CAMKIV - T200	Jurkat, M026
SPEVLLGS*AR	CDK1 - S178	Jurkat, PC9, M026
VYT*HEVVTLWYR	CDK1 - T161	Jurkat, HEK, PC9, M026
EYGS*PLK	CDK11A - S577 / CDK11B - S589	Jurkat, HEK, PC9, M026
AYT*PVVVTLWYR	CDK11B - T595	Jurkat, HEK, PC9, M026
NSS*PAPPQPAPGK	CDK12 - S1053	Jurkat, HEK, PC9, M026
TYT*HEVVTLWYR	CDK2 - T160	Jurkat, HEK, PC9, M026
SFGS*PNR	CDK7 - S164	Jurkat, HEK, PC9, M026
AYT*HQVVTR	CDK7 - T169	Jurkat, HEK, PC9, M026
AFS*LAK	CDK9 - S175	PC9, M026
ILGETS*LMR	CHK2 - s379	Jurkat, HEK, PC9, M026
IYQY*IQSR	DYRK1A - Y321	Jurkat, HEK, PC9, M026, Platelets
VYTY*IQSR	DYRK2 - Y382	Jurkat, HEK, PC9, M026
LYTY*IQSR	DYRK3 - Y369	Jurkat
WTAPEAIS*YR	EPHA2 - S790	M026
VLEDDPEAAY*TTR	EPHA3 - Y779 / EPHA4 - Y779 / EPHA5 - Y833	PC9
IADPEHDHTGFLTEY*VATR	ERK1 - Y204	Jurkat, PC9, M026, Platelets
VADPDHDHTGFLTEY*VATR	ERK2 - Y187	Jurkat, PC9, M026, Platelets
GHLS*EGLVTK	ERK3 - S189	Jurkat, HEK, PC9, M026
GYLS*EGLVTK	ERK4 - S186	HEK, M026R
Y*MEDSTYYK	FAK - Y570	HEK, PC9, M026
YIEDEDY*Y*K	FAK2 - Y579&Y580	Platelets
QEDGGVY*SSSGLK	FER - Y714	Platelets

Supporting Table S3: List of the detected t-loop phosphorylations across various cell lines. (continued)

Peptide Sequence measured	Kinase - t-loop phosphorylation site	Activity detected in:
LIEDNEY*TAR	FYN - Y420 / LCK - Y394 / YES1 - Y426 / SRC - Y419	Jurkat, HEK, PC9, M026, Platelets
WTAPEAALY*GR	FYN - Y440 / YES1 - Y446/ SRC - Y439	Jurkat, Platelets
GEPNVSY*ICSR	GSK3α - Y279 / GSK3β - Y216	Jurkat, HEK, PC9, M026, Platelets
VIEDNEY*TAR	HCK - Y411 / LYN - Y397	HEK, PC9, Platelets
TVCSTY*LQSR	HIPK3 - Y359	Jurkat, HEK, PC9, M026
FAQTVMTS*R	IRAK4 - S346	Platelets
FAQTVMT*SR	IRAK4 - T345	Platelets
TAGTSFMMT*PYVVTR	JNK1 - T183/ JNK3 - T221	PC9
TAGTSFMMTPY*VVTR	JNK1 - Y185 / JNK3 - Y224	PC9
TACTNFMMTPY*VVTR	JNK2 - Y185	Platelets
LDT*FCGSPPYAAPELFQGK	Mark1 - T215 / Mark2 - T208 / Mark3 - T211 / Mark4 - T214	Jurkat, HEK, PC9, M026, Platelets
EY*YSVHNK	MET - Y1234	PC9
EY*Y*SVHNK	MET - Y1234&Y1235	PC9
IDQGDLMT*PQFTPYYVAPQVLEAQR	MK5 - T182	Jurkat, PC9, M026
LCDFGISGQLVDS*IAK	MKK4 - S257	PC9, M026, Platelets
ATDS*FSGR	MNK2 - S74	HEK, PC9
NT*FVGTPFWMAPEVIK	MST3 - T190 / YSK1 - T174	PC9, M026
HMT*QEVVTQYYR	NLK - T298	Jurkat, M026, Platelets
S*VVGTPAYLAPEVLLNQGYNR	nPKC-D2 - S710	Jurkat, HEK, PC9, M026
S*VVGTPAYLAPEVLR	nPKCμ - S742 / nPKCv - S735	Jurkat, HEK, PC9, M026
HTDDEMTGY*VATR	P38A - Y181	Jurkat, HEK, PC9, M026, Platelets
HTDDEMT*GY*VATR	P38A - Y181&T179	Jurkat
S*LVGTPYWMAPELISR	PAK4 - S475	Jurkat, HEK, PC9, M026, Platelets
S*LVGTPYWMAPEVISR	PAK6 - S560 / PAK7 - S602	PC9
ANS*FVGTAQYVSPELLTEK	PDK1 - S241 / (PDK2 - S241)	Jurkat, HEK, PC9, M026, Platelets
TWT*LCGTPEYLAPEIILSK	PKA-Cα - T198[a]	Jurkat, PC9, M026
T*FCGTPDYIAPEIIAYQPYGK	PKCα - T497 / PKCβ - T500 / PKCγ - T514	Jurkat, PC9, M026, Platelets
ENIFGES*R	PKCδ - S503	Jurkat, HEK, PC9, M026, Platelets

Supporting Table S3: List of the detected t-loop phosphorylations across various cell lines. (continued)

Peptide Sequence measured	Kinase - t-loop phosphorylation site	Activity detected in:
TNT*FCGTPDYIAPEILLGQK	PKCθ - T538	Jurkat, PC9, Platelets
TSTFCGT*PEFLAPEVLTETSYTR	PKN2 - T816[a]	PC9, M026
TLCGT*PNYIAPEVLSK	PLK1 - T214	Jurkat, PC9
IADLGLAS*FK	RIPK1 - S161	Jurkat[b], Platelets
MMSLS*QSR	RIPK2 - S176	PC9, Platelets
AENGLLMT*PCYTANFVAPEVLK	RSK1 - T573 / RSK2 - T577	PC9, M026
YVLDDQY*TSSSGAK	TEC - Y519	Platelets
ALGADDSY*YTAR	ZAP70 - Y492	Jurkat

[a] unambiguous site localization not possible

[b] only detected upon TNFα treatment

A - List of phosphoisomers included in the assays

ADENY*YK	DVY*STDYYR	HADAEMT*GYVVTR	PPYT*DYVSTR	TST*FCGTPEFLAPEVLTETSYTR
ADENYY*K	DVYSTDY*YR	HADAEMTGY*VVTR	PPYTDY*VSTR	TSTFCGT*PEFLAPEVLTETSYTR
APEMVNLY*SGK	DVYSTDYY*R	HTDDEMT*GYVATR	QADEEMT*GYVATR	TST*FCGTPEFLAPEVLTQEAYTR
APEMVNLYS*GK	DY*YVVR	HTDDEMTGY*VATR	QADEEMTGY*VATR	TSTFCGT*PEFLAPEVLTQEAYTR
AY*TPVVVTLWYR	DYY*VVR	ILNHDT*SFAK	QADSEMT*GYVVTR	TWT*LCGTPEYLAPEILSK
AYT*PVVVTLWYR	EDVY*LSHDHNIPYK	ILNHDTS*FAK	QADSEMTGY*VVTR	TWTLCGT*PEYLAPEILSK
CTIS*YR	EDVYLSHDHNIPY*K	IY*SGDYYR	SDPSGHLT*GMVGTALYVSPEVQGSTK	VADPDHDHTGFLT*EYVATR
CTISY*R	EY*YSVHNK	IYSGDY*YR	SDPSGHLTGMVGT*ALYVSPEVQGSTK	VADPDHDHTGFLTEY*VATR
DESEVS*DEGGSPISSEGQEPR	EYY*SVHNK	IYSGDYY*R	S*FNSHINASNNSEPSR	VADPDHDHTGFLTEYVAT*R
DESEVSDEGGS*PISSEGQEPR	EY*YSVQQHR	LDT*FCGSPPYAAPELFQGK	SFNSHINASNNS*EPSR	VAGS*QPITVAWYK
DIHHDY*YK	EYY*SVQQHR	LDTFCGS*PPYAAPELFQGK	S*VVGTPAYLAPEVLLNQGYNR	VAGSQPIT*VAWYK
DIHHDYY*YK	EY*YTVK	LMTGDT*YTAHAGAK	SVVGTPAY*LAPEVLLNQGYNR	VYT*YIQSR
DINNIDY*YK	EYY*TVK	LMTGDTY*TAHAGAK	TACTNFMMT*PYVVTR	VYTY*IQSR
DINNIDYY*K	FAQTVMT*SR	LMTGDTYT*AHAGAK	TACTNFMMTPY*VVTR	WTAPEAIS*YR
DIY*ETDYYR	FAQTVMT*SR	MS*AAGTYAWMAPEVIR	TAGTSFMMT*PYVVTR	WTAPEAISY*R
DIYETDY*YR	FFSSETT*AAHSLVGTPYYMSPER	MSAAGT*YAWMAPEVIR	TAGTSFMMTPY*VVTR	YIEDEDY*YK
DIYETDYY*R	FFSSETTAAHS*LVGTPYYMSPER	NDS*NYVVK	T*FCGTPDYIAPEIIAYQPYGK	YIEDEDYY*K
DIY*STDYYR	GAILT*TMLATR	NDSNY*VVK	TFCGT*PDYIAPEIIAYQPYGK	Y*MEDSTYYK
DIYSTDY*YR	GAILTT*MLATR	NLY*AGDYYR	T*FVGTPYYMSPEQMNR	YMEDSTY*YK
DIYSTDYY*R	GVHHIDY*YK	NLYAGDY*YR	TFVGT*PYYMSPEQMNR	YMEDSTYY*K
DVY*EEDSYVK	GVHHIDYY*K	NLYAGDYY*R	T*LCGTPNYIAPEVLSK	
DVYEEDSY*VK	IADPEHDHT*GFLTEYVATR	NLY*SGDYYR	TLCGT*PNYIAPEVLSK	
DVYETDY*YR	IADPEHDHTGFLT*EYVATR	NLYSGDY*YR	TQTS*MSLGTTR	
DVYETDYY*R	IADPEHDHTGFLTEY*VATR	NLYSGDYY*R	TQTSMS*LGTTR	

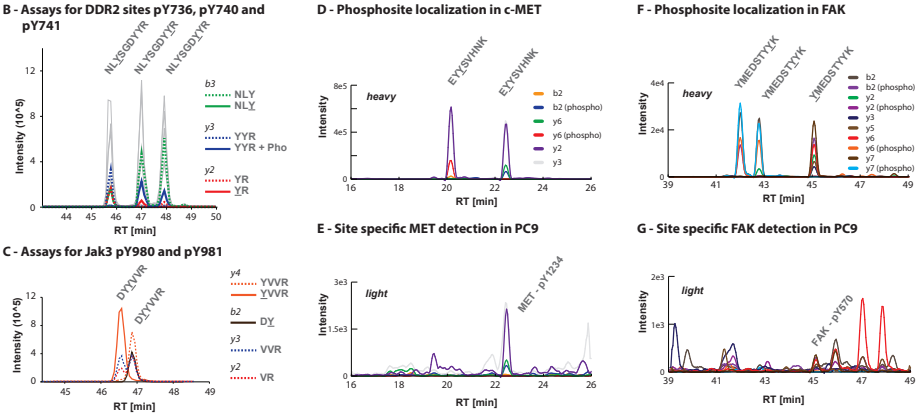


Figure S1: ***Differentiation of phosphoisomers.*** *(A) Eventually our assays comprised a total of 51 peptide sequences present in various positional isomers. 41 of them were present in 2 different isoforms, 10 were present in 3 different isoforms. (B) Key, for the successful differentiation of phosphosites localization isomers are the presence of transitions from site determining fragment ions and high performance online LC separation. Using synthetic heavy isotope labeled peptides, we were able to show baseline chromatographic separation for all positional isomers for 47 out of 51 sequences, illustrated for 3 different phosphosites in the t-loop of DDR2 (Y736, Y740, Y741). (C) Another two peptide sequences showed partial chromatographic separation enabling the clear distinction of different phosphosite localization in complex samples by the use of phosphosite localization specific transitions, illustrated for two different phosphosites in Jak3 (Y980, Y981). (D-G). Unambiguous phosphosite localization in complex sample, illustrated in PC9. (D) SRM signal from heavy peptides spiked into PC9 lysates representing phosphorylation at Y1234 and Y1235 of MET respectively. The two isomers are clearly separated and the presence of site determining fragment ions enables unambiguous phosphosite localization. (E) The presence of site localization specific transitions (e.g. y6 – green) and coelution with the corresponding heavy peptide show that endogenous MET is phosphorylated at Y1234, whereas no signal can be detected for phosphorylation at Y1235. (F-G) Upon high S/N for the measurement of low abundant peptides, co-elution can be used as a stand-alone criterion for phosphosite localization, exemplified with phosphosite Y570 in the Focal Adhesion Kinase 1 (FAK).*

References

[1] van den Eshof, B. L., Hoogendijk, A. J., Simpson, P. J., van Alphen, F. P. J., *et al.*, Paradigm of Biased PAR1 (Protease-Activated Receptor-1) Activation and Inhibition in Endothelial Cells Dissected by Phosphoproteomics. *Arterioscler Thromb Vasc Biol* 2017, *37*, 1891-1902.

[2] Post, H., Penning, R., Fitzpatrick, M. A., Garrigues, L. B., *et al.*, Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* 2017, *16*, 728-737.

[3] Cristobal, A., Hennrich, M. L., Giansanti, P., Goerdayal, S. S., *et al.*, In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. *Analyst* 2012, *137*, 3541-3548.

[4] Schmidlin, T., Garrigues, L., Lane, C. S., Mulder, T. C., *et al.*, Assessment of SRM, MRM(3) , and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer. *Proteomics* 2016, *16*, 2193-2205.

[5] Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 2007, *4*, 923-925.

[6] de Graaf, E. L., Kaplon, J., Mohammed, S., Vereijken, L. A., *et al.*, Signal Transduction Reaction Monitoring Deciphers Site-Specific PI3K-mTOR/MAPK Pathway Dynamics in Oncogene-Induced Senescence. *J Proteome Res* 2015, *14*, 2906-2914.

[7] Soste, M., Hrabakova, R., Wanka, S., Melnik, A., *et al.*, A sentinel protein assay for simultaneously quantifying cellular processes. *Nat Methods* 2014, *11*, 1045-1048.

[8] Maclean, B., Tomazela, D. M., Abbatiello, S. E., Zhang, S., *et al.*, Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal Chem* 2010, *82*, 10116-10124.

[9] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., *et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, *26*, 966-968.

[10] Chang, C. Y., Picotti, P., Hüttenhain, R., Heinzelmann-Schwarz, V., *et al.*, Protein significance analysis in selected reaction monitoring (SRM) measurements. *Mol Cell Proteomics* 2012, *11*, M111.014662.

[11] Marx, H., Lemeer, S., Schliep, J. E., Matheron, L., *et al.*, A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat Biotechnol* 2013, *31*, 557-564.

Chapter 4

Assessment of SRM, MRM^3, and DIA for the Targeted Analysis of Phosphorylation Dynamics in Non-small Cell Lung Cancer

Thierry Schmidlin[1], Luc Garrigues[1], Catherine S. Lane[2], T. Celine Mulder[1], Sander van Doorn[1], Harm Post[1], Erik L. de Graaf[1,3], Simone Lemeer[1], Albert J. R. Heck[1], A. F. Maarten Altelaar[1]

Published in Proteomics 2016, Volume 16, issue 15-16, pages 2193-2205

[1] *Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University and Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, The Netherlands*

[2] *SCIEX, Phoenix House, Lakeside Drive, Centre Park, Warrington WA1 1RX, United Kingdom*

[3] *Present address: Fondazione Pisana per la Scienza ONLUS, Via Panfilo Castaldi 2, 56121 Pisa, Italy*

Abstract

Hypothesis-driven MS-based targeted proteomics has gained great popularity in a relatively short timespan. Next to the widely established selected reaction monitoring (SRM) workflow, data-independent acquisition (DIA), also referred to as sequential window acquisition of all theoretical spectra (SWATH) was introduced as a high-throughput targeted proteomics method. DIA facilitates increased proteome coverage, however, does not yet reach the sensitivity obtained with SRM. Therefore, a well-informed method selection is crucial for designing a successful targeted proteomics experiment. This is especially the case when targeting less conventional peptides such as those that contain PTMs, as these peptides do not always adhere to the optimal fragmentation considerations for targeted assays. Here, we provide insight into the performance of DIA, SRM, and MRM cubed (MRM^3) in the analysis of phosphorylation dynamics throughout the phosphoinositide 3-kinase mechanistic target of rapamycin (PI3K-mTOR) and mitogen-activated protein kinase (MAPK) signaling network. We observe indeed that DIA is less sensitive when compared to SRM, however demonstrates increased flexibility, by postanalysis selection of alternative phosphopeptide precursors. Additionally, we demonstrate the added benefit of MRM^3, allowing the quantification of two poorly accessible phosphosites. In total, targeted proteomics enabled the quantification of 42 PI3K-mTOR and MAPK phosphosites, gaining a so far unachieved in-depth view mTOR signaling events linked to tyrosine kinase inhibitor resistance in non-small cell lung cancer.

1 Introduction

Protein phosphorylation is closely linked to multiple cellular processes such as enzyme activation or inhibition, protein–protein interaction, cell–cell signaling events, and protein degradation. MS has advanced to be the method of choice to study temporal dynamics of protein phosphorylation events at a proteome wide scale [1, 2].

The majority of MS-based phosphoproteomic research entails a relatively unbiased approach by combing phosphopeptide enrichment protocols with shotgun proteomics as reviewed extensively [3-6]. This enables the identification and quantification of phosphorylation events without any prior knowledge and usually achieves a substantially broad phosphoproteome coverage [7-9], although it introduces a bias toward high abundant phosphopeptides [10]. An alternative is the use of MS-based targeted proteomics, which can circumvent this bias and allows for a more hypothesis-driven approach [11]. Conventionally, MS-based targeted proteomics is performed on triple quadrupole mass spectrometers operated in SRM mode [12]. More recently, analogous methods were developed using high resolution instruments such as the Q Exactive (parallel reaction monitoring, PRM) [13-15] or the TripleTOF (HR-MRM, high resolution MRM) [16]. While SRM applies filters for both, precursor and fragment ions, PRM and HR-MRM rely solely on precursor ion isolation in combination with acquisition of high resolution/accurate mass MS/MS spectra. These full-scan fragment spectra are then used to extract traces of specific fragment ions and their chromatographic peak areas are used for quantification [17].

While SRM is generally regarded as the most sensitive and quantitatively most accurate MS-based method [18], it has drawbacks in terms of a comparably low resolution in both precursor and fragment *m/z* filtering. This lower resolution may cause interferences to occur, leading to potentially falsified quantitative readouts by integrating peak groups that do not originate from the peptide of interest. PRM and HR-MRM can reduce such effects by acquiring high resolution MS/MS. Another way to overcome this issue is by using a technique called MRM cubed (MRM^3), implemented on QTRAP instruments. Here, the third quadrupole is replaced by a linear ion-trap, enabling an existing SRM assay to be further refined by isolation and fragmentation of primary fragment ions. This enables monitoring of chromatographic traces originating from secondary fragment ions. While mainly used to increase the specificity of SRM assays so far [19-21], it has also been shown to increase the sensitivity of the assay [22-25].

Another more recent approach in MS-based targeted proteomics is referred to as data-independent acquisition (DIA) or sequential window acquisition of all theoretical spectra [26]. Compared to SRM the acquisition method itself is not targeted per se, but repeatedly cycles through broad consecutive isolation windows, thus fragmenting the whole MS1 range within the chromatographic time scale [27]. Based on prior information obtained from spectral libraries, peak groups for each peptide of interest can be extracted

from the highly multiplexed DIA data and quantified by using the sum of the integrated chromatographic fragment peak areas, similar to SRM. As data acquisition in DIA provides continuous fragment ion intensity information across time, it outperforms SRM in multiplexing capability, thus competing with shotgun proteomics in terms of proteome coverage. Moreover, DIA outperforms SRM in ease of use, as no time consuming method development for each peptide of interest is required [28]. However, it does not achieve the sensitivity of an optimized SRM assay yet [27].

So far MS-based targeted proteomics methods were mainly used to obtain quantitative information on protein expression level, with only limited studies focusing on quantifying specific protein PTMs. However, large-scale phosphoproteomics studies often lack the ability to comprehensively map known functional phosphosites throughout selected signaling cascades of interest. Therefore, MS-based targeted phosphoproteomics potentially presents a valuable alternative in the analysis of cellular signaling. An initial study using SRM to quantify phosphopeptides was presented by Unwin *et al*. in 2005 [29] followed by a handful of studies adopting the phospho-SRM technology to measure different target phosphopeptides (e.g. [30-33]). Approaches to quantify phosphorylation sites by targeted data extraction from DIA analyses are also starting to appear [34, 35]. MRM^3 has so far not been used for phosphopeptide quantification, however, MS3 has shown to be beneficial for targeted phosphopeptide analysis in a pseudo-SRM approach, increasing the dynamic quantification range [25].

Here, we evaluated in parallel SRM, DIA, and MRM^3 for the analysis of phosphorylation dynamics across selected nodes of the phosphoinositide 3-kinase mechanistic target of rapamycin (PI3K-mTOR)/mitogen-activated protein kinase (MAPK) signaling pathway. Involved in a variety of vital cellular processes such as cell proliferation, cell growth, and survival [36], the deregulation of PI3K-mTOR and MAPK-related signaling has been linked to several clinically relevant disorders such as Alzheimer's disease and cancer [37, 38]. Thus, a thorough understanding of disease-specific mTOR-related signaling dynamics can provide valuable information in terms of potential drug targets [36]. Previously we combined highly selective phosphopeptide enrichment, based on Ti^{4+}-IMAC [39] with SRM to monitor and quantify crucial phosphosites of the PI3K-mTOR/MAPK pathway upon senescence [31]. The selection of these phosphosites of interest was based on extensive data mining in publicly available shotgun proteomics datasets, hence increasing the overall success rate of phosphopeptide SRM assay development. In the current study, we used these SRM assays, supplemented with MRM^3, to assess their performance compared to targeted data extraction, as used in DIA experiments. As a model system we chose non-small cell lung cancer cell lines (NSCLC), that were shown to rely highly on mTOR signaling (reviewed multiple times, most recently by Yip 2015 [40]). Especially in epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) resistant cell lines, NSCLC were shown to drastically differ in respect to their PI3K-mTOR/MAPK signaling compared

to sensitive cell lines [41]. For this study, we chose to compare two NSCLC lines, PC9 that is sensitive to TKI treatment and H1975, which has T790M-mediated resistance to first-generation TKIs [42]. Combining highly selective phosphopeptide enrichment with SRM, DIA, and MRM^3 enabled us to determine the specific performance characteristics for these three MS acquisition modes as well as to gain insight into the mTOR signaling dynamics of different NSCLC lines.

2 Materials and methods

2.1 Cell culture

PC9 and H1975 NSCLC cells were purchased from Sigma-Aldrich and the American Type Culture Collection, respectively. PC9 cells contain a deletion (DelE746A750), where H1975 cells contain both the L858R and T780M mutations in the EGFR. Both cell lines were cultured in triplicates in standard Roswell Park Memorial Institute medium 1640 medium (Lonza), containing 10% FBS (Thermo), 2 mM L-glutamine and 1% penicillin/streptomycin (Lonza), at 37°C in a humidified atmosphere containing 5% CO2.

Cells were detached from the culture surface using trypsin (Lonza), and washed three times with PBS before lysis. Drug response of both cell lines was tested as described in the Supporting Information.

2.2 Sample preparation and phosphopeptide enrichment

Frozen cell pellets were subjected to standard phosphoproteomic sample preparation as described before [31]. In brief the protocol combined cell lysis by sonication in urea-based lysis buffer, protein reduction (DTT), and alkylation (iodoacetamide) followed by Lys-C/Trypsin double digestion and desalting. Prior to phosphopeptide enrichment samples subjected to SRM measurements were split in two series and complemented with heavy isotope-labeled standard peptides (JPT) at concentrations of 150 fmol/mg lysate and 3 pmol/mg lysate, respectively. For each mass spectrometric measurement 300 µg lysate were enriched for phosphopeptides using Ti^{4+}-IMAC columns. For each biological replicate two phosphopeptide enrichment replicas per measurement methods were prepared. Further details about sample preparation are provided in the Supporting Information, including a detailed scheme about biological and technical replicas prepared for each acquisition method depicted in Supporting Information Figure 1.

2.3 SRM measurements

Unless otherwise stated, all SRM experiments were conducted on a TSQ-Vantage (Thermo Fisher). Chromatographic separation was performed using an EASY-spray system containing an Easy-nLC 1000 coupled to a 25 cm, 75 µm ID PepMap RLSC, C18, 100 Å, 2 µm particle

size column (Thermo Scientific, Odense, DK). SRM assays were adapted and extended from de Graaf *et al*. [31] including values for optimized collision energy. Phosphopeptides were separated on a gradient from 0 to 25% B in 170 minutes. The expected retention time of each peptide was determined from heavy isotope-labeled standard peptides analyzed in multiple unscheduled measurements. Endogenous phosphopeptides were measured in scheduled acquisition mode (10 min RT window, 4 s cycle time, 971 transitions in total reaching a maximum number of 173 concurrent transitions at RT = 65.90–65.98 min). Resolution was set to 0.7 Da peak width (fwhm) for both Q1 and Q3. Collision gas pressure was set to a constant value of 1.5 mTorr. Further Details about LC-MS/MS setup are provided in the Supporting Information.

2.4 MRM^3 measurements

All MRM^3 measurements were performed on an ekspert nano-LC 425 (Eksigent) coupled to a QTRAP 6500 (SCIEX). LC separation was carried out using a nanoAcquity (75 µm x 25 cm, 1.8 mm, HSS T3) column with a nanoAcquity (180 µm x 20 mm, 5µm Symmetry C18) trap column (Waters) in trap-elute configuration. Detailed information about instrument setup and acquisition methods are provided in the Supporting Information.

2.5 High pH-fractionation

A total of 4 mg cell digest mixed from both cell lines was fractionated on a high-pH (HpH) reversed-phase C18 column (Gemini 3u C18 110 Å, 100 × 1.0 mm, Phenomenex) coupled to an Agilent 1100 series (Agilent Technologies) on a 60 min gradient. 67 fractions of 1 min each were collected and concatenated into five pools as previously described [43]. These were dried down in vacuo and subjected to phosphopeptide enrichment as described above, loading 1/3 of each fraction onto the Ti^{4+}-IMAC tip.

2.6 DIA library generation

Spectral libraries were acquired on a TripleTOF 5600 mass spectrometer (SCIEX) operated in data-dependent acquisition (DDA) mode. Upfront chromatographic separation was performed using an Agilent 1290 Infinity System (Agilent Technologies), adapted to nanoflow conditions by using a split flow setup as described in [44], coupled to in-house packed trap column (Dr. Maisch Reprosil C18, 3 µm, 2 cm × 100 µm) and analytical column (Agilent Poroshell 120 EC-C18, 2.7 µm, 50 cm × 75 µm) using a 155 min gradient. Further details about LC-MS/MS setup and detailed DDA criteria are provided in the Supporting Information.

2.7 DIA measurements

DIA measurements were acquired on a TripleTOF 5600 using the same instrument setup and gradients as described above. We used 64 variable DIA windows, each with 1 amu overlap (Supporting Information Figure 2) spanning an *m/z* range of 400–1250. The collision energy

for each window was determined based on the collision energy for a 2+ ion with a collision energy spread of 0 eV. An accumulation time of 50 ms was used for each fragment ion scan which in combination with a 100 ms survey scan resulted in a total cycle time of 3.4 s.

2.8 Data analysis

All SRM experiments were analyzed using Skyline [45]. To ensure reliable assessment of the SRM traces, signals were validated using mProphet [46], as implemented in the advanced peak picking option of Skyline. Decoy sequences were created by sequence shuffling and adding a precursor mass shift of 10 Da and measured in the same retention time window as their respective target peptide. The mProphet scoring model for SRM data was optimized according to an initial training sample and used without further optimization on all samples using a *q*-value cutoff of 0.01 (corresponding to a 1% FDR).

For DIA targeted data extraction, DDA measurements were searched using MASCOT [47] and combined to a spectral library in Skyline as described in more detail in the Supporting Information. Subsequent to targeted data extraction of both target and decoy sequences in Skyline, the DIA data were validated using an mProphet scoring model directly optimized on the data. Compared to the model used for SRM it also contained DIA-specific scoring features such as mass errors and precursor isotope dot products. Specific mProphet feature scores and score distributions for both SRM and DIA data can be found in Supporting Information Table 1 and Supporting Information Figure 3.

In both, SRM and DIA a few obvious wrongly assigned peaks were integrated manually, as indicated in Supporting Information Table 2.

Quantification was performed by integrating chromatographic fragment ion peak areas in Skyline. Pearson correlations were calculated and visualized with Perseus using intensity values for all transitions individually. CVs were determined by Skyline. Quantitative values were subsequently submitted to significance analysis using MSStats [48]. For SRM this included transforming ratios between analyte and internal standard to a log2-scale, for DIA global intensity normalization based on equalizing medians was performed, followed by log2-transformation of the intensity values. Peak groups not significantly annotated by mProphet were treated as missing values. A linear-mixed effects model was successively used to test for abundance changes between different cell lines. An FDR cutoff of ≤ 5% was considered significant. A list of the number of transitions used for each peptide for quantification and significance analysis is provided in Supporting Information Table 3.

3 Results and discussion

Here, we set out to assess and compare the performance of SRM, MRM^3, and DIA for MS-based targeted phosphoproteomics (Figure 1). As model system we chose two dif-

ferent NSCLC lines, namely H1975 and PC9 that differ substantially in sensitivity to EGFR tyrosine kinase inhibitors. In these cells, we monitored phosphorylation events involved in PI3K-mTOR/MAPK signaling. Both cell lines were grown in triplicates and tested for their sensitivity to Gefitinib and Afatinib. As expected, H1975 showed increased tolerance toward both tyrosine kinase inhibitors (Supporting Information Figure 4). All cells were subjected to alike sample preparation protocols (differing mainly in the amount of internal standards added in SRM) including the workflows used for tryptic digestion and subsequent phosphopeptide enrichment by Ti^{4+}-IMAC. For each biological replicate, two enrichment replicates for each MS method were prepared. A schematic of the sample preparation is depicted in Supporting Information Figure 1. In Supporting Information Figure 5 and 6 we show representative chromatographic elution profiles and information about retention time reproducibility for SRM and DIA, respectively. For SRM and DIA the maximal retention time shift is in the range of a minute within a 3 h gradient. Peak width in the SRM measurements range from 40 to 60 s and in DIA from roughly 30–120 s. Pearson correlations as well as CVs for each peptide are given in Supporting Information Figure 7.

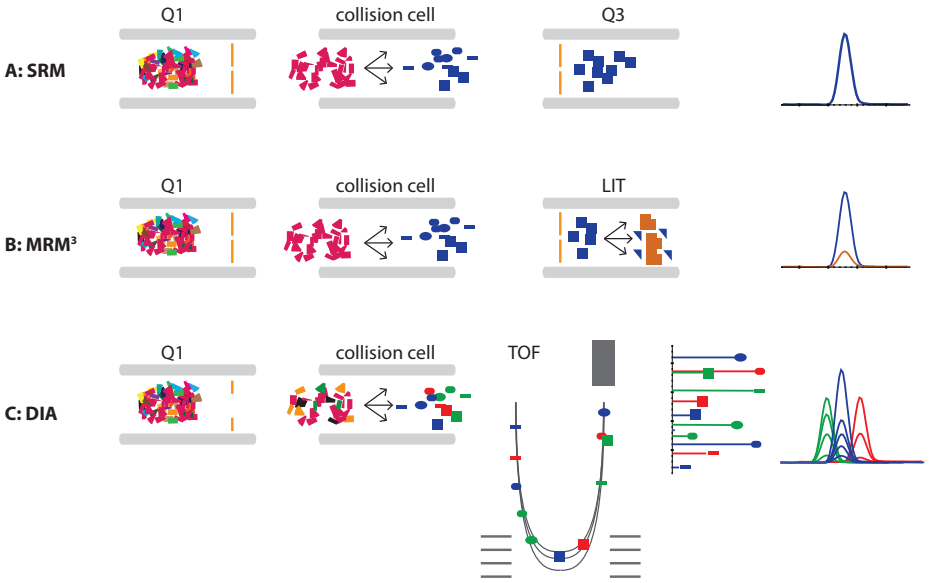


Figure 1: ***Schematic overview of three different MS-based targeted proteomics acquisition types evaluated for their application in phosphoproteomics****. (A) In a triple quadrupole mass spectrometer SRM combines specific sets of m/z filters for Q1 and Q3, representing highly specific precursor and fragment ions pairs, so called transitions. Chromatographic traces of these transitions are then recorded for a predefined retention time window. (B) Compared to SRM, MRM3 adds an additional fragmentation and filtering event monitoring also chromatographic traces of secondary fragment ions. (C) DIA combines wide isolation windows for precursor m/z selection with acquisition of high-resolution fragmentation spectra. The resulting highly multiplexed fragmentation spectra are subsequently analyzed by targeted data extraction based on prior established spectral libraries.*

3.1 Signal transduction monitoring by SRM

One of the major strengths of MS-based targeted proteomics is the capability to monitor specific molecular events of interest. In the case of phosphoproteomics, many global studies have been performed that give rise to large numbers of identified phosphosites [8, 9, 49, 50]. These studies give a global picture of phosphoproteome regulation, however, often lack the ability to comprehensively map all important phosphosites with reported function in signaling cascades or biological processes, caused by undersampling or lack of sensitivity. Here, we selected specific phosphorylation sites known to play important roles in the PI3K/mTOR and MAPK pathways as well as sites of so far unknown function, quantified in our earlier study [31] (Supporting Information Table 2). Sites of known function included activity inducing sites for the kinases AMPK-B1 (serine 108 [51]), PDK1 (serine 241 [52]), ERK1/2 (tyrosine 204 and 187, respectively [53]), p70S6K (serine 447 [54]), BRAF (serine 446 and serine 729 [55, 56]) and mTOR (serine 1291 [57]) as well as kinase activity inhibiting phosphorylation sites for CDK1 (threonine 14 [58]), c-RAF (threonine 259 [59]), and IRS1 (serine 1101 [60]). Additionally several phosphosites were monitored, which are not located on kinases but have established roles in signal transmission. These sites included phosphorylation of serine 65 of 4EBP1, triggering the release of EIF4B to an unbound state and thus activating translation and protein synthesis [61], two phosphorylation sites of Akt1S1 that promote its dissociation from mTORC1 and consequently repress its inhibitory function on mTORC1 kinase activity [62-64], RAPTOR serine 863 which, upon phosphorylation induces mTORC1 activity [65] and PIK3C2A serine 259, where phosphorylation promotes its own degradation [66].

To accurately quantify the selected phosphosites, heavy isotope-labeled standard peptides were added after cell lysis and tryptic digestion in two different concentrations (150 fmol/mg proteins and 3 pmol/mg proteins), following a workflow presented earlier by de Graaf *et al*. [31], after which cell lysates were subjected to Ti^{4+}-IMAC enrichment in duplicate for each of the two heavy isotope-labeled standard peptide concentrations. In total 48 phosphopeptides of the PI3K-mTOR/MAPK pathways were targeted by SRM. Based on mProphet scoring and manual data assessment, 40 of these targeted peptides could be significantly (mProphet q-value ≤ 0.01) detected in at least three out of six injections per cell line (Figure 2).

An exceptional case is the phosphorylation of serine 415 of RSK2, which according to mProphet, has been confidently identified in all three biological replicas (including both enrichment replicas for each) of H1975 but only in one enrichment replicate of PC9. Manual assessment of the data revealed consistent presence of the heavy isotope-labeled standard peptides across all samples. Peak groups for the endogenous peptide were detected in all replicas of H1975, whereas they hardly exceeded the noise threshold in PC9, thus suggesting a potential on/off situation for this particular phosphosite (Supporting Information Figure 8).

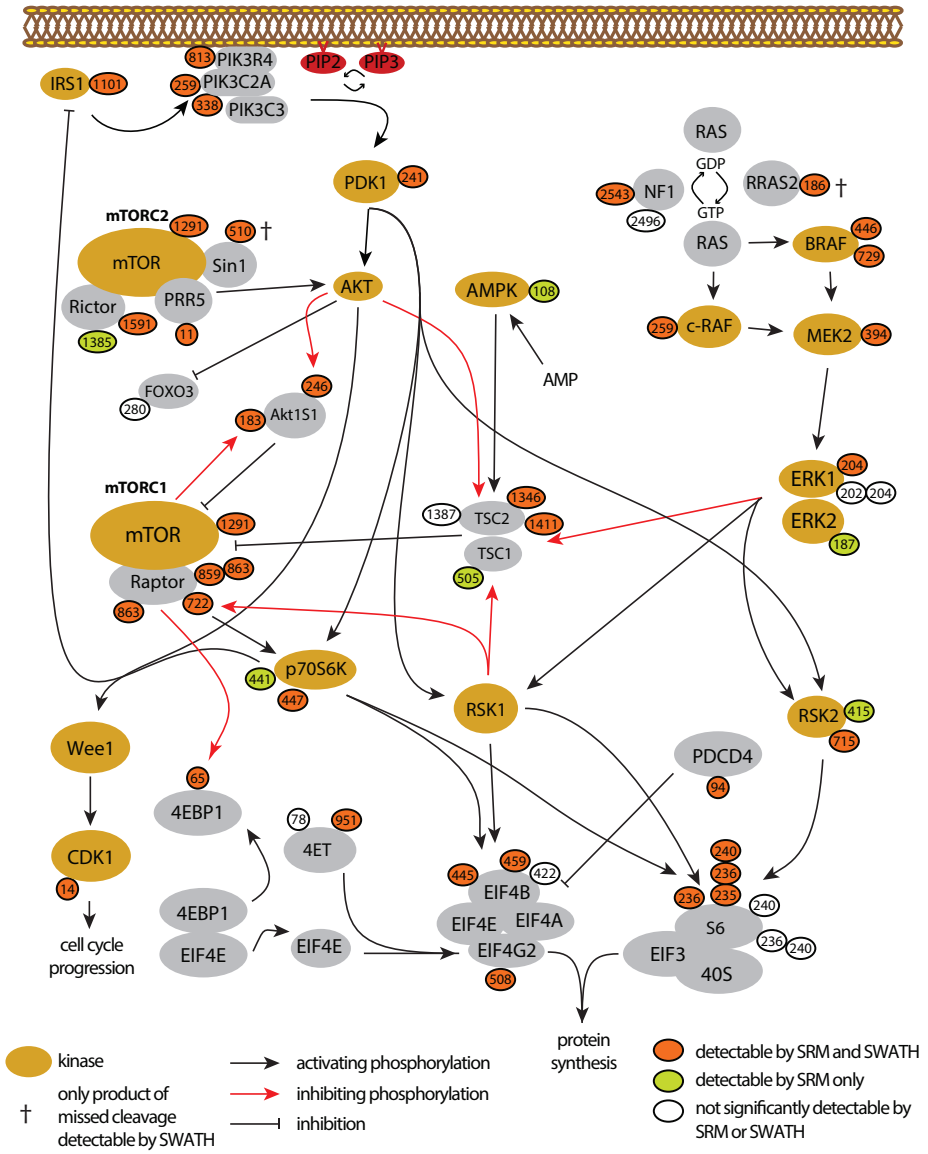


Figure 2: ***Overview of the detectability of phosphosites in the PI3K-mTOR/MAPK pathway in non-small cell lung cancer cell lines****. Thirty-four phosphosites (orange) were significantly detected with SRM as well as DIA. For two phosphosites (Sin1 S510 and RRAS2 S186 (†)) only an alternative peptide containing a missed cleavage was detectable in DIA. Six additional phosphosites (light green) could be confidently detected by SRM but not by DIA. In case of phosphorylation site S415 in RSK2 detection was only successful in the H1975 cells.*

3.2 MRM cubed

To increase our signaling pathway coverage we explored the use of MRM^3 for targeted quantification of phosphosites. In a previous study, Bauer *et al*. [25] showed increased

sensitivity by performing MS3 on the neutral loss ion of low abundant phosphopeptides in a pseudo SRM approach, in the linear iontrap of an LTQ-Orbitrap. As a first proof of principle we selected two phosphopeptides from an existing dataset that performed poorly under "normal" optimized SRM conditions and developed MRM^3 assays. One phosphopeptide (IS-(pho)S-PTETER) was selected based on an occurring interference and one (ALVHQL-(pho)S-NESR) based on its low S/N level. As can be seen in Supporting Information Figure 9, for both these phosphopeptides MRM^3 was able to drastically reduce interferences and improve the S/N levels. Based on these promising results we selected three specific phosphosites from the mTOR/PI3K pathway, with again either high interference or very low S/N ratio, to be targeted by MRM^3, specifically PIK3R4 (S813), FOXO3 (S280), and TSC2 (S1387), with the latter two not accessible in our SRM experiments. The MRM^3 experiments require MS instrumentation with trapping capabilities and were performed on a QTRAP 6500. Existing SRM assays were copied and reoptimized on the QTRAP, where the different LC-MS setup caused slightly different results, already in SRM mode. Still the selected sites performed poorly, as can be seen from TSC2 (S1387) in Figure 3A. Likewise the other two sites showed very low S/N (Supporting Information Figure 10). Next, these selected peptides were analyzed in MRM^3 mode, optimizing crucial instrument parameters such as secondary ion selection, linear ion-trap fill time and excitation energy (Supporting Information Table 4). With this approach we were able to substantially increase the assay quality for those three phosphopeptides in terms of interference and sensitivity (Figure 3B), expanding the coverage of nodes in the signal transduction cascade to 42 out of 48 targeted phosphopeptides. These results demonstrate the great potential of MRM^3 analysis of (low abundant) phosphosites. However, due to the increased measurement time required (acquisition times of up to 1 s for MRM^3), and the (so far) limited software support for scheduled MRM^3 methods, comprehensive pathway analysis by MRM^3 is not yet feasible. Yet, for most of the phosphosites investigated here this is not required as they perform well under SRM conditions. Thus we concluded that MRM^3 is best used as a complementary method in combination with SRM, in which MRM^3 is used specifically for a limited number of individual targets poorly detectable in SRM. A prerequisite for successful MRM^3, however, is a sufficiently efficient secondary peptide fragmentation event (MS3). The fragment ion usually giving rise to the most informative MS3 spectra is the neutral loss of phosphoric acid. However, we found the SRM trace resulting from neutral loss of phosphoric acid to be nonspecific. Accordingly, the gain in specificity associated with the second fragmentation step in MRM^3 was greatly reduced when targeting the neutral loss fragment, providing little increase in sensitivity or specificity over the best performing SRM traces (Supporting Information Figure 11).

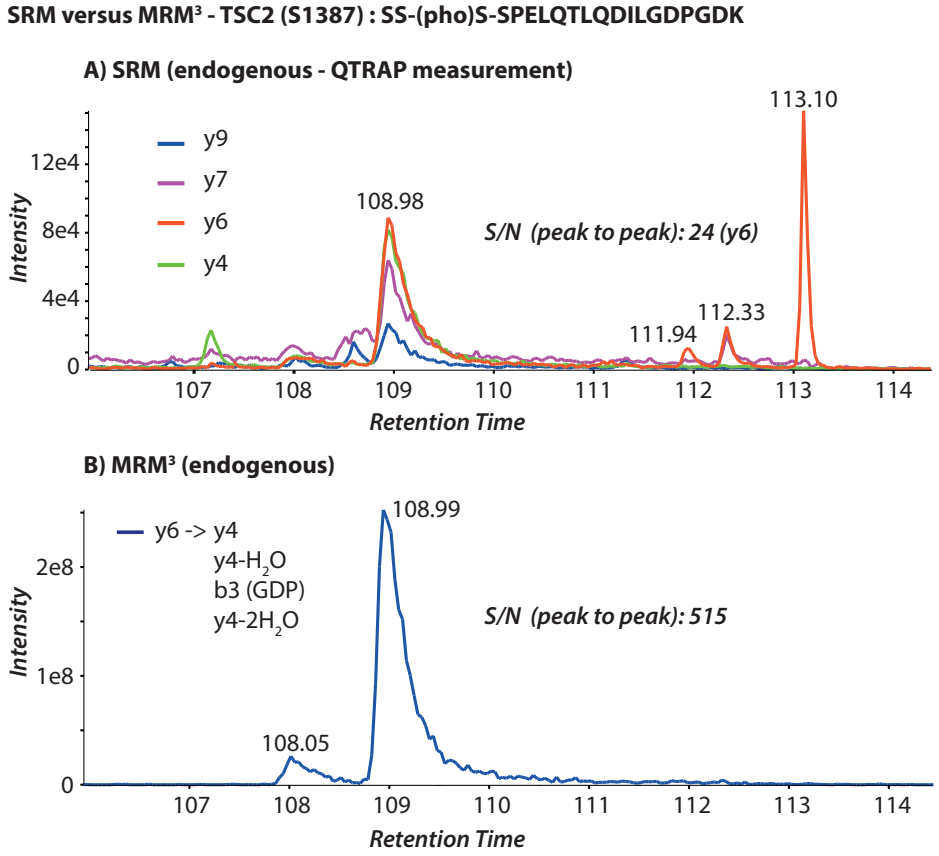


Figure 3: Illustrative chromatographic traces obtained for the phosphopeptide SS-(pho)S-SPELQTLQDILGDPGDK measured in a complex cell lysate using two different measurement methods: (A) SRM and (B) MRM³ were acquired using the same instrument setup (QTRAP 6500). Compared to conventional SRM, MRM³ is capable of reducing unspecific ion signals, as seen by the disappearance of the peak groups at retention times 107.1, 111.9, 112.3, and 113.1 in the MRM³ workflow. For MRM³ longer accumulation times in the trap also enable a drastic increase in sensitivity, improving the S/N by a factor of more than 40.

3.3 Signal transduction monitoring by data-independent acquisition

Next, we assessed the performance of DIA for the targeted analysis of the same 48 phosphopeptides, targeted by combining SRM and MRM³. The acquisition of a reference spectral library is an essential step for the targeted data extraction in DIA measurements. Here, we compared two different ways of acquiring this spectral library (Figure 4A). So far the most promising approaches described in literature include pooling all samples of interest followed by extensive fractionation [67, 68]. In view of these results, we performed high-pH reversed-phase (HpH) fractionation with subsequent phosphopeptide enrichment on a mixed PC9 and H1975 cell lysate. Using a concatenation strategy as demonstrated by Batth *et al*. [43] we identified 12 770 phosphopeptides from 5 HpH fractions. To investigate if the

increased proteomic depth of the library benefits the targeted data extraction we decided to acquire a second spectral library combining one single (unfractionated) DDA analysis of each cell line only, giving rise to a combined library containing 9569 phosphopeptides. Figure 4B shows the overlap between these two spectral libraries, with, as expected, the highest number of unique phosphosites in the HpH-generated library. Subsequently, similar as in the above-described SRM experiments, two Ti^{4+}-IMAC enrichments per biological replicate were prepared and analyzed in DIA mode. Based on the acquired spectral libraries, mProphet scoring and visual assessment of the DIA data, we were able to extract peak groups for 34 of the 48 phosphosites targeted by SRM (Figure 2). Thirty-three out of the 34 peptides detected by DIA could be observed in the smaller spectral library already.

3.4 Comparison between SRM and DIA in terms of sensitivity

The initial DIA study by Gillet *et al*. reported an approximately tenfold lower sensitivity of DIA compared to SRM, based on classical dilution curve experiments [27]. In this study, we show potential implications of this loss in sensitivity when aiming for comprehensive pathway analysis. Six phosphosites within the mTOR signaling network that were well accessible by optimized SRM assays remained inaccessible for DIA analysis, hence limiting information about pathway dynamics. To illustrate this loss in sensitivity, chromatographic traces of TSC1 (S505) measured in SRM and DIA are displayed in Figure 4C. To ensure an unbiased evaluation, SRM and DIA measurements were both performed on PC9 samples containing 50 fmol of the heavy isotope-labeled standard phosphopeptides. The synthetic peptide clearly shows the difference in sensitivity (in S/N) between SRM and DIA for this specific phosphopeptide. For the endogenous phosphopeptide this difference in sensitivity between the two methods leads to a situation in which a clear signal can be obtained with SRM whereas it remains below detection limit for DIA. In two other cases, specific phosphopeptides successfully targeted by SRM were not detectable by DIA, however, the phosphosites could be quantified by extracting peak groups for their corresponding phosphopeptides containing missed cleavages. This again highlights the difference in sensitivity between the two methods, but also demonstrates the benefit of DIA when it comes to target peptide selection. Whereas in DIA it is straightforward to refine the target peptide selection postacquisition this would require de novo assay development and data acquisition in SRM.

3.5 Quantification of PI3K/mTOR and MAPK phosphosites

Both the SRM dataset and the DIA dataset were separately tested to determine significant changes in abundance between the two cell lines (numeric results are provided in Supporting Information Table 5). Figure 5A depicts the results obtained from the SRM experiments. The volcano plot shows several phosphosites exhibiting significant differential abundances between the two investigated cell lines, with those annotated with their protein name

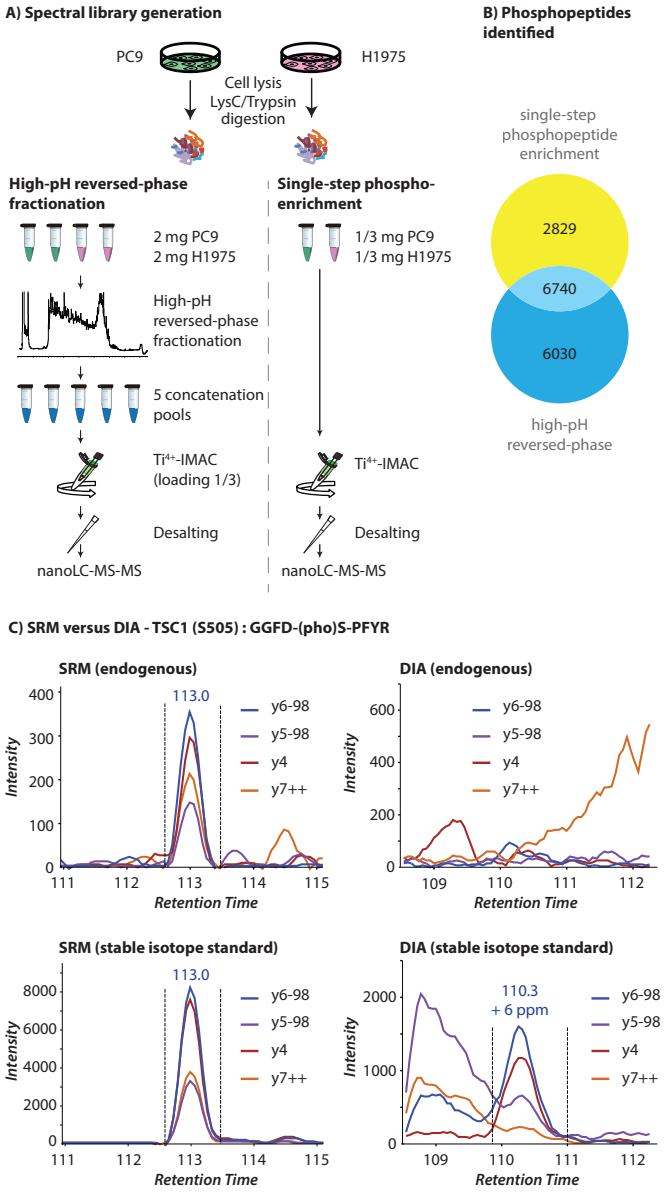


Figure 4: (A) Build-up of spectral libraries for the DIA analysis using two different strategies: Either pooled peptide digest of PC9 and H1975 cell lines were fractionated using high-pH reversed-phase chromatography and concatenated to five fractions followed by phosphopeptide enrichment (left side) or the digest of both cell lines were directly enriched (right side) after which samples were analyzed by DDA. (B) Venn diagram comparing the number of unique phosphopeptides identified in the two spectral libraries. (C) The reduced sensitivity of DIA compared to SRM exemplified by the phosphopeptide GGFD-(pho)S-PFYR. Samples for both measurements contained equal amounts of heavy isotope-labeled standard peptides for this example. Panels show chromatographic traces for SRM (left panels) and DIA (right panels), endogenous peptides (upper panels) and stable isotope standard (lower panels).

uniquely quantified by SRM. Comparison of the quantification results obtained with SRM and DIA shows a reasonably good correlation (R^2 = 0.82, Figure 5B), although lower than reported by Gillet *et al*. (R^2 = 0.95). This is mainly caused by two phosphosites (BRAF) exclusively regulated in the SRM experiment, and four regulated phosphosites exclusive to the DIA experiment. This discrepancy between the previously obtained correlation and the one reported here likely originates from the fact that (1) the previously reported SRM and DIA correlation was obtained at protein level analysis, generally involving more than one peptide per protein, thus strengthening the reliability of the quantification. (2) The quantitative fold changes reported here are relatively small, up to fourfold maximum, whereas Gillet *et al*. reported fold changes of up to 300. Overall the observed differences are small and both SRM and DIA largely agree quantitatively.

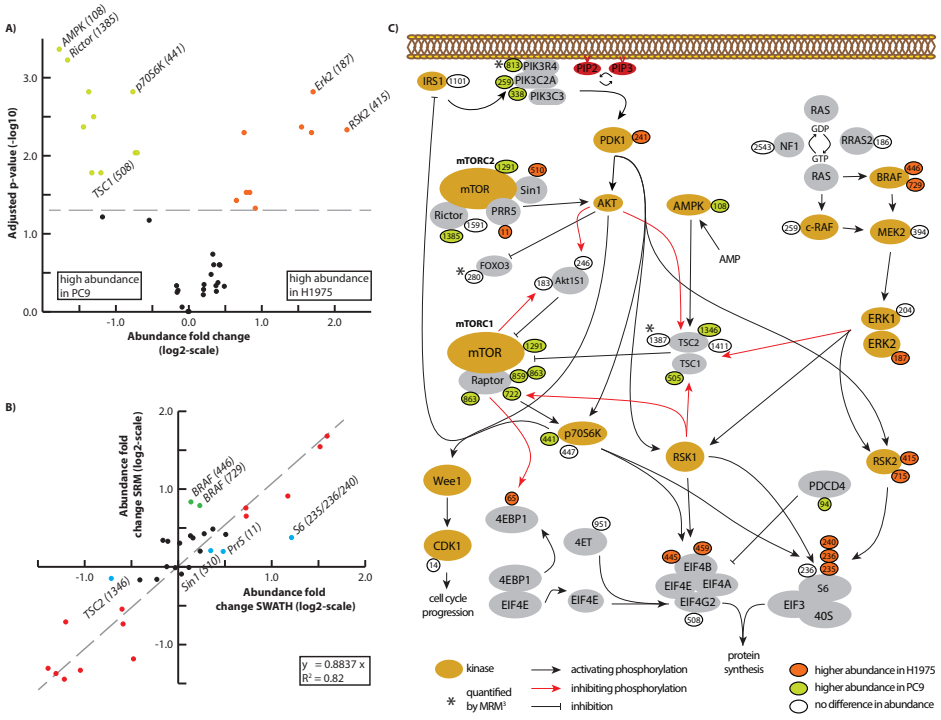


Figure 5: (A) Statistical analysis of the differential abundances of the detected phosphopeptides between H1975 cells and PC9 cells as detected by SRM displayed in a volcano plot. p values (–log10 of adjusted p-value) are plotted as a function of the differential abundance observed using p ≤ 0.05 as significance threshold. Significant changes in abundance are color coded (orange = higher abundance in H1975, light green = higher abundance in PC9). Labeled dots represent phosphosites not accessible by DIA. (B) Correlation between phosphopeptide abundance differences as determined by DIA and SRM. Abundance fold changes are defined as log2 (abundanceH1975/abundancePC9). Values in red represent a significant (p ≤ 0.05) differential abundance observed with both methods, dots marked in green and blue represent significant fold changes exclusively determined by SRM and DIA, respectively. (C) Phosphosite localization within the PI3K/mTOR/MAPK signaling pathway, color coded by abundance changes between H1975 cells and PC9 cells (p ≤ 0.05).

Of note, in some cases multiple potential peak groups show up in both SRM and DIA data (Supporting Information Figure 12). This might be caused by potential different phosphorylation sites on the same peptide, which do share most of the transitions but differ slightly in retention time. For an automated peak picking and scoring program such as mProphet or Spectronaut distinction of such features is highly challenging, demonstrating the importance of manual data assessment in both SRM and DIA, in particular when analyzing phosphopeptides.

Quantification by MRM^3 could confirm the significantly higher abundance of PIK3R4 (S813) in PC9, as previously determined by both SRM and DIA. For the remaining two phosphopeptides targeted by MRM^3 no significant difference in abundance across the two cell lines was observed (Supporting Information Table 6).

3.6 Implications of TKI resistance on mTOR signaling

We quantified 42 phosphopeptides across two different non-small cell lung cancer cell lines using three different MS-based targeted proteomics approaches. Of these monitored phosphosites, 25 showed significant differential regulation between the two cell lines, 12 displaying an increased and 13 a decreased abundance in H1975 compared to PC9 (Figure 5C). Overall the differences in phosphosite abundances observed here suggest a more complex change in mTOR signaling between the two cell lines than a mere up- or down-regulation of the whole pathway. It is striking that many phosphosites in the downstream region of the signaling pathway clearly suggest an elevated rate of protein synthesis in H1975 compared to PC9. This begins with phosphorylation of S65 of 4EBP1 releasing EIF4E to activate translation, followed by an increase in phosphorylation of S445 and S459 on EIF4B. Moreover, the phosphorylation of PDCD4 at serine 94 showed a decreased abundance in H1975 compared to PC9. PDCD4 has been shown to inhibit protein translation but gets rapidly degraded in proliferating cells [69], hence the decrease in abundance of its phosphorylation, caused by protein degradation, corroborating a potential increase in protein synthesis. In line with these observations, we find the abundance of ERK2 S187 and the two phosphosites of RSK2 to be more abundant upon TKI resistance (with RSK S415 exclusively present in H1975), which again is consistent with an increase in protein synthesis. It is however interesting that the entire BRAF/MAPK pathway upstream of ERK2 does not show any difference in abundance between the two cell lines, except for a slight upregulation of BRAF detected in SRM but not confirmed by DIA.

Remarkably, many other phosphosites localized in upstream regions of the PI3K-mTOR pathway show opposing trends, namely higher abundance in the TKI sensitive PC9 cells. These sites include several phosphosites in the mTORC1 pathways as well as the phosphoinositide 3-kinase (PI3K) complex. It has been shown in previous studies performed on tumor biopsies of patients treated with mTORC1 inhibitors, that there is a negative feedback loop between activation of PI3K/mTORC1 and the MAPK pathway [70] suggest-

ing a certain negative cross-control between the two prosurvival pathways. An opposing trend in the two pathways as observed in our data is therefore plausible and might be part of two different prosurvival mechanisms.

In many cases, we observed that multiple phosphosites on the same protein, or in case of mTORC1 and PI3K the same protein complex, show very similar quantitative behavior (summarized in Supporting Information Table 7). This suggests that molecular changes rather happen at the protein expression than protein phosphorylation level. In a study on the effects of TKI resistance on mTOR signaling, Fei *et al*. detected substantial difference in kinase activity between mTORC1 and mTORC2 in different NSCLC lines [41]. Using a combination of pull downs and in vitro kinase assays they observed an increased kinase activity of mTORC2 in TKI resistant cells as well as an increased kinase activity of mTORC1 in TKI sensitive cells, which nicely is in accordance with our present study.

4 Concluding remarks

MS-based targeted proteomics has become a valuable proteomics tool including its current variants SRM, MRM^3, and DIA. In this study, we evaluated the performance of these three popular methods for the targeted analysis of phosphorylation events in specific signaling networks. By monitoring the PI3K-mTOR/MAPK pathway dynamics in different NSCLC cell lines we could pinpoint advantages and disadvantages of all three methods. SRM outperforms DIA in terms of sensitivity resulting in increased pathway coverage by approximately 15%. DIA, however, exceeds SRM in terms of flexibility, as shown here by quantifying alternative tryptic cleavage products, a strategy that would involve tedious de novo assay development in SRM. By quantifying three phosphosites using MRM^3 we show the additional benefit of this feature over a conventional SRM assay in terms of sensitivity and selectivity, increasing the phosphopeptide coverage in our selected pathway even further.

These considerations can be of great value for successful experiment planning, as all three methods can prove optimal depending on the specific question asked and the resources available. In a purely hypothesis-driven approach, SRM (likely in combination with MRM^3) is likely still the most promising approach, although it comes at the cost of time-consuming assay development and the requirement of heavy isotope-labeled standard peptides for each analyte of interest. In exchange all parameters such as collision energy or dwell time can be iteratively optimized for each transition individually, resulting in highly sensitive assays. In DIA internal standards are mostly omitted, mainly because the measurement of a survey scan enables very robust normalization based on total ion signal as implemented in many label-free shotgun quantification approaches. This renders DIA the cheaper method in terms of both cost and time. However, the unspecific isolation

windows and the acquisition of full-scan MS/MS spectra result in way more restrictions in terms of parameter optimization for DIA compared to SRM, which is one of the reasons for the lower sensitivity compared to SRM.

Next to the reduced sensitivity one of the biggest disadvantages of DIA is the requirement of a spectral library, which is time consuming to build. However, the increasing availability of proteomics sequencing data in online repositories might rapidly overcome this disadvantage in the near future. Hence, our observations suggest a great potential for DIA to grow to an "easy to use" alternative to SRM, if the cost of reduced sensitivity does not conflict with the question at hand. Moreover, it remains to be seen whether DIA approaches can outperform "single shot" label-free quantification in terms of quantitative depth and throughput [71].

Using MS-based targeted proteomics on phosphopeptides does pose additional challenges, especially in terms of peak picking, which requires a thorough manual data assessment in addition to automatized data analysis pipelines such as mProphet or Spectronaut. Nonetheless, the three methods described here all exhibit great potential for accurate quantification, showing sufficient reproducibility between different measurement strategies. Thus, altogether our data shows that MS-based targeted proteomics methods have matured, allowing the measurement of even subtle differences in protein phosphorylation.

Acknowledgments

A.F.M.A. is supported by the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (723.012.102). S.L. is supported by the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (723.013.008). This work is part of the project Proteins at Work, a program of the Netherlands Proteomics Centre financed by the Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-scale Research Facilities of the Netherlands (project number 184.032.201). We like to thank Dr. Séverine Clavier for initial help with the culturing of the cells.

Associated Contend

Raw data can be found in http://www.peptideatlas.org/PASS/PASS00770 including:

- Raw files (SRM), wiff files (DIA, MRM^3).
- Skyline daily files for SRM (two different concentrations of heavy isotope-labeled standard peptides) and DIA including mProphet scores.
- Transition lists for SRM and DIA.

- Raw quantitative values as exported from Skyline and subjected to MSStats analysis.

Supporting Information can be found in http://onlinelibrary.wiley.com/doi/10.1002/pmic.201500453/full

References

[1] White, F. M., Quantitative phosphoproteomic analysis of signaling network dynamics. *Curr Opin Biotechnol* 2008, *19*, 404-409.

[2] Altelaar, A. F., Munoz, J., Heck, A. J., Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 2013, *14*, 35-48.

[3] Lemeer, S., Heck, A. J., The phosphoproteomics data explosion. *Curr Opin Chem Biol* 2009, *13*, 414-420.

[4] Macek, B., Mann, M., Olsen, J. V., Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu Rev Pharmacol Toxicol* 2009, *49*, 199-221.

[5] Thingholm, T. E., Jensen, O. N., Larsen, M. R., Analytical strategies for phosphoproteomics. *Proteomics* 2009, *9*, 1451-1468.

[6] Riley, N. M., Coon, J. J., Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal Chem* 2016, *88*, 74-94.

[7] Giansanti, P., Aye, T. T., van den Toorn, H., Peng, M.*, et al.*, An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas. *Cell Rep* 2015.

[8] de Graaf, E. L., Giansanti, P., Altelaar, A. F., Heck, A. J., Single-step enrichment by Ti4+-IMAC and label-free quantitation enables in-depth monitoring of phosphorylation dynamics with high reproducibility and temporal resolution. *Mol Cell Proteomics* 2014, *13*, 2426-2434.

[9] Bodenmiller, B., Wanka, S., Kraft, C., Urban, J.*, et al.*, Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci Signal* 2010, *3*, rs4.

[10] Michalski, A., Cox, J., Mann, M., More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* 2011, *10*, 1785-1793.

[11] Sajic, T., Liu, Y., Aebersold, R., Using data-independent, high-resolution mass spectrometry in protein biomarker research: perspectives and clinical applications. *Proteomics Clin Appl* 2015, *9*, 307-321.

[12] Picotti, P., Rinner, O., Stallmach, R., Dautel, F.*, et al.*, High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods* 2010, *7*, 43-46.

[13] Gallien, S., Duriez, E., Crone, C., Kellmann, M.*, et al.*, Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol Cell Proteomics* 2012, *11*, 1709-1723.

[14] Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., Coon, J. J., Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics* 2012, *11*, 1475-1488.

[15] Gallien, S., Domon, B., Detection and quantification of proteins in clinical samples using high resolution mass spectrometry. *Methods* 2015, *81*, 15-23.

[16] Tong, L., Zhou, X. Y., Jylha, A., Aapola, U.*, et al.*, Quantitation of 47 human tear proteins using high resolution multiple reaction monitoring (HR-MRM) based-mass spectrometry. *J Proteomics* 2015, *115*, 36-48.

[17] Gallien, S., Bourmaud, A., Kim, S. Y., Domon, B., Technical considerations for large-scale parallel reaction monitoring analysis. *J Proteomics* 2014, *100*, 147-159.

[18] Picotti, P., Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* 2012, *9*, 555-566.

[19] Jones, J. W., Pierzchalski, K., Yu, J., Kane, M. A., Use of fast HPLC multiple reaction monitoring cubed for endogenous retinoic acid quantification in complex matrices. *Anal Chem* 2015, *87*, 3222-3230.

[20] Fortin, T., Salvador, A., Charrier, J. P., Lenz, C., *et al.*, Multiple reaction monitoring cubed for protein quantification at the low nanogram/milliliter level in nondepleted human serum. *Anal Chem* 2009, *81*, 9343-9352.

[21] Lakowski, T. M., Szeitz, A., Pak, M. L., Thomas, D., *et al.*, MS³ fragmentation patterns of monomethylarginine species and the quantification of all methylarginine species in yeast using MRM³. *J Proteomics* 2013, *80*, 43-54.

[22] Lemoine, J., Fortin, T., Salvador, A., Jaffuel, A., *et al.*, The current status of clinical proteomics and the use of MRM and MRM(3) for biomarker validation. *Expert Rev Mol Diagn* 2012, *12*, 333-342.

[23] Jaffuel, A., Lemoine, J., Aubert, C., Simon, R., *et al.*, Optimization of liquid chromatography-multiple reaction monitoring cubed mass spectrometry assay for protein quantification: application to aquaporin-2 water channel in human urine. *J Chromatogr A* 2013, *1301*, 122-130.

[24] Jeudy, J., Salvador, A., Simon, R., Jaffuel, A., *et al.*, Overcoming biofluid protein complexity during targeted mass spectrometry detection and quantification of protein biomarkers by MRM cubed (MRM3). *Anal Bioanal Chem* 2014, *406*, 1193-1200.

[25] Bauer, M., Ahrné, E., Baron, A. P., Glatter, T., *et al.*, Evaluation of data-dependent and -independent mass spectrometric workflows for sensitive quantification of proteins and phosphorylation sites. *J Proteome Res* 2014, *13*, 5973-5988.

[26] Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A., Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* 2004, *1*, 39-45.

[27] Gillet, L. C., Navarro, P., Tate, S., Röst, H., *et al.*, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* 2012, *11*, O111.016717.

[28] Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., *et al.*, OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* 2014, *32*, 219-223.

[29] Unwin, R. D., Griffiths, J. R., Leverentz, M. K., Grallert, A., *et al.*, Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol Cell Proteomics* 2005, *4*, 1134-1144.

[30] Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., White, F. M., Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci U S A* 2007, *104*, 5860-5865.

[31] de Graaf, E. L., Kaplon, J., Mohammed, S., Vereijken, L. A., *et al.*, Signal Transduction Reaction Monitoring Deciphers Site-Specific PI3K-mTOR/MAPK Pathway Dynamics in Oncogene-Induced Senescence. *J Proteome Res* 2015, *14*, 2906-2914.

[32] Jin, L. L., Tong, J., Prakash, A., Peterman, S. M., *et al.*, Measurement of protein phosphorylation stoichiometry by selected reaction monitoring mass spectrometry. *J Proteome Res* 2010, *9*, 2752-2761.

[33] Soste, M., Hrabakova, R., Wanka, S., Melnik, A., *et al.*, A sentinel protein assay for simultaneously quantifying cellular processes. *Nat Methods* 2014, *11*, 1045-1048.

[34] Zawadzka, A. M., Schilling, B., Held, J. M., Sahu, A. K., *et al.*, Variation and quantification among a target set of phosphopeptides in human plasma by multiple reaction monitoring and SWATH-MS2 data-independent acquisition. *Electrophoresis* 2014, *35*, 3487-3497.

[35] Parker, B. L., Yang, G., Humphrey, S. J., Chaudhuri, R., *et al.*, Targeted phosphoproteomics of insulin signaling using data-independent acquisition mass spectrometry. *Sci Signal* 2015, *8*, rs6.

[36] Chiarini, F., Evangelisti, C., McCubrey, J. A., Martelli, A. M., Current treatment strategies for inhibiting mTOR in cancer. *Trends Pharmacol Sci* 2015, *36*, 124-135.

[37] Xu, K., Liu, P., Wei, W., mTOR signaling in tumorigenesis. *Biochim Biophys Acta* 2014, *1846*, 638-654.
[38] Liko, D., Hall, M. N., mTOR in health and in sickness. *J Mol Med (Berl)* 2015, *93*, 1061-1073.
[39] Zhou, H., Ye, M., Dong, J., Corradini, E.*, et al.*, Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat Protoc* 2013, *8*, 461-480.
[40] Yip, P. Y., Phosphatidylinositol 3-kinase-AKT-mammalian target of rapamycin (PI3K-Akt-mTOR) signaling pathway in non-small cell lung cancer. *Transl Lung Cancer Res* 2015, *4*, 165-176.
[41] Fei, S. J., Zhang, X. C., Dong, S., Cheng, H.*, et al.*, Targeting mTOR to overcome epidermal growth factor receptor tyrosine kinase inhibitor resistance in non-small cell lung cancer cells. *PLoS One* 2013, *8*, e69104.
[42] Sharma, S. V., Bell, D. W., Settleman, J., Haber, D. A., Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* 2007, *7*, 169-181.
[43] Batth, T. S., Francavilla, C., Olsen, J. V., Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res* 2014, *13*, 6176-6186.
[44] Cristobal, A., Hennrich, M. L., Giansanti, P., Goerdayal, S. S.*, et al.*, In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. *Analyst* 2012, *137*, 3541-3548.
[45] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M.*, et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, *26*, 966-968.
[46] Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R.*, et al.*, mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods* 2011, *8*, 430-435.
[47] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551-3567.
[48] Chang, C. Y., Picotti, P., Hüttenhain, R., Heinzelmann-Schwarz, V.*, et al.*, Protein significance analysis in selected reaction monitoring (SRM) measurements. *Mol Cell Proteomics* 2012, *11*, M111.014662.
[49] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B.*, et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, *127*, 635-648.
[50] Smit, M. A., Maddalo, G., Greig, K., Raaijmakers, L. M.*, et al.*, ROCK1 is a potential combinatorial drug target for BRAF mutant melanoma. *Mol Syst Biol* 2014, *10*, 772.
[51] Sanders, M. J., Ali, Z. S., Hegarty, B. D., Heath, R.*, et al.*, Defining the mechanism of activation of AMP-activated protein kinase by the small molecule A-769662, a member of the thienopyridone family. *J Biol Chem* 2007, *282*, 32539-32548.
[52] Casamayor, A., Morrice, N. A., Alessi, D. R., Phosphorylation of Ser-241 is essential for the activity of 3-phosphoinositide-dependent protein kinase-1: identification of five sites of phosphorylation in vivo. *Biochem J* 1999, *342 (Pt 2)*, 287-292.
[53] Jurek, A., Amagasaki, K., Gembarska, A., Heldin, C. H., Lennartsson, J., Negative and positive regulation of MAPK phosphatase 3 controls platelet-derived growth factor-induced Erk activation. *J Biol Chem* 2009, *284*, 4626-4634.
[54] Lekmine, F., Sassano, A., Uddin, S., Smith, J.*, et al.*, Interferon-gamma engages the p70 S6 kinase to regulate phosphorylation of the 40S S6 ribosomal protein. *Exp Cell Res* 2004, *295*, 173-182.
[55] Tran, N. H., Wu, X., Frost, J. A., B-Raf and Raf-1 are regulated by distinct autoregulatory mechanisms. *J Biol Chem* 2005, *280*, 16244-16253.
[56] MacNicol, M. C., Muslin, A. J., MacNicol, A. M., Disruption of the 14-3-3 binding site within the B-Raf kinase domain uncouples catalytic activity from PC12 cell differentiation. *J Biol Chem* 2000, *275*, 3803-3809.

[57] Acosta-Jaquez, H. A., Keller, J. A., Foster, K. G., Ekim, B*., et al.*, Site-specific mTOR phosphorylation promotes mTORC1-mediated signaling and cell growth. *Mol Cell Biol* 2009, *29*, 4308-4324.

[58] Chow, J. P., Siu, W. Y., Ho, H. T., Ma, K. H*., et al.*, Differential contribution of inhibitory phosphorylation of CDC2 and CDK2 for unperturbed cell cycle control and DNA integrity checkpoints. *J Biol Chem* 2003, *278*, 40815-40828.

[59] Kilili, G. K., Kyriakis, J. M., Mammalian Ste20-like kinase (Mst2) indirectly supports Raf-1/ERK pathway activity via maintenance of protein phosphatase-2A catalytic subunit levels and consequent suppression of inhibitory Raf-1 phosphorylation. *J Biol Chem* 2010, *285*, 15076-15087.

[60] Li, Y., Soos, T. J., Li, X., Wu, J*., et al.*, Protein kinase C Theta inhibits insulin signaling by phosphorylating IRS1 at Ser(1101). *J Biol Chem* 2004, *279*, 45304-45307.

[61] Peter, D., Igreja, C., Weber, R., Wohlbold, L*., et al.*, Molecular architecture of 4E-BP translational inhibitors bound to eIF4E. *Mol Cell* 2015, *57*, 1074-1087.

[62] Vander Haar, E., Lee, S. I., Bandhakavi, S., Griffin, T. J., Kim, D. H., Insulin signalling to mTOR mediated by the Akt/PKB substrate PRAS40. *Nat Cell Biol* 2007, *9*, 316-323.

[63] Wang, L., Harris, T. E., Roth, R. A., Lawrence, J. C., PRAS40 regulates mTORC1 kinase activity by functioning as a direct inhibitor of substrate binding. *J Biol Chem* 2007, *282*, 20036-20044.

[64] Wang, L., Harris, T. E., Lawrence, J. C., Regulation of proline-rich Akt substrate of 40 kDa (PRAS40) function by mammalian target of rapamycin complex 1 (mTORC1)-mediated phosphorylation. *J Biol Chem* 2008, *283*, 15619-15627.

[65] Wu, X. N., Wang, X. K., Wu, S. Q., Lu, J*., et al.*, Phosphorylation of Raptor by p38beta participates in arsenite-induced mammalian target of rapamycin complex 1 (mTORC1) activation. *J Biol Chem* 2011, *286*, 31501-31511.

[66] Didichenko, S. A., Fragoso, C. M., Thelen, M., Mitotic and stress-induced phosphorylation of HsPI3K-C2alpha targets the protein for degradation. *J Biol Chem* 2003, *278*, 26055-26064.

[67] Zi, J., Zhang, S., Zhou, R., Zhou, B*., et al.*, Expansion of the ion library for mining SWATH-MS data through fractionation proteomics. *Anal Chem* 2014, *86*, 7242-7246.

[68] Shao, S., Guo, T., Koh, C. C., Gillessen, S*., et al.*, Minimal sample requirement for highly multiplexed protein quantification in cell lines and tissues by PCT-SWATH mass spectrometry. *Proteomics* 2015.

[69] Dorrello, N. V., Peschiaroli, A., Guardavaccaro, D., Colburn, N. H*., et al.*, S6K1- and betaTRCP-mediated degradation of PDCD4 promotes protein translation and cell growth. *Science* 2006, *314*, 467-471.

[70] Carracedo, A., Ma, L., Teruya-Feldstein, J., Rojo, F*., et al.*, Inhibition of mTORC1 leads to MAPK pathway activation through a PI3K-dependent feedback loop in human cancer. *J Clin Invest* 2008, *118*, 3065-3074.

[71] Sharma, K., Schmitt, S., Bergner, C. G., Tyanova, S*., et al.*, Cell type- and brain region-resolved mouse brain proteome. *Nat Neurosci* 2015, *18*, 1819-1831.

Chapter 5

Analysis of Phosphorylated Peptides by Data-independent Acquisition: Considerations Regarding HCD-, ETD-, and EThcD-Fragmentation, Data Analysis and Phosphosite Localization

Thierry Schmidlin[1], Juan M. Valverde[1], A. F. Maarten Altelaar[1]

[1] *Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University and Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, The Netherlands*

Abstract

In the area of quantitative proteomics the use of data-independent acquisition (DIA) methods is getting increasingly popular and is applied to an ever increasing variety of sample types ranging from peptidomics up to the analysis of post-translational modifications. Several of these sample types have been previously shown to profit from electron driven fragmentation methods such as ETD and EThcD. These fragmentation methods have so far not been implemented in DIA analysis. Here, we show the feasibility of combining DIA with electron-driven fragmentation methods by analyzing HeLa phosphopeptides. We show how the longer reaction time of ETD/EThcD affects overall performance resulting in lower cycle time and show how the use of wider isolation windows can be used to compensate these effects. Additionally we demonstrate how the lower fragmentation efficiency experienced in ETD affects data analysis by mProphet and how custom made integration models can resolve these issues through MS1 and LC data.

1 Introduction

In 2012 Gillet *et al*. introduced a new concept for consistent and accurate proteome analysis by LC-MS termed sequential window acquisition of all theoretical spectra (SWATH) or data-independent acquisition (DIA) [1]. In contrast to the more commonly used data-dependent acquisition (DDA) analysis, DIA consecutively cycles through a series of wide isolation windows within a chromatographic time scale, giving rise to highly multiplexed MS2-maps. These can be queried for fragment ion traces for numerous peptides by using pre-acquired spectral libraries. This strategy results in quantitative consistency and accuracy comparable to those of Selected Reaction Monitoring while providing proteomic depth comparable to DDA LC-MS [2].

Several studies have since then successfully applied the DIA technology for the analysis of various sample types, ranging from patient biopsies [3] and plasma [4] up to the analysis of N-glycosylation [5] and protein phosphorylation [6, 7]. At the same time, extensive method development has been performed, seeking for improvements in terms of specificity and sensitivity. The proposed solutions range from the use of variable Q1 isolation window sizes [8] up to the simultaneous analysis of multiple Q1 isolation windows in combination with subsequent data deconvolution [9].

Surprisingly however, no in-depth investigation of using alternative peptide fragmentation methods such as ETD or EThcD in DIA has been presented thus far. This despite the known benefit of electron driven fragmentation methods in many aspects of proteomics such as peptidomics [10], glycoproteomics [11] and phosphoproteomics [12]. Therefore, the present study sets out to investigate different aspects of DIA in terms of electron driven peptide fragmentation and data analysis. We show the principal feasibility of combining ETD and EThcD with DIA for the analysis of phosphopeptides, but also report the shortcomings in terms of speed and sensitivity when compared to standard HCD fragmentation. Additionally, we carefully investigate the effects of various DIA parameters on the post-acquisition data analysis process, thus demonstrating how these steps can be specifically optimized with respect to the actual data at hand. Lastly, we demonstrate the great potential DIA might eventually have for the analysis of phosphopeptides in terms of phosphosite localization and which steps we believe need to be taken by the DIA community in order to fully exploit this potential.

2 Materials and Methods

2.1 Cell Cultures

HeLa cells were grown in Dulbeco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 10 mM glutamine (all from Lonza, Braine-l'Alleud, Belgium).

Six hours before harvesting, the medium was replaced by fresh medium. Cells were harvested and the cell pellets were immediately washed two times with phosphate-buffered saline buffer (PBS) and stored at −80 °C until further usage.

2.2 Protein Lysis and Digestion

Cells were lysed, reduced, and alkylated in lysis buffer (1% sodium deoxycholate (SDC), 10 mM tris(2-carboxyethyl)phosphinehydrochloride (TCEP)), 40 mM chloroacetamide (CAA), and 100 mM TRIS, pH 8.0 supplemented with phosphotase inhibitor (PhosSTOP, Roche) and protease inhibitor (cOmplete mini EDTA-free, Roche). Cells were heated for 5 min at 95 °C, sonicated with a Bioruptor Plus (Diagenode) for 15 cycles of 30 s and cell debris was removed by centrifugation at 20 000 g for 10 min. A Bradford protein assay was used to quantify protein amount. Prior to digestion, samples were diluted 1:10 with 50 mM ammoniumbicarbonate, pH 8.0. Proteins were digested overnight at 37 °C with trypsin (Sigma-Aldrich) with an enzyme/substrate ratio of 1:50 and lysyl endopeptidase (Wako) with an enzyme/substrate ratio of 1:75. SDC was precipitated with 2% formic acid (FA) and samples were desalted using Sep-Pak C18 cartridges (Waters), dried in vacuo and stored at -80 °C until further use.

2.3 High pH-fractionation

A total of 4mg HeLa cell digest was fractionated on a high-pH (HpH) reversed-phase C18 column (Gemini 3u C18 110 Å, 100 × 1.0 mm, Phenomenex) coupled to an Agilent 1100 series (Agilent Technologies) on a 60 min gradient. 67 fractions of 1 min each were collected and concatenated into five pools as previously described [13].

2.4 Phosphopeptide Enrichment

Phosphopeptide enrichment was performed as previously described [14]. In brief phosphorylated peptides were enriched using Fe(III)-NTA 5 µL (Agilent technologies) in an automated fashion using the AssayMAP Bravo Platform (Agilent Technologies). Fe(III)-NTA cartridges were primed with 200 µL of 0.1% TFA in ACN and equilibrated with 250 µL of loading buffer (80% ACN/0.1% TFA). Samples were dissolved in 200 µL of loading buffer and loaded onto the cartridge at a loading speed of 5 µL/min. Loading amounts per tip were 250 µg for unfractionated HeLa lysate. Each HpH fraction was divided into two samples enriched on one cartridge each. After sample loading columns were washed with 250 µL loading buffer and eluted with 35 µL of 10 % ammonia directly into 35 µL of 10% formic acid. Samples were dried down and stored at −80 °C until LC–MS analysis.

2.5 LC-MS setup

Unless otherwise indicated, all experiments were performed on an Orbitrap Fusion mass spectrometer (Thermo Scientific) coupled to an Agilent 1290 Infinity System (Agilent

Technologies) adapted to nanoflow conditions by using a split flow setup as described in [15]. The system was operated with in-house packed trap column (Dr. Maisch Reprosil C18, 3 µm, 2 cm × 100 µm) and analytical column (Agilent Poroshell 120 EC-C18, 2.7 µm, 50 cm × 75 µm). The split flow was adapted to achieve 300 nl/min flow at the front end of the column upon applying a flow rate of 0.2 mL/min. 0.6 % acetic acid in water (Milli-Q, Millipore) was used as buffer A an 0.6 % acetic acid, 80 % ACN was used as buffer B.

2.6 Data acquisition

All samples were reconstituted in 10 % formic acid in water and injected at a volume of 10 µL. Upon injection peptides were trapped for 5 min at a 5 µL/min. Subsequently peptides were chromatographically separated on a 100 min gradient running from 8 % B to 32 % B followed by column washing (ramping up to 100 % B for 3 min followed 100 % B for 1 min) and equilibration (ramping down to 0 % B for 1 min followed by 0 % B for 10 min). Parameters used for data-dependent acquisition were set as follows: Survey scans were acquired in the Orbitrap at 60K resolution spanning the 375-1500 *m/z* range using an AGC target of 4e5 and a maximum fill time of 50 ms. Data-dependent MS2 scans were acquired for charge states 2-6 using an isolation width of 1.6 Th and a 12 s dynamic exclusion. HCD, EThcD and ETD fragmentations were used respectively. All MS2 scans were recorded in the Orbitrap in centroid mode, scanning the 350-1000 Th range at 30K resolution using an AGC target of 5e4 and a maximum fill time of 35 ms. Parameters used for data-independent acquisition were set as follows: MS scans were acquired in the Orbitrap at 120K resolution covering the 400-1000 *m/z* range. Orbitrap filling was controlled by an AGC target of 2e5 and a maximum injection time 100 ms. DIA scans were recorded using quadrupole isolation of 30 equal sized windows throughout the 400-1000 *m/z* range. All MS2 scans were recorded in the Orbitrap measuring the 350-1000 *m/z* range in profile mode at 30K resolution. Orbitrap fill time was controlled through an AGC target of 5e4 and a maximum fill time of 100 ms. Three different fragmentation modes were used, HCD with stepped collision energy (25% +/- 5), ETD (with supplemental activation) and EThcD (stepped CE 5%).

2.7 Data analysis

Raw files from data-dependent acquisition analyses were searched against a concatenated database consisting of the SwissProt database (version 56.2) and the Biognosys HRM Calibration Kit peptide sequences using Mascot accessed by Proteome Discoverer (version 1.4). Parameters were set to tryptic digest, allowing for up to three missed cleavages, using carbamidomethyl cysteine as fixed modification and allowing for serine/threonine/tyrosine phosphorylation and methionine oxidation. Precursor mass tolerance and MS/MS tolerance were set to 50 ppm and 0.05 Da respectively. Results were filtered using Percolator [16] to an FDR below 1 %. PhosphoRS [17] was used to control phosphosite localization. Search results were subsequently used to create spectral libraries in Skyline

[18]. These spectral libraries were used for subsequent DIA data extraction. Unless otherwise stated we attempted to extract all identified peptides within the DIA *m/z* range from the DIA runs alongside with an equally sized decoy group. MS1 filtering was set to include 3 isotope peaks using a resolving power of 60 000 at 400 *m/z*. MS/MS filtering was used in DIA mode, following multiple isolation modes as specified in detail for each experiment, using a resolving power of 60 000 at 400 *m/z*. mProphet as implemented in the Skyline advanced peak picking option was used for FDR control of the targeted data extraction (q ≤ 0.01 was used as significance threshold throughout the whole study).

3 Results

3.1 Spectral Library Acquisition for Targeted Data Extraction

Here, we set out to investigate the contribution of different fragmentation methods for the analysis of phosphopeptides by DIA. As a model system we chose HeLa cell lysates enriched for phosphopeptides by Fe(III)-IMAC. Key step to any DIA analysis is the acquisition of reference spectral libraries. This usually entails fractionation of the sample to obtain substantial proteomic depth [19, 20]. Here, HeLa lysates were fractionated by high-pH reverse phase (HpH) in combination with a concatenation strategy as described by Batth *et al*. [13]. 5 concatenated fractions were each enriched for phosphopeptides by Fe(III)-IMAC. Each of the 5 fractions was analyzed in DDA mode by HCD, EThcD and ETD separately (Figure 1A). This enabled us to create spectral libraries containing spectra for 6624 (HCD), 5318 (EThcD) and 3758 (ETD) phosphopeptides respectively. Figure 1B depicts the overlap between the 3 spectral libraries.

3.2 Method development for DIA methods using HCD, ETD and EThcD

For DIA analyses, HeLa samples were subjected to phosphopeptide enrichment without prior fractionation and directly analyzed in DIA using a 2h LC separation. 3 different DIA methods were tested, all scanning the 400-1000 Th range in 30 steps of 20 Th, using HCD, EThcD and ETD-fragmentation respectively (Figure 1A). DIA by nature relies on high speed instrumentation capable of providing a good balance between a short cycle time and a good enough dwell time for a high number of isolation windows. Compared to flow-through instruments such as Q-TOFs, iontrap based instruments provide less control over cycle time. This is primarily due to the dynamic regulation of the orbitrap fill time which is controlled by an AGC target and a maximum fill time. Additionally, electron based fragmentation methods are slower than HCD and can thus substantially contribute to an increased cycle time. Due to these effects we set out to determine the actual cycle times of all three methods empirically. We observed a cycle time of roughly 3 s for HCD fragmentation and 6 s for ETD and EThcD fragmentation.

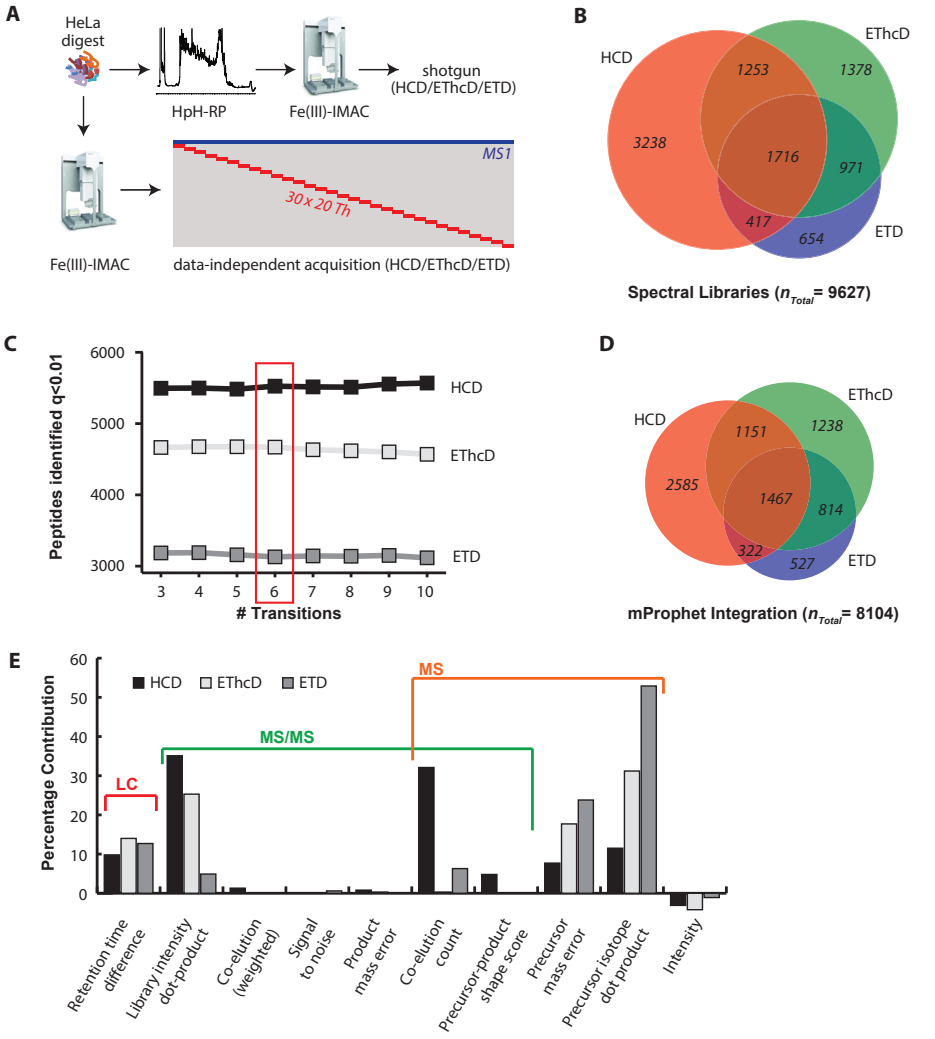


Figure 1: ***Comparison of HCD-DIA, EThcD-DIA and ETD-DIA.*** *(A) Workflow used for the analysis of three different fragmentation methods for the analysis of phosphopeptides by DIA. Spectral libraries were acquired by subjecting a tryptic digest of HeLa cells to high-pH fractionation collecting 5 concatenated fractions subsequently enriched for phosphopeptides by Fe(III)-IMAC. Each fraction was analyzed three times by shotgun LC-MS using HCD, ETD and EThcD respectively (upper panel). For DIA analysis no fractionation was performed prior to phosphopeptide enrichment. DIA methods were acquired using 30 x 20 Th isolation windows and analyzed three times using HCD, EThcD and ETD respectively (lower panel). (B) Number of phosphopeptides identified in the HCD-, EThcD-, and ETD-library respectively, depicted as Venn diagram. (C) Optimization of the data analysis for HCD-DIA, EThcD-DIA and ETD-DIA in terms of transition numbers used for targeted data extraction. Red bar indicates the number of transitions used for subsequent analyses (6). (D) DIA identification (mProphet q-value ≤ 0.01) for each of the three fragmentation methods depicted as Venn diagram. (E) Contribution of multiple mProphet sub-scores to the composite score used for FDR-controlled re-integration.*

For the initial data analysis this difference in cycle time has not been taken into account. Instead the effect of the 3 different fragmentation methods on the subsequent DIA-data analysis has been analyzed in more detail. As the analysis was not performed in replicas, the data presented below should not be considered in terms of actual numbers and statistical significance. It is more our intention to show general trends and effects observed in DIA data analysis that might be taken into consideration when planning DIA experiments.

3.3 Optimizing Data Analysis by Applying mProphet

In a first step we performed targeted data extraction for all three DIA analyses using their respective spectral library as a reference. Through simultaneous extraction of decoy sequences from the data, this step could subsequently be assessed in an error controlled manner. The method of choice for this control step is mProphet, a software package that scores multiple data quality criteria, provides an FDR control and can be optimized in a sample specific way by semi-supervised learning [21]. Initially we set out to determine the optimal amount of transitions to be extracted from the MS2-map for each of the 3 methods individually. Therefore, all peptides present in each library together with an equal amount of decoy sequences were extracted from each DIA run using the 3 most abundant isotopic peaks of the precursor ions (M, M+1, M+2) in combination with 3 to 10 fragment ions. For each amount of transitions an individual mProphet model was created and extracted peak groups were reintegrated using a $q \leq 0.01$ (1 % FDR) cutoff. Surprisingly no drastic changes were observed upon increasing the number of transitions in all 3 methods (Figure 1C). Therefore, all subsequent analyses were performed using a standard number of 6 transitions unless otherwise described. The number of peptides significantly identified ($q \leq 0.01$) in each DIA run are depicted in Figure 1D, as well as their overlap. Interestingly, EThcD-DIA shows the highest recovery of library peptides (4670 of 5318 = 87 %) whereas in case of both ETD-DIA and HCD-DIA the recovery is in the range of 83 %. In total numbers however, HCD offers the most identification (5525 phosphopeptides in 2 h). It is however noteworthy that also in DIA the use of alternative fragmentation methods can give rise to largely different subsets of peptides identified (Figure 1D).

To obtain a better understanding of the effects of the various MS parameters on DIA data and their analysis, we decided to take a deeper look into the actual process of mProphet scoring for HCD-DIA, ETD-DIA and EThcD-DIA. The actual criteria scored for each sub-score can be grouped into LC-dependent (RT-deviation), MS1-dependend (precursor mass error, precursor isotope dotP, etc.), and MS2-dependent (library dotP, product mass error, etc.) as well as combinations thereof. Through semi-supervised learning these sub-scores are weighted differently according to the actual quality of the data at hand. When comparing the 3 fragmentation methods it is striking that the contribution of the MS2-dependent sub-scores, especially the library dotP, are high in HCD-DIA, lower in EThcD-DIA and drop substantially in ETD-DIA (Figure 1E). In case of ETD this is likely caused by low fragment ion

intensity, as evident by the presence of many remaining precursor ions in the MS2 spectra (Illustrated in Supporting Information Figure S1). To compensate for this, sub-scores relying on MS1 data and LC quality however increase dramatically. Similar trends could be observed for all 3 methods upon optimizing transition numbers. Extracting less transitions from the MS2-map results in higher contribution of MS1- and LC-based sub-scores (data not shown). Hence it seems like the occurrence of less efficient fragmentation does not *per se* hamper DIA analysis – although the reliance on MS1 data and LC accuracy automatically increases.

3.4 Applying MS1 quantification in DIA

This observation primed us to investigate the effect of completely missing MS2-data on the analysis of DIA. A few studies in recent years have exclusively employed MS1 data for quantification in DIA. This strategy is based on using spectral libraries to extract MS1 peak traces from MS1-only analyses assisted by high-precision RT analysis [22]. In a first step we assessed how a further reduction of transitions affects DIA data analysis. Figure 2A depicts how the overall identification levels in DIA behave upon reducing the number of transitions stepwise to zero. Interestingly the complete omission of fragment spectra exclusively affects the analysis of HCD-DIA resulting in roughly 10 % less identifications. Identification rates for EThcD-DIA and ETD-DIA however stay stable.

This surprising result prompted us to further investigate this effect. We hence built a combined spectral library comprising identifications from all 3 initial libraries and performed an MS1-only extraction on all 3 DIA runs. A total of 5100 peptides were successfully identified in this analysis, 4724 of them in all three runs (92.6 % overlap - Figure 2B). Figure 2C depicts in which spectral library these 4724 peptides were initially identified. Surprisingly the largest part originates from the EThcD library, implying that simple HCD based DIA analysis misses out on a large subset of phosphopeptides. Potential applications of this approach could be the use of decision-tree based library acquisition in combination with HCD only DIA analysis.

However, the lack of MS2 information in the targeted data extraction is expected to largely compromise confidence in both, peptide identification and quantification. We tested this effect by performing a negative control experiment by analyzing a tryptic digest of *E coli* by DIA in triplicates. For the subsequent data analysis, the same mProphet model was used as for the MS1-based analysis of the HeLa phosphopeptides. Figure 2E-G depicts the score distributions of target and decoy peptides resulting from this analysis. The data analysis of 3 HeLa samples results in a clearly separated score distribution of target and decoy peptides (Figure 2E). However, by including the 3 *E coli* measurements into the data analysis the score distribution of the target peptides changes greatly. The number of target peptides scored in the same score range as the decoy peptides increases, implying that the MS1 traces of the human phosphopeptides extracted from the *E coli* samples score largely

similar to the decoys (Figure 2F). The analysis of the 3 *E coli* datasets alone results in an almost perfect overlap of the score distribution for target and decoy peptides (Figure 2G). By using a cutoff of q ≤ 0.01, mProphet reports only 13-15 significant identification of human phosphopeptides in each *E coli* run. As they are known not be present in the analyzed sample, they can be used as a readout for false discoveries. In comparison to the 5100 peptides identified from the HeLa samples, this corresponds to a FDR of roughly 0.2-0.3 %.

This dataset suggests that in the particular case of our experiments MS1-based targeted data extraction is relatively robust. In itself the approach is also not new *per se* and has been previously used under the name accurate mass tag [23]. Also, most of the shotgun data analysis software makes use of linking ID information to MS1 data of different runs (e. g. match-between-runs in MaxQuant [24]). However, these algorithms are usually only reliable within data sets with highly similar runs. The analysis performed here has certainly also benefited from the fact that all LC-MS runs were carried out on the same instrument platform within a relatively short time span (3-5 days).

In the bigger picture, however, the idea of DIA is to provide the possibility of high throughput. This entails centralized spectral libraries (e. g. online repositories) that can be directly accessed for targeted data extraction of any experiment to avoid the tedious task of generating spectral libraries [25]. Thus the similarities between analyses used for spectral libraries and for the actual quantitative experiment will likely decrease. Therefore, a direct and robust application of MS1 quantification seems unlikely and more confidence through MS2 data is prerequisite.

3.5 Considerations about assay specificity and sensitivity

Both ETD and EThcD are affected by a relatively low duty cycle [26, 27]. DIA however greatly relies on acquisition speed. In the data analysis described above this was not taken into consideration and cycle times differed between 3 s for HCD and 6 s for ETD/EThcD. In order to decrease the cycle time for EThcD-DIA we performed EThcD-DIA methods with fewer isolation windows (Figure 3A). Additional to the 30 x 20 Th windows we performed the same experiment with 15 x 40 Th (cycle time 3 s) and 5 x 120 Th (cycle time 1 s). To investigate the effect of the loss in MS2 specificity we first applied an mProphet model trained on the 30 x 20 Th window dataset to all three EThcD-DIA runs. As seen in Figure 3B, increasing the isolation width from 20 Th to 40 Th results in a slight decrease from 4670 to 4258 peptides (-9 %). This effect is even more pronounced upon increasing isolation widths to 120 Th, where identifications drop to 2695 peptides. In both cases retraining the mProphet model could be used to increase the successful identifications to 4277 peptides and 3108 peptides respectively (Figure 3B). It is noteworthy that upon increasing isolation widths the acquired MS2-data loses specificity. Hence mProphet sub-scores contributions associated with MS2-data decrease with increasing isolation width. At the same time the contribution of MS1- and LC-based sub-scores increases (Figure 3C).

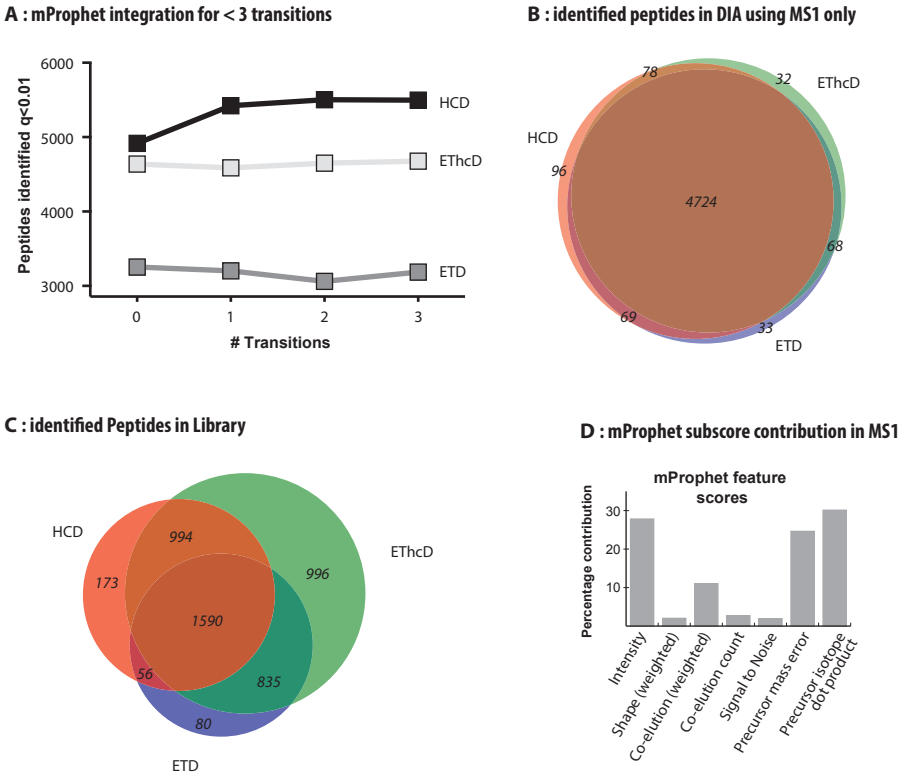


Figure 2: ***MS1-based targeted data extraction****. (A) Number of significantly (q ≤ 0.01) identified phosohopeptides upon stepwise reduction of transition numbers to zero, resulting in exclusively extracting MS1-based data. (B) A combined peptide library comprising DDA runs with HCD, ETD and EThcD was used to perform MS1 only targeted data extraction on 3 DIA analyses. This results in the significant (q ≤ 0.01) identification of 5 100 phosphopeptides depicted here as Venn diagram to showing an overlap of roughly 92.6 %. (C) Venn diagram grouping the 4 724 peptides identified in all three analyses to their spectral library of origin. (D) Contribution of multiple mProphet sub-scores to the composite score used for FDR-controlled reintegration of MS1 traces. (E-G) mProphet score distribution for the extraction of MS1 ion traces for target and decoy sequences of all HeLa-library phosphopeptides from: (E) HeLa tryptic digests enriched for phosphopeptides analyzed by DIA in triplicates, (F) HeLa tryptic digests enriched for phosphopeptides analyzed by DIA in triplicates and* E coli *tryptic digest analyzed by DIA in triplicates, (F)* E coli *tryptic digest analyzed by DIA in triplicates.*

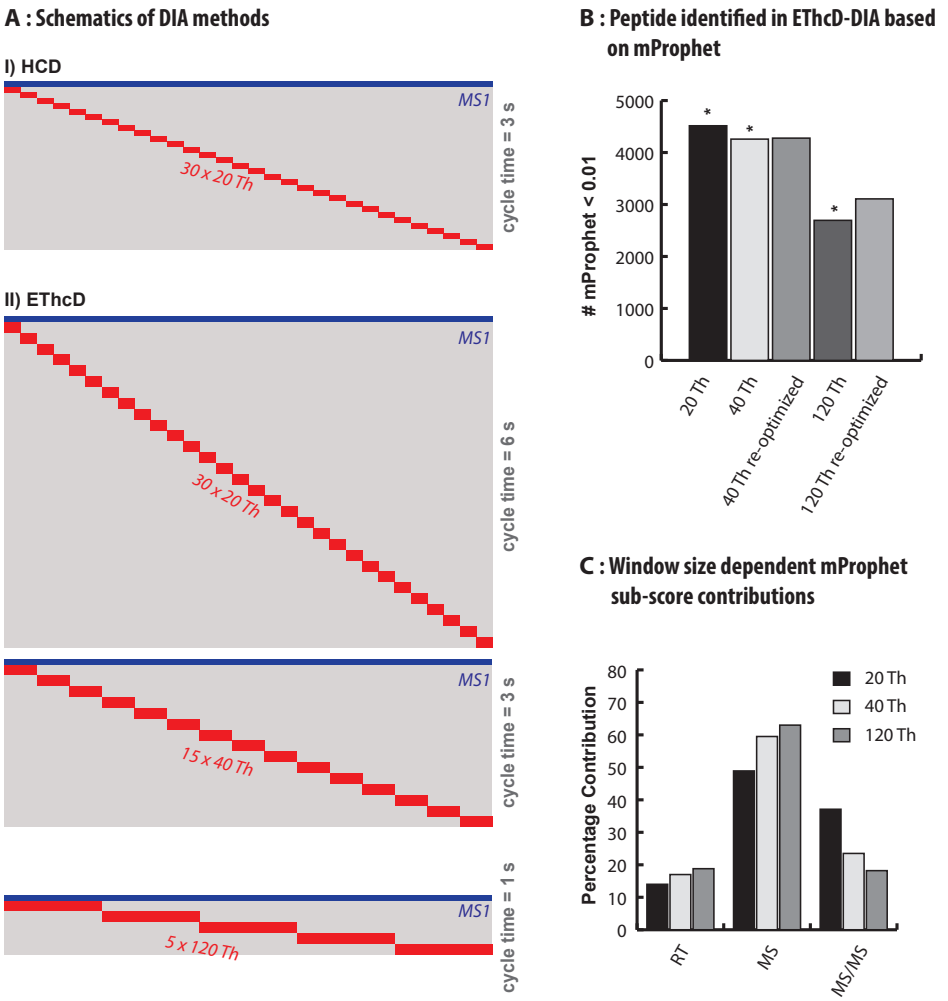


Figure 3: ***Method setup to compensate for longer ETD- and EThcD reaction time****. (A) A reference method used for HCD-DIA comprising 30 x 20 Th windows gives rise to a 3 s cycle time. Through longer reaction time of EThcD and ETD, an alike window distribution gives rise to a 6 s cycle time. Potential alternatives tested here apply lower numbers of larger isolation windows, specifically 15 x 40 Th (cycle time 3 s) and 5 x 120 Th (cycle time 1 s). (B) mProphet analysis of methods using wider windows result in less identifications. Numbers of significantly identified peptides are shown for all different window sizes analyzed with a standard mProphet model optimized for the 20 x 30 Th method (asterisk) and with retrained mProphet models specifically optimized for each data set. (C) mProphet subscore contribution grouped in LC-, MS1-, and MS2-dependent subscores optimized for different isolation window sizes shows an increasing importance of LC- and MS1-based sub-scores upon increasing isolation width.*

Additionally, the choice of fragmentation method can greatly affect the sensitivity of the DIA method. The intensity of the extracted fragment ion traces in an EThcD-DIA method are lower than in HCD, while noise levels stay similar. This is likely due to the AGC con-

trolled fill time, keeping total ion numbers injected into the Orbitrap similar. Due to the fact that EThcD gives rise to c- and z-ions in addition to the common b- and y-ions common to HCD, the intensity of each ion decreases. This effect is exemplified for two peptides in Figure 4 (a global analysis of integrated peak areas for precursor and fragment ions can be found in Supporting Information Figure S2). This however can partially be corrected for by summing up the AUC of a higher number of transitions.

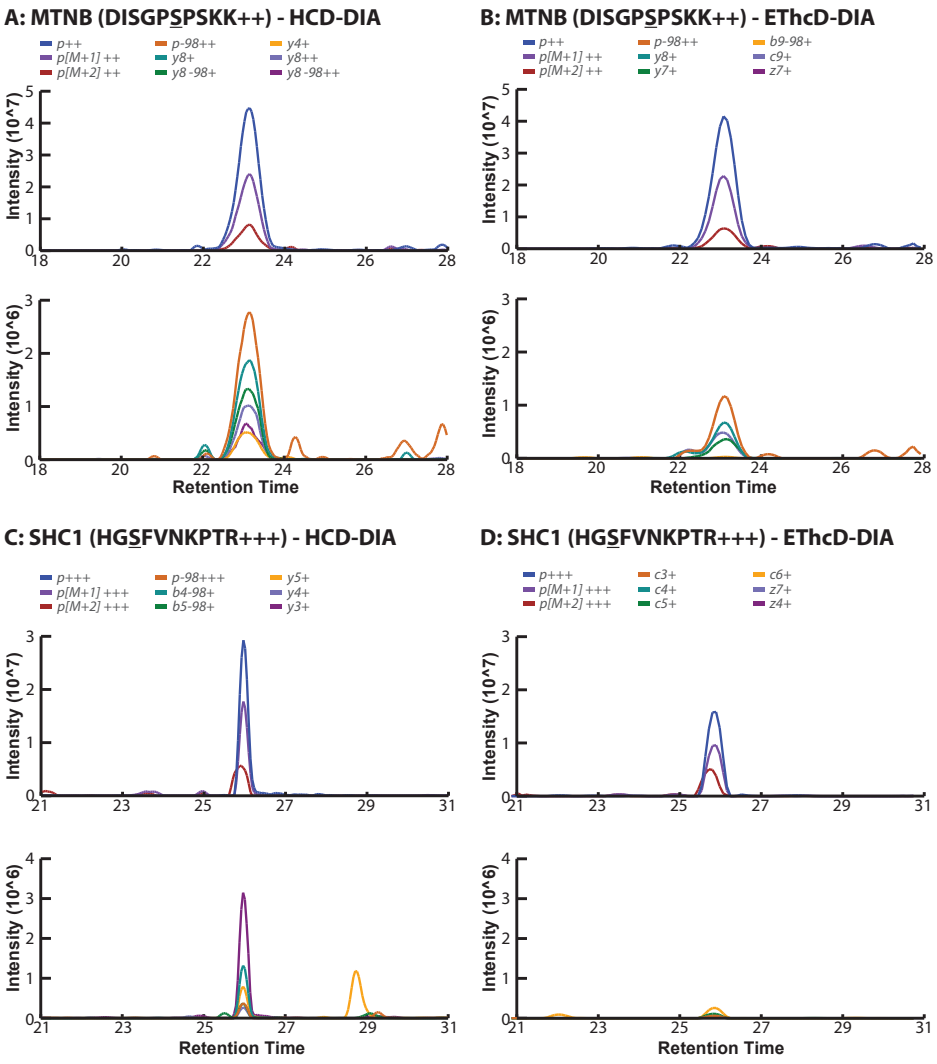


Figure 4: ***Two representative peptides measured by HCD-DIA and EThcD-DIA demonstrate how fragment ion sensitivity can decrease upon EThcD fragmentation.*** *(A-B) MNTB peptide DISGPSPSKK shows similar XIC intensity for precursor ions (upper panels), however roughly 2-fold lower intensity in MS2 XICs upon using EThcD compared to HCD (lower panel). (C-D) SHC1 peptide HGSFVNKPTR shows a similar trend.*

3.6 Considerations about phosphosites localization

A general challenge in the analysis of phosphopeptides by DDA LC-MS is the localization of the phosphorylation site on the peptide sequence. This is especially pronounced once a mixture contains several phosphoisomers, as these peptides are identical in precursor mass. MS1 based quantification can thus only be applied successfully if the different phosphoisomers are completely separated by LC. This however is highly dependent on the specific peptide sequences and the separation power of the chromatography used. Therefore, DIA presents a promising tool for the quantification of different phosphoisomers by extracting chromatographic traces of isomer specific fragment ions. Thus, phosphoisomers could be differentiated even if no chromatographic separation is achieved. However, the current data analysis strategy relies on shotgun based spectral library acquisition. Therefore, the DIA data analysis is directly linked to the challenges occurring in shotgun. Usually phosphosite localizations arising from database searches are scored and assigned with a localization probability by dedicated software tools such as PhosphoRS [17]. However, effects such as dynamic exclusion can drastically hamper the fragmentation of phosphoisomers eluting later in the chromatogram. Additionally, multiplexed MS2-specta containing various isoforms exist. All these effects can greatly influence the quality of the spectral libraries used for DIA targeted data analysis.

To better understand this effect, we designed a controlled experiment in which defined amounts of synthetic peptides were spiked into a tryptic HeLa digest at different concentrations, resulting in the injection of 8 fmol and 800 amol respectively. The synthetic peptide library used for these experiments were the same as used in **Chapter 3** of this thesis. They contain a total of 127 phosphoisomers and were therefore considered suitable for the experiment planned. Figure 5 illustrates the shortcomings of phosphosite localization in DIA for a peptide sequence representing the activation loop of the DDR2 kinase. The sequence is present in three different phosphorylation states, phosphorylated once at each of the 3 tyrosine residues. The 3 phosphoisomers are known to be baseline resolved in RP-chromatography (Figure 5A). At high concentration all 3 isoforms of the peptide were identified by Mascot (Figure 5B) and could be successfully extracted from the DIA data, showing chromatographic baseline separation (Figure 5C). Upon reducing the injection amount to 800 amol none of the 3 peptides was identified in shotgun anymore, however targeted data extraction from the DIA data was still successful (Figure 5D). An identical data analysis has been performed for three phosphisomers of synthetic peptides representing the DDR1 kinase t-loop. At 800 amol only the earliest eluting isoform was identified in DDA, thus preventing the targeted data extraction of the other two isoforms in DIA despite clear evidence of 3 different DIA-traces (Figure 6).

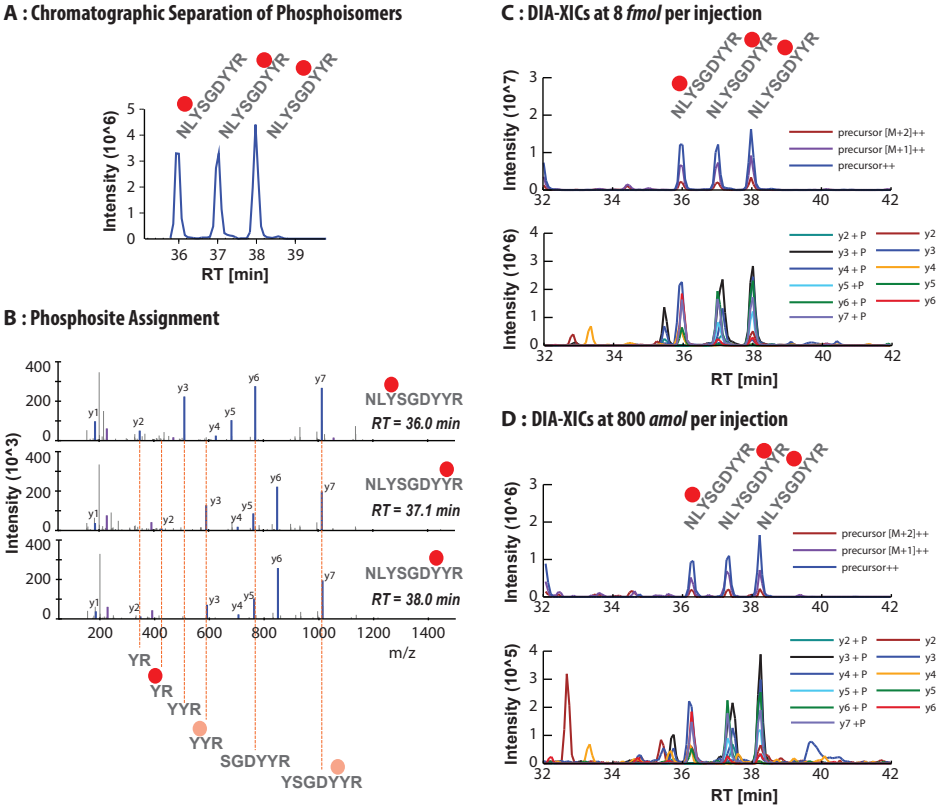


Figure 5: ***DDR2 – an example to visualize the potential benefit of DIA for the analysis of phosphoisomers and the challenges associated with it.*** *(A) Synthetic peptides for three phosphoisomers of a peptide representing the DDR2 t-loop are used for this analysis. They differ only in the position of their phosphotyrosine, present at three different positions within the peptide sequence. The 3 peptides are known to be chromatographically baseline separated. (B) Upon spiking the synthetic peptides into a HeLa digest background at high concentration (8 fmol per LC-MS injection), all three phosphoisomers are identified correctly by MASCOT at their respective retention time. All three identifications are also unambiguous in terms of phosphosite localization. (C) Targeted data extraction of the three isoforms results in three baseline separated peak groups clearly attributable to the three phosphosite localization isomers by RT and determining fragment ions alike. (D) Upon reduced injection of synthetic peptides in the same HeLa background, none of the 3 peptides could be identified by shotgun LC-MS anymore. DIA measurements however still result in high quality extracted ion traces enabling identification and confident site localization of all three phosphoisomers.*

Altogether this data shows the potential capability of DIA to distinguish different phosphosite localization isomers. However, due to the tightly interwoven data analysis of shotgun data and subsequent targeted data extraction, DIA data cannot be used up to its full potential yet.

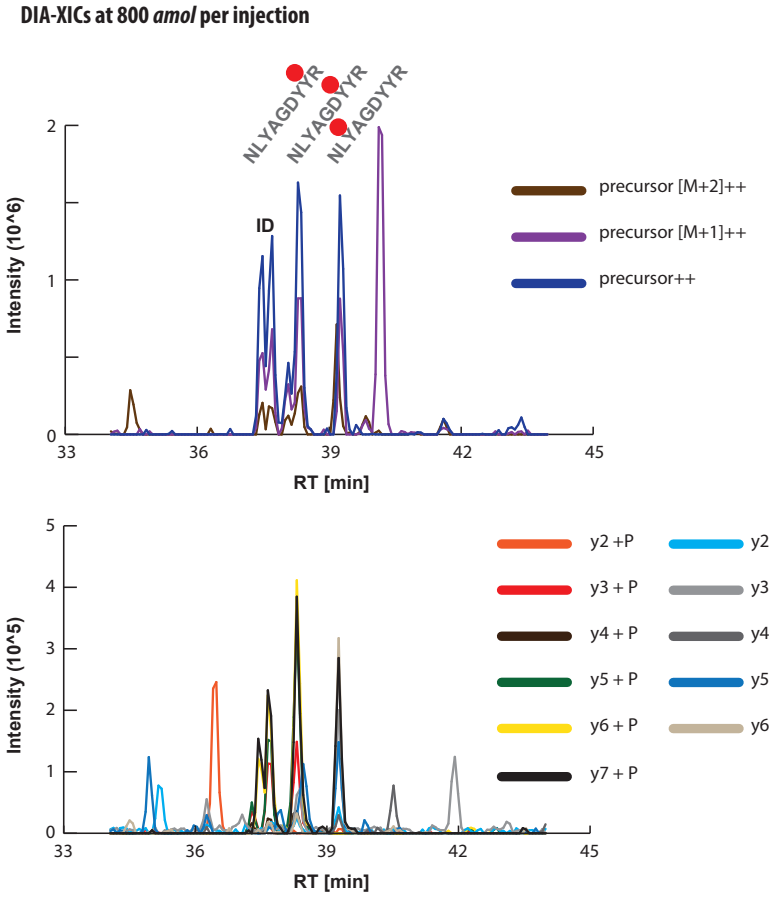


Figure 6: ***DIA signals for three phosphoisomers representing DDR1 t-loop phosphorylations.*** *Synthetic stable isotope labeled peptides for three phosphosites localization isomers were spiked into a background of digested HeLa lysates. At an injection amount of 800 amol only the isomer eluting earliest was identified (indicated as ID). Targeted data extraction, however, gives rise to three individual traces for precursor ions (top panel) and fragment ions (bottom panel).*

4 Discussion

Since the initial report in 2012, the application range of DIA has been expanding, ranging from classical proteome quantification up to PTM analysis and peptidomics [2]. As several of these applications have greatly profited from electron driven fragmentation methods, we set out to perform a feasibility study to their use in DIA measurements. To the best of our knowledge this is the first time the effect of different fragmentation methods on DIA measurements has been investigated.

Several key points stand out from this analysis. First it is evident, that in a classical DIA method using subsequent isolation windows of the same size, the need for speed becomes predominant. Both ETD and EThcD are known for long reaction times, which increases the cycle time drastically. Additional factors have to be considered when performing the experiments on Orbitrap based instruments. Trap-based instruments are intrinsically slow compared to quadrupole based fragmentation. Specifically, the requirement of high resolution/accurate mass MS2 data results in a longer cycle time, as each MS2 ion bundle has to be analyzed in the Orbitrap. It is striking that the use of 30 acquisition windows on an Orbitrap Fusion results in a cycle time of 3 s while similar analyses performed on a QTOF instrument can make use of 64 isolation windows in combination with only a slightly longer cycle time of 3.4 s (**Chapter 4** of this thesis). Reducing these cycle times to a chromatographic time scale, however, likely entails the use of either wider isolation windows, which comes at the cost of a reduction in specificity. Other workarounds could be the use of multiple methods spanning smaller MS1 ranges. This would retain the method's specificity, but also substantially increase the total analysis time and sample amount required. An alternative could be the use of MSX methods [9], which are capable of multiplexing several acquisition windows into one single MS2 spectrum and de-multiplex them again at the data analysis level. This should enable the specificity of relatively narrow isolation windows in combination with reasonably short cycle times. Via this approach, the lack of speed during the fragmentation reaction should not affect the interplay of cycle time and specificity. Although, it remains to be investigated if high multiplexing affects the sensitivity of the measurement. Essentially, the fill time and the AGC target control the number of ions entering the Orbitrap. In a highly multiplexed method, the composition of each ion bundle will be highly complex, and could thus drastically reduce the signal intensity of each individual ion. Increasing the AGC-target might circumvent this effect, but might result in space charge effects, thus compromising the mass accuracy of the acquired MS2-scan [28]. The tight control of the Orbitrap fill time likely also causes the loss of sensitivity for EThcD-DIA compared to HCD-DIA. The fact that each precursor ion gives rise to a z-, and c-ion series in addition to the y- and b-ion series also results in a lower intensity of each individual ion originating from the same amount of precursor ions. In ETD the poor fragmentation of the precursor in general largely compromises MS2 based quantification.

Most of these shortcomings are not necessarily unique to DIA analysis, as they can affect DDA alike. Nonetheless, especially EThcD is widely used in shotgun for certain sample types, because it outperforms HCD- or CID-fragmentation. Examples are the analysis of endogenous peptides [10], HLA peptides [29, 30] or glycopeptides [11]. It is not inconceivable that the analysis of these samples by EThcD-DIA is going to be implemented for the quantitative analysis of large cohorts of samples, for instance occurring in biomarker studies or even for diagnostic purposes [2]. A complete reliance on MS1 based data for identification and quantification as described here however is unlikely, because of the

strong reliance on reproducible LC-MS configurations over prolonged periods. It is however probable that in this kind of analysis the score contribution of MS1 data will be of increasing importance.

Especially for the analysis of phosphopeptides however, we envision that DIA can develop into an indispensable tool within the upcoming years. MS1 based quantification of phosphopeptides is of limited use when it comes to the analysis of different phosphosite isomers. Successful differentiation of these phosphosite isomers by DDA MS requires complete chromatographic separation and unambiguous identification of all isoforms involved. This however only rarely reflects what happens in reality, where chromatographic peaks might overlap or identification might be ambiguous in terms of phosphosite localization. The complete MS2-map provided by DIA analyses should theoretically be impartial to these effects. However, as of to date, the shortcomings of shotgun are still directly transferred into the DIA data analysis through the use of shotgun derived spectral libraries. A possibility to circumvent this problem could be the use of synthetic peptides for library generation, as they can be analyzed also at high concentrations. Initial efforts to generate peptide libraries for all hypothetically possible phosphopeptides of the human proteome are already under way [31]. Alternatively, the issue of phosphosite localization in DIA could be tackled by the development of dedicated pieces of software. Key to success in this case would be to disconnect phosphosite localization from the database search and integrate it into the DIA targeted data extraction. A potential piece of software envisioned here would consist of the following steps: (1) Combinatorial distribution of the localization of each phosphosite along the peptide sequence. (2) Simulation of MS2 spectra for each phosphosite localization isomer based on the MS2 spectra of the detected phoshopeptide. (3) Targeted data extraction for all possibilities. (4) mProphet sub-score for the occurrence of phosphosite determining ions and exclusion of non-occurring isomers. Each of these steps however would require tight control and new statistical tools to avoid an explosion of newly reported phosphosites that later turn out to be false positives. Alternatively, tools to control the assignment of phosphosite localization could be implemented within the framework of spectral library-free DIA data analysis workflows such as DIAUmpire [32] or PECAN [33].

5 Conclusion and Outlook

Altogether this study demonstrates the versatility DIA methods can provide these days. Through careful optimization of parameters such as fragmentation technique, isolation window size and distribution the success rate of DIA experiments can be greatly improved in a sample specific way. Specifically, on Orbitrap based instruments the balance between speed and sensitivity needs to be carefully tailored and will eventually require rational

choices based on factors such as sample type, sample complexity, need for sensitivity, and throughput. All of these factors eventually also influence data analysis, as programs such as mProphet can be dynamically adapted. This effect can definitely be exploited as shown here for the analysis of DIA runs with less informative MS2 data. Awareness of those effects however seems pivotal, as a greater reliance on MS1 and LC data also requires a high level of confidence in this data. Extensive instrument maintenance is thus absolute prerequisite as any kind of mass deviation or LC fluctuation drastically scrutinizes the subsequent data analysis. Bearing these factors in mind DIA however provides a great expansion of the proteomics toolbox. Today mostly used for high throughput analysis, it is not inconceivable that future improvements of the data analysis pipeline enable DIA to outperform shotgun in many aspects. Especially in terms of phosphoproteomics we envision DIA to become the tool of the future.

Acknowledgements

A.F.M.A. is supported by The Netherlands Organization for Scientific Research (NWO) through a VIDI grant (Grant 723.012.102). This work is part of the project Proteins At Work, a program of The Netherlands Proteomics Centre financed by The Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-Scale Research Facilities of The Netherlands (Project Number 184.032.201).

References

[1] Gillet, L. C., Navarro, P., Tate, S., Röst, H., *et al.*, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* 2012, *11*, O111.016717.

[2] Anjo, S. I., Santa, C., Manadas, B., SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications. *Proteomics* 2017, *17*.

[3] Guo, T., Kouvonen, P., Koh, C. C., Gillet, L. C., *et al.*, Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat Med* 2015, *21*, 407-413.

[4] Liu, Y., Buil, A., Collins, B. C., Gillet, L. C., *et al.*, Quantitative variability of 342 plasma proteins in a human twin population. *Mol Syst Biol* 2015, *11*, 786.

[5] Liu, Y., Hüttenhain, R., Surinova, S., Gillet, L. C., *et al.*, Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *Proteomics* 2013, *13*, 1247-1256.

[6] Zawadzka, A. M., Schilling, B., Held, J. M., Sahu, A. K., *et al.*, Variation and quantification among a target set of phosphopeptides in human plasma by multiple reaction monitoring and SWATH-MS2 data-independent acquisition. *Electrophoresis* 2014, *35*, 3487-3497.

[7] Schmidlin, T., Garrigues, L., Lane, C. S., Mulder, T. C., *et al.*, Assessment of SRM, MRM(3) , and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer. *Proteomics* 2016, *16*, 2193-2205.

[8] Zhang, Y., Bilbao, A., Bruderer, T., Luban, J., *et al.*, The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *J Proteome Res* 2015, *14*, 4359-4371.

[9] Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., *et al.*, Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods* 2013, *10*, 744-746.

[10] Parker, B. L., Burchfield, J. G., Clayton, D., Geddes, T. A., *et al.*, Multiplexed temporal quantification of the exercise-regulated plasma peptidome. *Mol Cell Proteomics* 2017.

[11] Yu, Q., Wang, B., Chen, Z., Urabe, G., *et al.*, Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *J Am Soc Mass Spectrom* 2017.

[12] Frese, C. K., Zhou, H., Taus, T., Altelaar, A. F., *et al.*, Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J Proteome Res* 2013, *12*, 1520-1525.

[13] Batth, T. S., Francavilla, C., Olsen, J. V., Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res* 2014, *13*, 6176-6186.

[14] Post, H., Penning, R., Fitzpatrick, M. A., Garrigues, L. B., *et al.*, Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* 2017, *16*, 728-737.

[15] Cristobal, A., Hennrich, M. L., Giansanti, P., Goerdayal, S. S., *et al.*, In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. *Analyst* 2012, *137*, 3541-3548.

[16] Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 2007, *4*, 923-925.

[17] Taus, T., Köcher, T., Pichler, P., Paschke, C., *et al.*, Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res* 2011, *10*, 5354-5362.

[18] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., *et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, *26*, 966-968.

[19] Zi, J., Zhang, S., Zhou, R., Zhou, B., *et al.*, Expansion of the ion library for mining SWATH-MS data through fractionation proteomics. *Anal Chem* 2014, *86*, 7242-7246.

[20] Shao, S., Guo, T., Koh, C. C., Gillessen, S., *et al.*, Minimal sample requirement for highly multiplexed protein quantification in cell lines and tissues by PCT-SWATH mass spectrometry. *Proteomics* 2015.

[21] Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., *et al.*, mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods* 2011, *8*, 430-435.

[22] Bruderer, R., Bernhardt, O. M., Gandhi, T., Reiter, L., High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. *Proteomics* 2016, *16*, 2246-2256.

[23] Smith, R. D., Anderson, G. A., Lipton, M. S., Pasa-Tolic, L., *et al.*, An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* 2002, *2*, 513-523.

[24] Bielow, C., Mastrobuoni, G., Kempa, S., Proteomics Quality Control: Quality Control Software for MaxQuant Results. *J Proteome Res* 2016, *15*, 777-787.

[25] Rosenberger, G., Koh, C. C., Guo, T., Röst, H. L., *et al.*, A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci Data* 2014, *1*, 140031.

[26] Frese, C. K., Altelaar, A. F., van den Toorn, H., Nolting, D., *et al.*, Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem* 2012, *84*, 9668-9673.

[27] Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 2004, *101*, 9528-9533.

[28] Kharchenko, A., Vladimirov, G., Heeren, R. M., Nikolaev, E. N., Performance of Orbitrap mass analyzer at various space charge and non-ideal field conditions: simulation approach. *J Am Soc Mass Spectrom* 2012, *23*, 977-987.

[29] Mommen, G. P., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., *et al.*, Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc Natl Acad Sci U S A* 2014, *111*, 4507-4512.

[30] Marino, F., Mommen, G. P., Jeko, A., Meiring, H. D., *et al.*, Arginine (Di)methylated Human Leukocyte Antigen Class I Peptides Are Favorably Presented by HLA-B*07. *J Proteome Res* 2017, *16*, 34-44.

[31] Zolg, D. P., Wilhelm, M., Schnatbaum, K., Zerweck, J., *et al.*, Building ProteomeTools based on a complete synthetic human proteome. *Nat Methods* 2017, *14*, 259-262.

[32] Tsou, C. C., Avtonomov, D., Larsen, B., Tucholska, M., *et al.*, DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* 2015, *12*, 258-264, 257 p following 264.

[33] Ting, Y. S., Egertson, J. D., Bollinger, J. G., Searle, B. C., *et al.*, PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data. *Nat Methods* 2017, *14*, 903-908.

Supporting Information

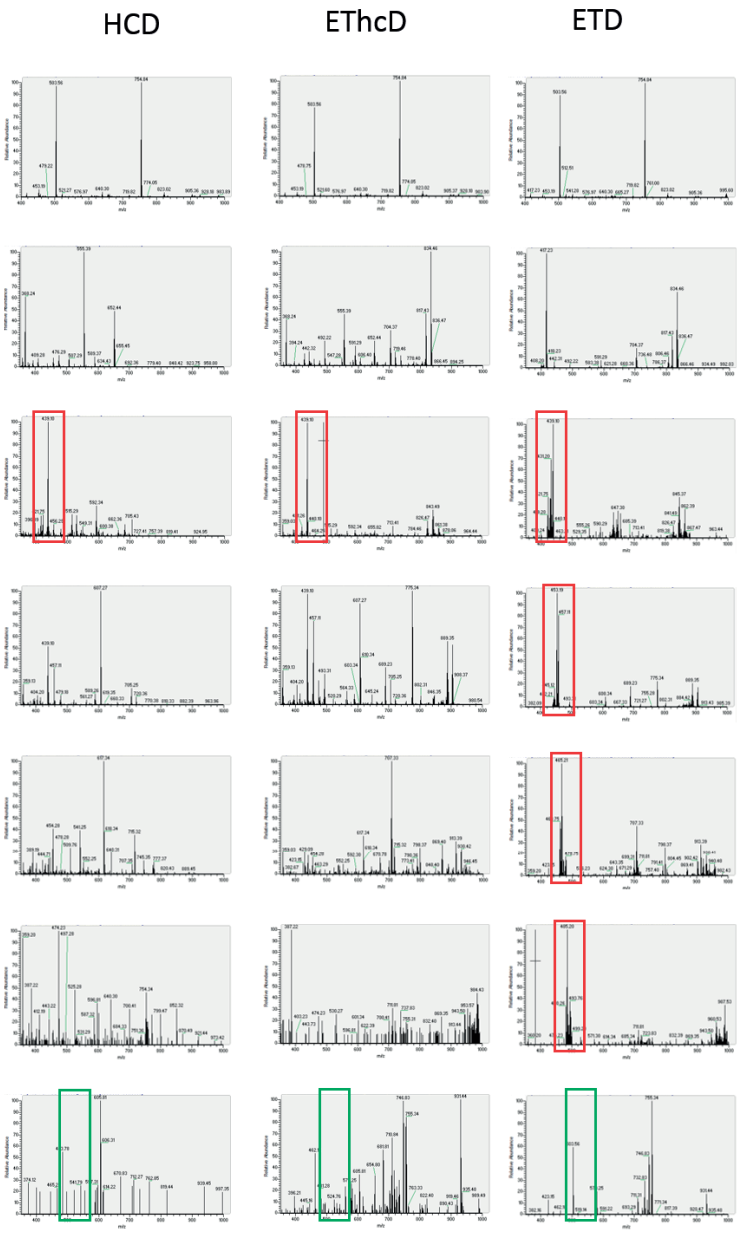


Figure S1: Illustation of typical fragment spectra obtained for HCD-DIA, EThcD-DIA and ETD-DIA across one DIA cycle. To exclude any bias from the chromatography, each cycle depicted here was extracted from alike chromatographic regions. The apex of the MS1 signal at *m/z* = 754.84 (2+) was randomly chosen as a starting point. This results in minor RT differences between the 3 runs (HCD-DIA: RT = 53.63 min, EThcD-DIA: RT = 53.56 min, ETD-DIA: RT = 53.51 min). Red squares indicate areas in the MS2 maps dominated by non-fragmented signals, evidenced by the increasing signal in the *m/z* range of the isolation window. This effect is predominantly observed in ETD-DIA. Green squares indicate the isolation of highly abundant ions observed in the MS1 measurement (*m/z* = 503.56, 719.82, 754.84, 823.02, 905.36).

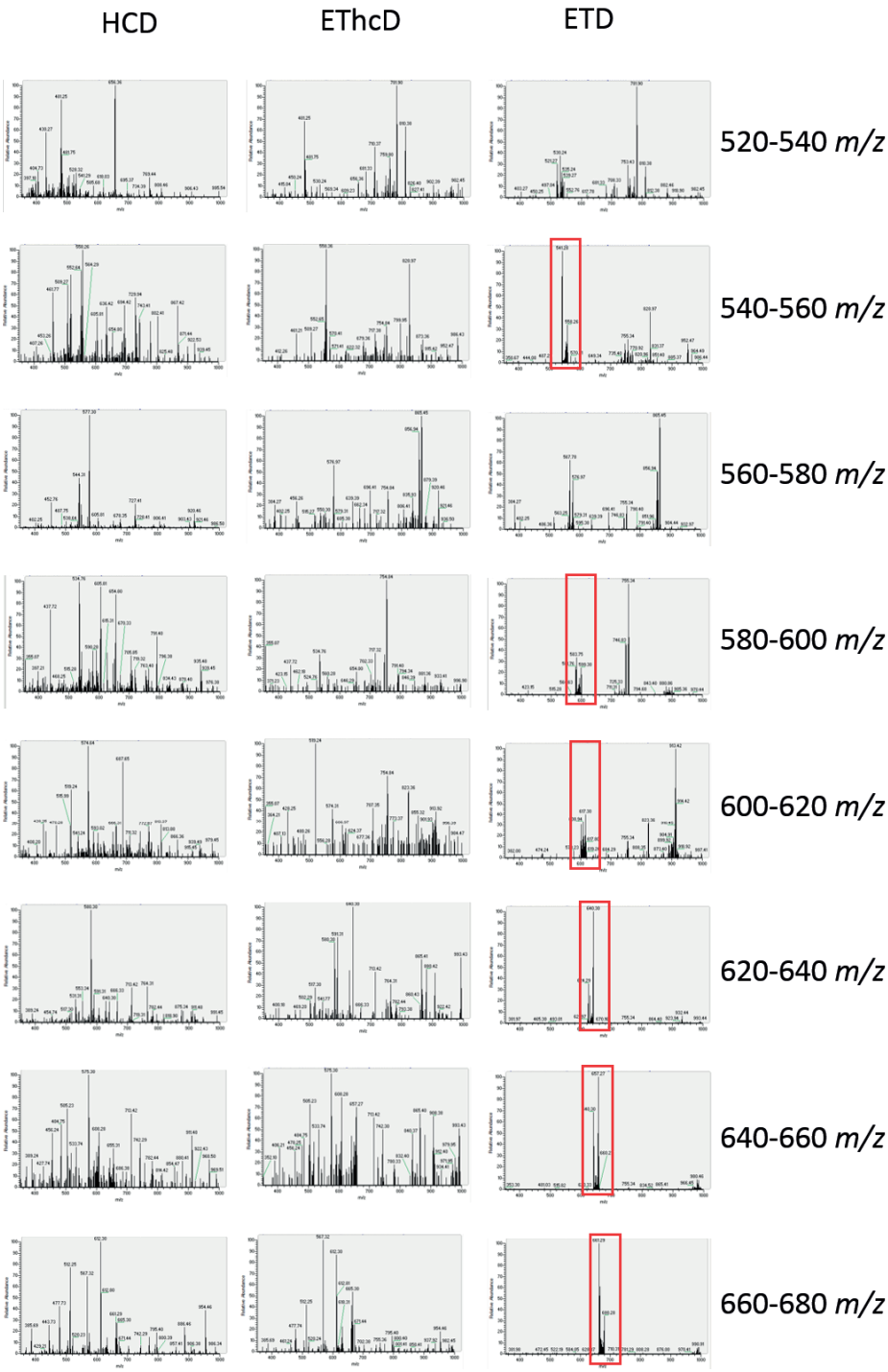

HCD
EThcD
ETD
520-540 m/z
540-560 m/z
560-580 m/z
580-600 m/z
600-620 m/z
620-640 m/z
640-660 m/z
660-680 m/z

Figure S1

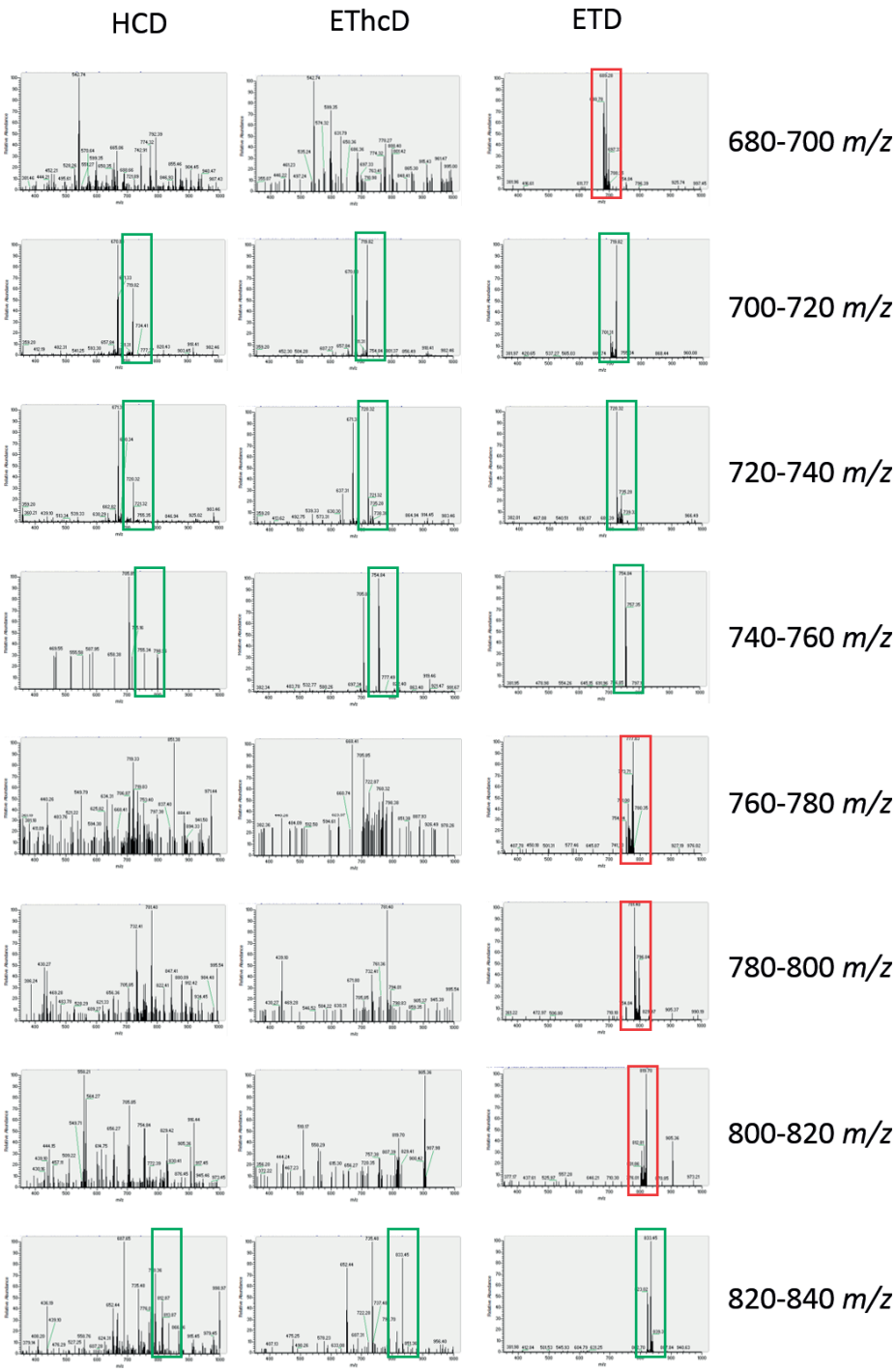

HCD
EThcD
ETD
680-700 m/z
700-720 m/z
720-740 m/z
740-760 m/z
760-780 m/z
780-800 m/z
800-820 m/z
820-840 m/z

Figure S1

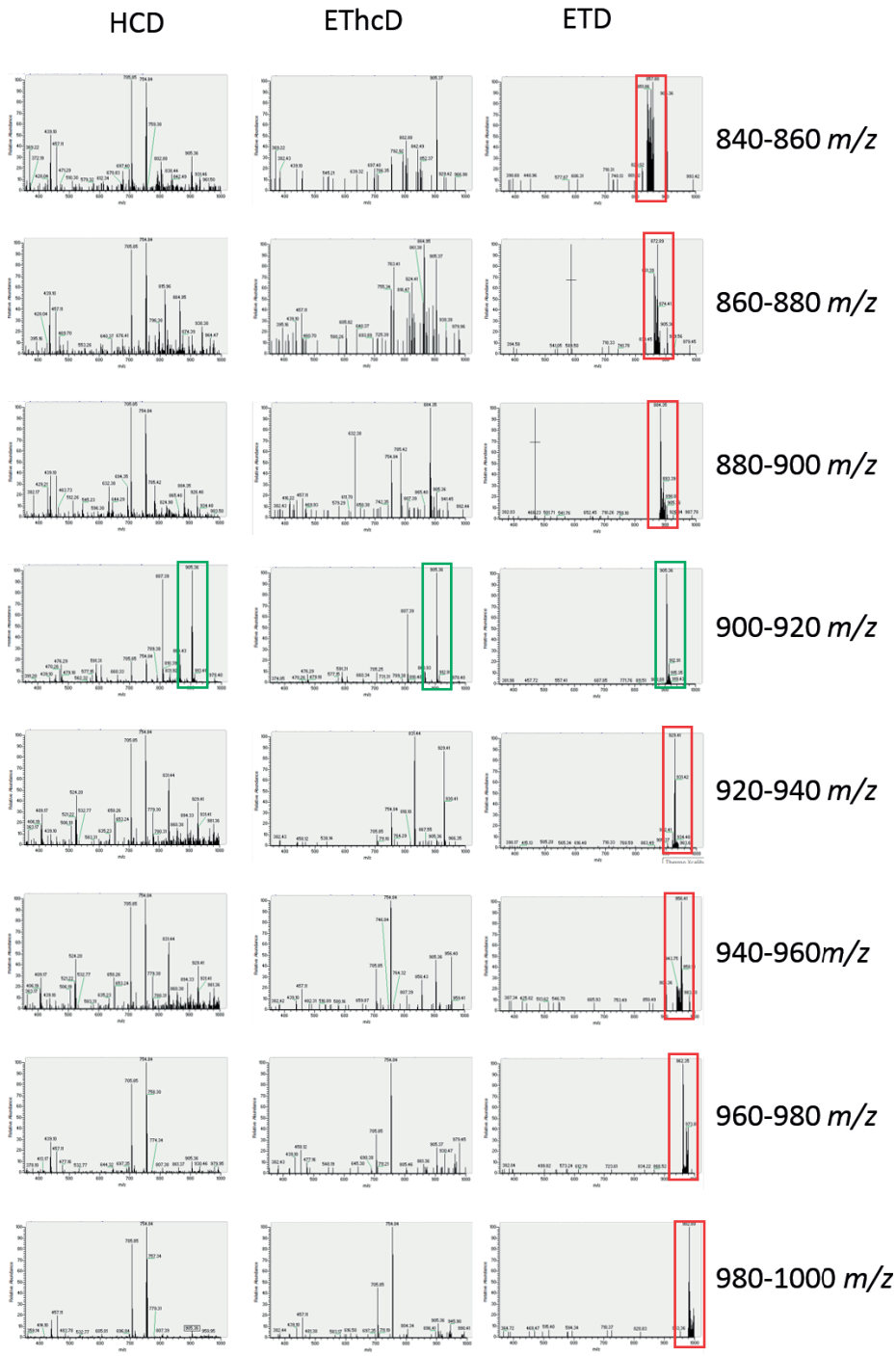

HCD
EThcD
ETD
840-860 m/z
860-880 m/z
880-900 m/z
900-920 m/z
920-940 m/z
940-960m/z
960-980 m/z
980-1000 m/z

Figure S1

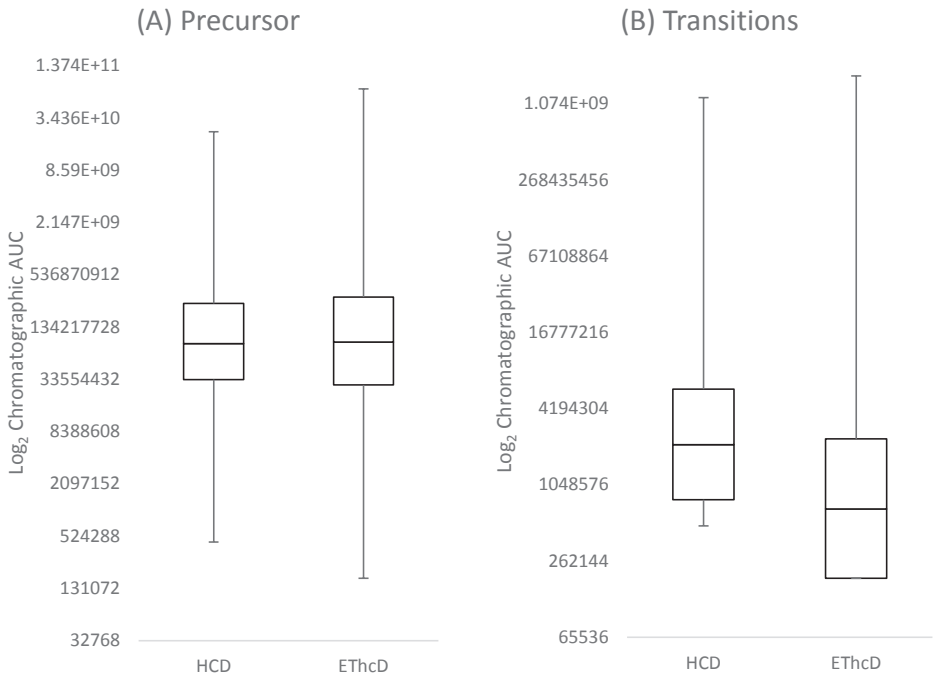


Figure S2: Intensity distribution of fragment ions in HCD and EThcD for all significantly identified peptides per method. (A) Chromatographic peak areas for precursor ions (M, M+1, M+2) show alike distribution for HCD-DIA and EThcD-DIA, proofing robust chromatography and MS1 behavior. (B) Integrated peak area for all transitions show a different intensity distribution. EThcD results in decreasing transition intensity compared to HCD. This is likely due to the occurrence of c- and z-ions additional to the b- and y-ions in combination with a less efficient fragmentation.

Chapter 6

Outlook, Summary, Deutsche Zusammenfassung, Nederlandse Samenvatting, List of Publications, Curriculum Vitae, Acknowledgements

Outlook – Where are Proteomics and Mass Spectrometry heading?

By the time I defend my PhD thesis it will be roughly seven years ago that I made the first step of the journey towards becoming a mass spectrometrist. Back then, I planned to perform my first research internship for my master program in a proteomics lab focusing on using DDA methods to analyse *in vitro* kinase reactions. I later on followed up on that topic during my master thesis within the same group, this time primarily working with SRM. It is probably due to a combination of the Zeitgeist back in 2011 and my lack of experience as a master student that I thought proteomics meant doing either DDA or SRM. The concept of data-independent acquisition – though in general already established in the form of MS^E [1], original DIA [2] or AIF [3] – had not been a major force in MS-based proteomics, at least not to my knowledge. The currently applicable SWATH technology was only in development back then, by a postdoc working next door. His process of preparing what is currently considered the landmark 2012 SWATH paper [4] could back then only be described as stumbling from frustration to frustration. However, since then the field of DIA has substantially matured with numerous novel DIA acquisition methods and data analysis strategies emerging.

The establishment and assessment of novel LC-MS/MS acquisition methods has afterwards also been the major red line throughout my PhD projects as evidenced by the four research chapters of this thesis ranging from the adaptation of SRM to the analysis of endogenous peptides up to the implementation of electron driven fragmentation in DIA. In this outlook I will thus first discuss my view on the future developments in MS acquisition and technology before moving into a broader discussion about MS-based proteomics in the context of biological and clinical research.

Where is MS-technology heading?

My personal vision for the future of MS-based proteomics is that DIA type of methods will eventually completely replace routine DDA measurements. Although I do not believe that these DIA methods have much in common with the currently used large window isolation methods. A major technological advantage of today's DDA methods over DIA is the relatively small Q1 isolation window, which allows to connect fragment ions to a specific precursor ion with great confidence. In practice this is even enhanced in a way that database searches assume all fragment ions are originating from the MS1 signal that has been used to define the isolation window. It is not unlikely that further developments in instrument speed will eventually allow for a continuous scanning of equally narrow isolation windows across the entire mass range on a chromatographic time scale. At this point DIA will essentially catch up with current DDA technology in terms of precursor selectivity whilst omitting the heuristic peak picking process. At this point the higher sensitivity, quantitative accuracy and specificity of MS2 versus MS1 quantification should ensure the

superiority of DIA over DDA. Efforts into this direction have already been made by current developments such as MSX, which essentially allows the use of isolation windows as narrow as in shotgun [5]. This method however exhibits two major drawbacks. (i) At the current state the method still lacks a robust data analysis pipeline for the deconvolution process. (ii) The acquisition encompasses the simultaneous readout of MS2 spectra from precursors of multiple isolation windows. These spectra contain fragment ions of various precursors which is prone to result in a general loss in sensitivity due to the tight control of total ion numbers injected into the Orbitrap (see **Chapter 5**). Therefore, I envision that the key technological advances in DIA instrumentation will focus primarily on increasing the speed of MS2 readouts without compromising the high resolution and mass accuracy that are currently state of the art in modern Orbitrap or TOF instruments.

Undoubtedly the discrete nature of the isolation windows cycling through the MS1 space will continue to pose certain challenges. Even if DIA isolation windows reach the narrow width used in DDA methods, it is very unlikely that they are perfectly centered around the precursor mass of the peptide analyzed. Historically the ion transmission rapidly deteriorates for ions when they are situated at the edge of the isolation window. Only recent developments in quadrupole technology made a more square-like isolation possible. For ions located directly at the edge of the isolation window, however, poor transmission still presents a major drawback. A very elegant way of solving this issue has been presented by Moseley *et al.* in 2017 [6]. Their method replaces the use of discrete isolation windows by continuous scanning of a 24 Th window across the entire precursor mass range. This ensures that proper placement of the precursor ion in the center of the isolation window happens for each analyte once per cycle. Additionally, fragment ions can be reliably aligned to precursor masses as their appearance and disappearance from the MS2 space can be directly linked to a precursor mass based on the known Q1 settings. Key for a successful implementation of this method, however, is a high speed MS2 readout.

The question remains whether DIA will eventually have the power to completely replace targeted proteomics. I am convinced that a large amount of experiments to date carried out by SRM/PRM will eventually be more easily applicable through targeted data extraction from DIA data. Nonetheless the analyte specific added sensitivity through CE and DP optimization or the added selectivity through the use of MRM^3 will likely result in the continued use of targeted proteomics methods for dedicated applications. The same is likely true for highly specific DDA methods implementing features such as MS^3, MS^n or neutral loss scans, or for methods that require lengthy fragmentation times.

Where are the applications heading? And how can we turn data into biological significance?

Recent technology developments in MS and LC have made it possible to identify and quantify thousands of proteins and PTMs within increasingly short time frames. This has

led to a virtual data explosion of information about protein abundance, their chemical modifications and interactions in a multitude of different biological model systems, patient samples, primary cells etc. This data explosion is not necessarily limited to proteins as also genomics, transcriptomics and metabolomics have matured substantially at the same time. Therefore, to date researchers and clinicians are presented with a wealth of data potentially relevant for drug development, patient diagnostics and treatment. This however can be both, a curse or a blessing. It becomes increasingly difficult to filter out biologically relevant data. In its essence, MS-based data can only deliver quantitative information. Thus, it is often unclear if the observed upregulation of a protein in a disease state is the cause or simply the result of the disease. Additionally, the sample sizes usually used in proteomics are comparatively small. Therefore, they often cannot account for the tremendous variations observed when comparing samples derived from different individuals rather than cell line based model systems. In its 'easiest' application an upregulated protein can be used as a diagnostic marker, in which case the question if it is the cause or the consequence of the disease is irrelevant. This however already involves lengthy validation processes including large independent cohorts of patients to ensure sufficient diagnostic power. Exploiting the predictive power of proteomics, e. g. for the discovery of novel vulnerabilities and drug targets, is even more difficult, as it requires a full picture of the biological functionality of a protein in an intra- and inter-cellular context. This layer of information is usually not included within the purely quantitative readout of proteomics. Therefore, each potential drug target discovered by proteomics needs to be carefully analyzed and validated by alternative biochemical methods in terms of its functionality within the biological system at hand before even initiating the tedious task of drug development. In this respect the majority of proteomics experiments will likely remain in the role of a hypothesis generator within the upcoming years, requiring extensive follow up validation experiments.

Nevertheless, multiple strives have been made in recent years to convert the purely quantitative information of MS-based proteomics into information about the behavior of proteins within the cellular context. This entailed information such as changes in protein-interaction as determined by AP-MS [7], cross-linking MS [8], thermal proteome profiling [9] or limited proteolysis [10]. Phosphoproteomics has also been increasingly used to link MS-based quantitative information to functionality. However, this mostly relies on the premise that any phosphorylation leads to protein activation. This however is a gross simplification as the majority of phosphosites reported from proteome analyses have not been studied in terms of functionality. Additionally, the few phosphosites that have been characterized in depth by biochemical methods often deliver a confusing or even self-contradictory picture as their functionality varies greatly depending on experimental setup, model system and external conditions. It is thus prerequisite to disentangle this situation. One way to do so is to reduce the information of global phosphorylation events

to the activity of it effectors, the protein kinases. Initially this has been performed in an indirect way by exploiting the concept of the consensus sequence. Through sequence motif analysis of the detected phosphosites, it is possible to discover peptide motifs clearly under- or overrepresented in specific conditions. Through prior knowledge of the kinases' substrate specificity, this can lead to information about differences in kinase activity. This approach however suffers from substantial drawbacks, most prominently the fact that consensus motifs are not known for the majority of protein kinases and the ones known might be redundant. Additionally, this approach requires highest standards of data quality. Motif analysis perform best upon large data input, therefore the identification of a large number of phosphopeptides is prerequisite for success. More importantly however, the identified phosphopeptides require highest confidence of phosphosite localization, as incorrect information about the site localization can largely compromise the sequence motif detection. **Chapter 3** presented in this thesis provides an alternative approach to link abundance with functionality by specifically quantifying known activation sites of kinases on a large scale. This approach alleviates most of the above-mentioned shortcomings. The requirements for experimental success however entail highest sensitivity and specificity in all aspects of the workflow including phosphopeptide enrichment, online HPLC and MS and can therefore so far only be performed in highly specialized labs. Further strives in instrument development - specifically in the direction of robust nano-flow HPLC – however might render the approach more accessible for broad use, providing a valuable tool in the toolbox of proteomics.

In summary, the field of MS-based proteomics has seen a wealth of technological innovations in recent years. Mass spectrometers have been improved in terms of sensitivity and speed, sophisticated sample preparation protocols have been developed for numerous different analyses and fractionation methods have been optimized to increase the proteomics depth. New innovations will continue this trend and open doors to new kinds of analyses. The sheer body of information about protein abundances will continue to grow exponentially. The big task for the future, however, is to figure out how to best exploit this wealth of information and converting it into new drugs, new diagnostics or new functional knowledge, thus contribution to general welfare.

References

[1] Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P., Geromanos, S. J., Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* 2006, *5*, 144-156.

[2] Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A., Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* 2004, *1*, 39-45.

[3] Geiger, T., Cox, J., Mann, M., Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol Cell Proteomics* 2010, *9*, 2252-2261.

[4] Gillet, L. C., Navarro, P., Tate, S., Röst, H.*, et al.*, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* 2012, *11*, O111.016717.

[5] Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W.*, et al.*, Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods* 2013, *10*, 744-746.

[6] Moseley, M. A., Hughes, C. J., Juvvadi, P. R., Soderblom, E. J.*, et al.*, Scanning Quadrupole Data Independent Acquisition - Part A. Qualitative and Quantitative Characterization. *J Proteome Res* 2017.

[7] Gingras, A. C., Gstaiger, M., Raught, B., Aebersold, R., Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* 2007, *8*, 645-654.

[8] Liu, F., Rijkers, D. T., Post, H., Heck, A. J., Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Methods* 2015, *12*, 1179-1184.

[9] Reinhard, F. B., Eberhard, D., Werner, T., Franken, H.*, et al.*, Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. *Nat Methods* 2015, *12*, 1129-1131.

[10] Feng, Y., De Franceschi, G., Kahraman, A., Soste, M.*, et al.*, Global analysis of protein structural changes in complex proteomes. *Nat Biotechnol* 2014, *32*, 1036-1044.

Summary

This thesis describes novel methods and applications of data-independent and targeted mass spectrometry. It employs Selected Reaction Monitoring (SRM) and data-independent acquisition (DIA) as core tools to build sensitive and robust assays to quantify molecular signaling events.

In **Chapter 1** a general introduction is given into the methods and systems used in this thesis. The chapter starts with an in-depth discussion about liquid chromatography (LC) and mass spectrometry (MS) based proteome research. Detailed information is given about different kinds of LC and MS instrumentation followed by a discussion about various data acquisition methods. Subsequently, the role of protein phosphorylation in cellular signaling is extensively reviewed. This includes a detailed description about function and structure of protein kinases as well as their interplay in a cellular context. A detailed discussion about LC-MS based phosphoproteomics is given, specifically focusing on studies, which successfully applied SRM or DIA to analyze protein phosphorylation events. Lastly, the biological role of neuropeptides and their analysis by LC-MS is discussed.

Chapter 2 focuses on exploiting SRM for the analysis of neuropeptides from rat hypothalami. Neuropeptide regulation is key in understanding a variety of functional processes, such as energy balance, drug addiction, etc. The biosynthesis of neuropeptides however involves multiple trimming steps and PTMs, impeding the prediction of the active peptide from the genome sequence or the precursor protein. Hence MS-driven analyses of undigested endogenous peptides are vital to determine the biologically active compounds. For SRM however, the chemical characteristics of endogenous peptides pose a substantial challenge to the MS analysis, as they do not always adhere to the optimal ionization and fragmentation considerations for a triple quadrupole instrument. The chapter describes a proof of principle study carried out on 15 neuropeptides to demonstrate the potential of SRM to also quantify long and highly basic endogenous peptides. Despite common practice it is shown that precursor and fragment charges of up to 5 and 3 respectively prove not to limit the quality of the SRM peak group and can be used for accurate quantification. However careful parameter optimization, especially values for transition-specific collision energies are key for sensitivity as default parameters are only optimized for tryptic peptides. Using this approach to quantify neuropeptide levels in hypothalami of rats subjected to different diets revealed several significant changes, most notably an upregulation of the VGF-derived signaling peptide AQEE-30 upon high caloric feeding.

Chapter 3 describes a novel approach to determine system wide kinase activation states based on direct quantification of phosphorylations in the t-loop region. This is highly challenging as t-loop phosphorylations are notoriously underrepresented in shotgun proteome analyses due to their low abundance. Therefore, the approach described employs the unparalleled sensitivity of SRM in combination with highly selective and sensitive phos-

phopeptide enrichment by Fe(III)-IMAC. Careful optimization of the SRM setup allowed assay development for 248 peptides, covering 221 phosphorylation sites in the t-loop region of 178 kinases. The targeted kinases included numerous clinically relevant kinases with FDA approved inhibitors such as MET, ABL, SRC, BTK, JAK3 and KIT. Global screening of t-loop phosphorylations, however, requires a robust number of transitions per peptide to achieve highest sensitivity and selectivity for accurate phosphosite localization. To be able to combine the system-wide screening approach even upon working with limited sample amounts such as primary cells, a two-phase strategy was developed. The strategy comprised a 'low-confidence' survey phase and a 'high-confidence' quantification phase for the initially observed phosphopeptides. Using the assays on different cell lines, patient-derived melanoma cells and primary blood platelets allowed monitoring activation of 54 kinases through 66 t-loop phosphosites. Stimulation of blood platelets revealed activation of cell type-specific kinases and profiling of activated kinases between drug resistant and sensitive melanoma cells revealed rewiring of the cancer signaling environment, highlighting potential novel drug targets, such as ERK4. The extreme sensitivity of the assay is highlighted through the reproducibly detected RIPK1 phosphorylation at serine 161 after TNFα stimulation, which has thus far only been reported using a protein expression system in combination with IP, in vitro kinase assay and phosphopeptide enrichment.

Chapter 4 employs SRM, DIA and MRM^3 to study alteration in the PI3K-mTOR/MAPK pathway in EGFR inhibitor (EGFRi) sensitive and resistant non-small cell lung cancer cell lines. Aiming at gaining new insights into the performance characteristics of the three types of MS-acquisition methods and into the molecular processes triggered by drug resistance in lung cancer we set out to analyze a total of 42 PI3K-mTOR/MAPK phosphosites by targeted proteomics. As reported before, we observed that DIA is less sensitive when compared to SRM. However, DIA exhibits increased flexibility, allowing for post-analysis selection of alternative phosphopeptide precursors. Additionally, we demonstrate the added benefit of MRM^3, allowing the quantification of two poorly accessible phosphosites. Unlike shotgun phosphoproteomics, where the vast majority of the observed phosphosites have an unknown function, SRM was exploited to monitor those sites known to be relevant, such as kinase activating or inhibiting sites as well as phosphosites with established roles in signal transduction. The targets included activity-inducing sites of several protein kinases (e. g. PKD1, BRAF), phosphorylation sites affecting the protein's binding partners (4EBP1, AKT1S1) and phosphorylation sites promoting the protein's degradation (PIK3C2A). Globally, we provide evidence that the EGFRi resistant cell line relies more on the MAPK cascade to maintain prosurvival signaling whereas the EGFRi sensitive cell line shows elevated mTORC1 levels.

Chapter 5 exclusively focusses on DIA method development and characterization for the analysis of protein phosphorylations. Primary focus is the attempt of combining DIA with electron-driven fragmentation and to investigate the effects this has on the data and

the data analysis. Major bottleneck of DIA combined with electron-driven fragmentation are the longer cycle times, which are generally unfavorable for DIA-type of methods. Throughout the chapter, we show various workarounds how to potentially cope with the longer duty cycle. Most prominently, these workarounds involve the use of wider isolation windows, which in turn result in a loss of specificity. An in-depth look was taken into the post-acquisition data analysis driven by mProphet. The dynamic scoring model was used to demonstrate how the weight of MS1 and LC characteristics such as retention time accuracy and precursor isotope distribution increase in the data analysis process when the quality of the fragment spectra decreases. In addition, the work demonstrates how the use of EThcD affects DIA signal intensity resulting in a general drop in sensitivity due to the occurrence of a higher number of fragment ions generated from the same amount of precursor ions. Lastly, synthetic phosphopeptides containing numerous different phosphoisomers were used to illustrate the potential advantage of DIA over shotgun for accurate quantification of different phosphoisomers. However, we also show how the shortcomings of shotgun in terms of phosphoisomer analysis are directly transferred to DIA when using spectral library based DIA data analysis approaches.

Chapter 6 provides an outlook, containing my prediction about the future of mass-spectrometry and proteomics including main challenges that need to be addressed. In the first part of this outlook I discuss my prediction for the future of DIA, which is currently mainly suffering from shortcomings in terms of selectivity during precursor ion isolation and the requirement of spectral libraries for data analysis. Alleviating these shortcomings is crucial for DIA to become a widely accepted tool in the future. I discuss recent promising studies heading into this direction as well as so far unsolved bottlenecks associated with DIA. In the second part of this outlook I highlight common issues with transferring the purely quantitative data obtained from MS-based proteomics into biological information useful for diagnosis, drug development, etc. In this process I discuss several approaches on how to link an MS-based quantitative readout to specific biological information such as structure, protein-protein interaction or enzymatic activity.

Deutsche Zusammenfassung

In der vorliegenden Dissertation beschreibe ich neue Methoden und Anwendungen im Bereich Massenspektrometrie-basierter Proteom-Forschung. Die beiden Schlüsseltechniken aller beschriebenen Experimente sind Selected Reaction Monitoring (SRM) und Data-independent Acquisition (DIA). Diese wurden angewendet, um robuste und sensitive Messmethoden für wichtige zelluläre Signal-Moleküle zu entwickeln.

Kapitel 1 gibt einen allgemeinen Überblick über die Methoden und Messinstrumente welche in dieser Dissertation zur Anwendung kamen. Das Kapitel beginnt mit einer Diskussion über Proteom-Forschung mit Hilfe von Flüssigchromatographie (liquid chromatography – LC) und Massenspektrometrie (MS). Dabei werden verschiedene Arten von LC- und MS-Instrumenten vorgestellt und mögliche Datenerfassungsmethoden aufgezeigt. Des Weiteren wird die Rolle von Proteinphosphorylierungen im Rahmen von zellulärer Signal-Übermittlung erklärt. Dieser Abschnitt beinhaltet eine detaillierte Beschreibung der Funktionen und des Aufbaus von Protein-Kinasen sowie deren Interaktionen im zellulären Umfeld. Darauf folgt eine detaillierte Diskussion über die Analyse des Phosphoproteoms mit Hilfe von LC-MS. Ein spezieller Fokus dieser Diskussion liegt auf jüngsten Studien, welche erfolgreich DIA oder SRM für die Quantifizierung von Proteinphosphorylierung angewendet haben. Zuletzt wird die biologische Bedeutung von Neuropeptiden beschrieben und aufgezeigt, wie diese mittels LC-MS basierter Methoden identifiziert und quantifiziert werden können.

Kapitel 2 fokussiert auf die Analyse von Neuropeptiden im Hypothalamus von Ratten mit Hilfe von SRM. Die Regulierung von Neuropeptiden im biologischen System ist ein wichtiger Schlüssel um eine Vielfalt von Prozessen, wie z. B. den Energieausgleich oder das Entstehen von Arzneimittel-Abhängigkeiten zu verstehen. Die Biosynthese von Neuropeptiden im Nervensystem ist allerdings ein komplizierter Prozess, welcher verschiedenste chemische Modifikationen auf Protein- und Peptidebene beinhaltet. Solche Prozesse beinhalten z. B. das gezielte Zerschneiden von Peptidbindungen durch Enzyme oder das Anhängen von Posttranslationalen Modifikationen. Diese Komplexität erschwert die Vorhersage der genauen Peptidsequenz von biologisch aktiven Neuropeptiden nur auf Basis der Gensequenz oder des Ursprungsproteins. Daher ist die Analyse von Neuropeptiden mit Hilfe von LC-MS-basierten Proteomanalysemethoden unabdingbar um biologisch relevante Peptide zu identifizieren. Allerdings muss darauf geachtet werden, dass während des Probenvorbereitungsvorgangs kein enzymatischer Verdau mit Trypsin ausgeführt wird. Dieser enzymatische Verdau ist in der klassischen Proteom-Forschung üblich, würde allerdings im Falle von Neuropeptiden zu einem Informationsverlust über die tatsächliche Beschaffenheit der Peptide führen. Die chemisch-physikalischen Eigenschaften von natürlich vorkommenden Peptiden, welche nicht mit Trypsin verdaut wurden, unterscheiden sich allerdings deutlich von den klassischen tryptischen Peptiden. Für

die Analyse von Neuropeptiden mittels SRM bedeutet dies eine grosse Herausforderung, da die Peptide in Bezug auf Ionisierung und Fragmentierung nicht mit tryptischen Peptiden vergleichbar sind. Das vorliegende Kapitel beschreibt den konzeptionellen Beweis, dass SRM fähig ist lange und hoch-basische Peptide robust zu detektieren und zu quantifizieren, was anhand von 15 Neuropeptiden demonstriert wurde. Es gelang uns zu zeigen, dass Ladungszustände von bis zu fünffach positiv auf Peptid-Ebene und bis zu dreifach positiv auf Fragment-Ebene keine Qualitätseinbusse für das SRM Signal nach sich ziehen und daher für eine akkurate Quantifizierung genutzt werden können. Allerdings bedingt dies eine sorgfältig durchgeführte Optimierung von LC-MS Parametern. Hauptsächlich die Parameter für die Kollisionsenergie müssen deutlich verändert werden, da die Standardwerte normalerweise nur bei tryptischen Peptiden optimale Resultate erzielen. Durch sorgfältige Optimierung erreichten wir so eine substantiell höhere Sensitivität unserer Messungen. Die optimierte Methode wurde in der Folge angewendet um Neuropeptide im Hypothalamus von Ratten zu quantifizieren. Als Modelsystem untersuchten wir Ratten, welche unterschiedlichen Nahrungsregimes ausgesetzt worden waren um so Fettleibigkeit zu simulieren. Wir konnten feststellen, dass das Neuropeptid AQEE-30 in übergewichtigen Ratten im Vergleich zu normalgewichtigen Ratten signifikant erhöhte Werte aufwies.

Kapitel 3 beschreibt eine neue Methode um Aktivierungszustände von Kinasen zu bestimmen. Dazu quantifizierten wir Phosphorylierungen in der t-loop Region von Kinasen. Diese Phosphorylierungen sind essentieller Bestandteil im Aktivierungsprozess von Kinasen, daher kann die Menge an t-loop Phosphorylierung direkt als Messwert für die Aktivität einer Kinase verstanden werden. Allerdings stellt diese Aufgabe eine drastische analytische Herausforderung dar, da t-loop Phosphorylierungen nur in äusserst geringen Mengen in der Zelle vorkommen und daher in klassischen Proteom- oder Phosphoproteomanalysen klar untervertreten sind. Aus diesem Grund nutzten wir die beispiellose Sensitivität von SRM in Kombination mit hochselektiven Phosphopeptid-Anreicherungsmethoden aus. Durch sorgfältige Optimierung von LC- und MS-Parametern gelang es uns, spezifische und sensitive Testmethoden für 248 Phosphopeptide zu entwickeln, welche 221 Phosphorylierungsstellen in der t-loop Region von 178 Kinasen abdecken. Dies entspricht rund einem Drittel aller bekannten Kinasen im menschlichen Proteom. Die angepeilten Kinasen beinhalten zahlreiche Kinasen welche in der klinischen Forschung als hochrelevant gelten und für welche von der FDA autorisierte Inhibitoren auf dem Markt sind, z. B. MET, ABL, SRC, BTK, JAK3 und KIT. Unser Ziel war, eine Methode zu entwickeln, welche in einer beliebigen menschlichen Probe, in kürzester Zeit, anhand von möglichst wenig Ausgangsmaterial möglichst sensitiv die t-loop Phosphorylierungen von allen von uns abgedeckten 178 Kinasen überprüfen kann. Dies stellt allerdings eine gewaltige analytische Herausforderung dar, da Sensitivität und genaue Lokalisierung der Phosphorylierungsposition eine beträchtliche Zahl an Messpunkten verlangt. An einem bestimmten Punkt übersteigt dies allerdings die Kapazität des Massenspektrometers.

Klassischerweise werden in einem solchen Fall die anvisierten Peptide auf mehrere Methoden verteilt, was allerdings die gesamte Messzeit wie auch die benötigte Probenmenge erhöht. Vor allem Letzteres ist nicht immer möglich, z. B. wenn es um die Analyse von primären Zellen geht und dadurch das vorhandene Probenmaterial sehr limitiert ist. Um den Aspekt der systemweiten Analyse auch dann aufrecht erhalten zu können, wenn nur wenig Probenmaterial zur Verfügung steht, entwickelten wir eine zwei-Phasen Strategie. Die Strategie besteht aus einer Übersichtsphase mit geringer Anzahl an Messpunkten (und folglich leicht verringerter Messsicherheit) und einer Quantifizierungsphase, in welcher nur noch bereits detektierte Kinasen mit hoher Messsicherheit gemessen und quantifiziert werden. Die beschriebene Methode wurde verwendet um verschiedenste Laborzelllinien sowie von Patienten abgeleiteten Hautkrebs-Zelllinien und primäre Blutplättchen zu analysieren, was es uns ermöglichte die Aktivierung von 54 Kinasen über 66 t-loop Phosphorylierungsreaktionen nachzuweisen. Wir konnten zeigen, dass die Stimulierung von Blutplättchen spezifische Aktivierungen von zelllinien-spezifischen Kinasen auslöst, und wie behandlungsresistente Hautkrebszellen im Vergleich zu behandelbaren Hautkrebszellen gezielt ihre Signalkaskaden umorganisieren. Diese Information ermöglichte es uns, potenzielle Kandidaten für alternative Medikamente zu erkennen, z. B. die Kinase ERK4. Die beispiellose Sensitivität unserer Analysemethode konnte des Weiteren durch die Detektion einer Phosphorylierung am Serin 161 der Kinase RIPK1 hervorgehoben werden, nachdem die Zellen mit dem Protein TNFα stimuliert wurden. Die Detektion dieser Phosphorylierung mittels Massenspektrometrie gelang bisher nur unter Verwendung von Protein-Expressionssystemen in Kombination mit Immunopräzipitierung, *in vitro* Kinase Reaktionen und Phosphopeptidanreicherung. Damit zeigt unsere Studie die erste Detektion dieses funktionell charakterisierten Aktivierungsmechanismus einer für die Krebsforschung hochrelevanten Kinase direkt aus einem Zelllysat.

Kapitel 4 beschreibt Anwendungen von SRM, DIA und MRM^3 um Veränderungen in der PI3K-mTOR/MAPK Signalkaskade in zwei Lungenkrebszelllinien zu studieren. Eine der beiden Zelllinien spricht auf Behandlung mit EGFR Inhibitoren (EGFRi) an, während die andere eine Resistenz dagegen entwickelt hat. Mit dem Ziel neue Einsichten in die Charakteristika der drei genannten MS-Methoden und in den molekularen Prozess von Behandlungsresistenzen in Lungenkrebs zu gewinnen, quantifizierten wir 42 Phosphorylierungen in der PI3K-mTOR/MAPK Signalkaskade. Wie bereits früher gezeigt, konnten wir beobachten, dass DIA weniger sensitiv ist als SRM. Allerdings konnten wir auch zeigen, wie die erhöhte Flexibilität von DIA im Vergleich zu SRM ausgenutzt werden kann um spezifische Phosphorylierungen mittels Peptiden zu quantifizieren, denen bestimmte enzymatische Schnittstellen fehlen. Zusätzlich konnten wir den Zusatznutzen von MRM^3 aufzeigen, indem wir zwei zusätzliche, schwer zu analysierende Phosphorylierungspositionen quantifizierten. Im Vergleich zur klassischen Phosphoproteomanalyse, welche darunter leidet, dass die grosse Mehrheit der detektierten Phosphorylierungen funktionell nicht

charakterisiert sind, nutzten wir die SRM Technologie um gezielt Phosphorylierungspositionen zu quantifizieren, von denen bekannt ist, dass sie biologisch relevant sind. Die anvisierten Phosphorylierungen beinhalteten Phosphorylierungen, welche die Aktivität von Kinasen induzieren (z. B. PDK1, BRAF), Phosphorylierungen welche darüber entscheiden welche Bindungspartner ein Protein besitzt (z. B. 4EBP1, AKT1S1) und Phosphorylierungen welche dazu führen, dass ein Protein degradiert wird (z. B. PIK3C2A). Gesamthaft gesehen können wir in dieser Studie zeigen, dass EGFRi resistente Zellen vermehrt auf die MAPK Signalkaskade setzen um ihre Überlebenssignale aufrecht zu erhalten, während EGFRi sensitive Zellen dies über erhöhte mTORC1-Levels erreichen.

Kapitel 5 fokussiert auf DIA-Methodenentwicklung und -charakterisierung in Bezug auf die Analyse von Proteinphosphorylierungen. Primärer Fokus ist der Versuch, DIA mit Elektronen-basierten Fragmentierungsmethoden zu kombinieren und zu untersuchen, wie dies die Datenanalyse beeinflusst. Grösster Hinderungsgrund von DIA in Kombination mit Elektronen-basierter Fragmentierung ist der hohe Auslastungsgrad des MS-Gerätes. Dies rührt daher, dass Messungen, welche Elektronen-basierte Fragmentierungsmethoden anwenden, länger dauern. Dieser Effekt ist für DIA Methoden, welche primär auf Messgeschwindigkeit setzen, generell unerwünscht. Im Verlauf des Kapitels zeigen wir verschiedene Wege dieses Problem zu umgehen. Unter Anderem beinhaltet dies die Verwendung von breiteren Isolationsfenstern, welche allerdings zu einer reduzierten Selektivität führen. Dies veranlasste uns dazu, den Prozess der Datenanalyse mittels mProphet genauer unter die Lupe zu nehmen. In einem dynamischen Machine-Learning Prozess optimiert mProphet die Datenanalyse, indem es unterschiedliche Charakteristika des Datensatzes unterschiedlich gewichtet. Wir zeigten auf, wie mit Hilfe der dynamisch gewichteten Bewertungskriterien die Charakteristika von LC und MS1 – wie zum Beispiel die Akkuranz der Retentionszeit und die Isotopen Verteilung des Peptidsignals – eine höhere Gewichtung erhalten sobald die Qualität des Fragmentspektrums abnimmt. Zusätzlich konnten wir zeigen, wie die Verwendung von EThcD anstelle von HCD zu einem Sensitivitätsverlust führt, da aus derselben Anzahl von Vorläufer-Ionen eine doppelte Anzahl an Fragment-Ionen gebildet wird. Zuletzt nutzten wir synthetische Phosphopeptide um aufzuzeigen, welches Potenzial wir in DIA sehen, wenn es um die Analyse von Phospho-Isomeren geht. Dies sind Peptide, welche die gleiche Sequenz und die gleiche Anzahl Phosphorylierungen aufweisen, sich allerdings in der genauen Position dieser Phosphorylierungen unterscheiden. Allerdings zeigten wir auch, wie die schwerwiegenden Nachteile der klassischen Proteomanalyse für die Quantifizierung von Phospho-Isomeren direkt auf die DIA-Datenanalyse transferiert wird, sobald mit Spektren-Bibliotheken gearbeitet wird.

Die Dissertation endet mit einem Ausblick in **Kapitel 6,** in welchem ich mich an eine Vorhersage über die Zukunft der Massenspektrometrie und der Proteomanalyse wage und in dem ich aufzeige, welche Herausforderungen in den nächsten Jahren angegangen werden müssen. Im ersten Teil diskutiere ich meine Vision über die Zukunft von DIA, welche

zurzeit hauptsächlich darunter leidet, dass die Selektivität auf MS1 Ebene nicht sehr hoch ist und dass die Datenanalyse an Spektren-Bibliotheken geknüpft ist. Um ein weitverbreitetes Werkzeug der Zukunft zu werden, müssen diese Nachteile ausgemerzt werden. In meinem Ausblick stelle ich einige kürzlich publizierte Studien vor, die sich einzelner dieser Probleme annehmen und meiner Meinung nach interessante Lösungsvorschläge anbieten. Ausserdem werden auch bisher ungelöste Probleme diskutiert. Im zweiten Teil meines Ausblicks zeige ich auf, welche Schwierigkeiten es mit sich bringt, die rein quantitativen Daten aus einem MS-Experiment in eine biologische Information zu übersetzen, welche klinische Relevanz besitzt, z. B. für eine Krankheitsdiagnose oder für eine medikamentöse Behandlung. Auch in dieser Diskussion zeige ich einige jüngste Entwicklungen auf, welche es erlauben ein quantitatives MS Resultat direkt in Informationen über Proteinstruktur, Protein-Protein-Interaktion oder enzymatische Aktivität zu übersetzen.

Nederlandse Samenvatting

Dit proefschrift beschrijft nieuwe methoden en applicaties van data-onafhankelijke en doelgerichte massa spectrometrie. Er wordt gebruik gemaakt van Selectieve Reactie Monitoring (SRM) en data-onafhankelijke acquisitie (DIA) als belangrijkste applicaties om sensitieve en robuuste analyses te ontwerpen met als doel moleculaire signalering te kwantificeren.

In **hoofdstuk 1** wordt een algemene introductie gegeven over de methoden en instrumenten die worden gebruikt in dit proefschrift. Het hoofdstuk start met een nauwgezette uiteenzetting omtrent onderzoek gebaseerd op vloeistofchromatografie (LC) en massa spectrometrie (MS). Er wordt gedetailleerde informatie verstrekt over verschillende soorten chromatografie en MS instrumenten, gevolgd door een overzicht van verschillende methoden van data acquisitie. Vervolgens wordt de functie van eiwit fosforylering in cellulaire signalering uitgebreid besproken. Dit bevat onder anderen een detailleerde beschrijving van de functie en structuur van eiwit kinases en hun invloed in een cellulaire context. Verder wordt er aandacht besteed aan LC-MS fosfo-proteomics, waarin wordt gefocust op studies die succesvol SRM en DIA hebben toegepast om eiwit fosforylering te analyseren. Als laatste wordt de biologische rol van neuropeptiden en de analyse door middel van LC-MS besproken.

Hoofdstuk 2 is gefocust op het gebruik van SRM om neuropeptiden uit rat hypothalami te analyseren. Kennis omtrent neuropeptide regulatie is belangrijk voor het begrijpen van verschillende functionele processen, zoals energie balans, drugs verslaving, enz. De biosynthese van neuropeptiden beslaat meerdere bewerking stappen en post-translationele modificaties (PTMs), waardoor het voorspellen van het actieve peptide vanuit de genoom sequentie of het intacte eiwit bemoeilijkt wordt. Om deze reden zijn MS gebaseerde analyses van ongedigesteerde endogene peptiden nodig om de biologisch actieve moleculen te bepalen. Echter, de chemische karaktereigenschappen van endogene peptiden zorgen voor een substantiële uitdaging in het gebruik van SRM, omdat zij niet altijd samengaan met de optimale ionisatie en fragmentatie afwegingen betreffende een triple quadrupole instrument. Dit hoofdstuk beschrijft een proof-of-principle studie waarin door middel van het bestuderen van 15 neuropeptiden de potentie van SRM wordt aangetoond om lange en zeer basische endogene peptiden te kunnen kwantificeren. Ondanks gangbare praktijken is er aangetoond dat peptide en fragment ladingen tot en met 5 en 3, respectievelijk, de kwaliteit van de SRM piek groep niet limiteren en dat ze gebruikt kunnen worden voor accurate kwantificatie. Echter, zorgvuldige parameter optimalisatie is nodig voor sensitiviteit, vooral voor waardes van transitie specifieke collisie energieën, aangezien de default parameters geoptimaliseerd zijn voor met trypsine gedigesteerde eiwitten. Met deze methode om neuropeptide te kwantificeren in hypothalimi van ratten

onderhevig aan verschillende diëten zijn er significante verschillen gevonden, met name een vermeerdering van VGF afhankelijke signaal peptide AQEE-30 na een calorierijk dieet.

Hoofdstuk 3 beschrijft een nieuwe methode om op een systeem-brede schaal kinase activatie status te bepalen, gebaseerd op directe kwantificatie van fosforylering van de t-loop regio. Dit is zeer lastig in shotgun proteomic omdat de t-loop fosforylering erg ondervertegenwoordigd is door zijn lage abundantie. De beschreven methode gebruikt de on-parallelle gevoeligheid van SRM in combinatie met zeer selectieve en gevoelige fosfopeptide verrijking door middel van Fe(III)-IMAC. Zorgvuldige optimalisatie van de SRM setup heeft het mogelijk gemaakt om assays te ontwikkelen voor 248 peptiden, die samen 221 fosforylatie sites beslaan in de t-loop van 178 kinases. Onder de bestudeerde kinases waren er meerdere klinisch zeer relevant met verschillende remmers die zijn goed gekeurd door de FDA, waaronder MET, ABL, SRC, BTK, JAK3 en KIT. Echter, globale analyse van t-loop fosforyleringen heeft een robuuste hoeveelheid transities per peptide nodig om zo de hoogste gevoeligheid en selectiviteit te behalen voor de meest accurate fosfosite lokalisatie. Om een brede analyse te combineren met een lage hoeveelheid materiaal, zoals in het geval van primaire cellen, is een twee-fase strategie ontwikkeld. De strategie bestaat uit een “lage-overtuigende” onderzoek fase en een “hoge-overtuigende” kwantificatie fase voor de initiële geobserveerde peptiden. Door deze assay toe te passen op verschillende cellijnen, patiënt afkomstig melanoom cellen en primaire bloed plaatjes is het gelukt om de activatie van 54 kinases door middel van 66 t-loop fosfosites te analyseren. Stimulatie van bloed plaatjes demonstreerde activatie van cel specifieke kinases, en het karakteriseren van geactiveerde kinases in zowel drug resistente als sensitieve melanoom cellen liet een verandering van de kanker signalering omgeving zien, met als hoogtepunt nieuwe geneesmiddel doelwitten, zoals ERK4. De enorme gevoeligheid van de assay is benadrukt door de reproduceerbare detectie van RIPK1 fosforylering op serine 161 na TNFα stimulatie, welke hiervoor slechts was laten zien door middel van eiwit expressie systemen in combinatie met een IP, een in vitro kinase assay en fosfopeptiden verrijking.

In **hoofdstuk 4** worden SRM, DIA en MRM gebruikt om veranderingen in de Pi3K-mTOR/MAPK signalering te bestuderen in EGFR remmer (EGFRi) gevoelige en resistente niet-kleincellige longcarcinoom cellijnen. Om nieuwe inzichten te verkrijgen in de karaktereigenschappen van de drie MS-acquisitie methoden en in de moleculaire processen die veroorzaakt worden door geneesmiddel resistentie, hebben we een totaal van 42 Pi3K-mTOR/MAPK gerelateerde fosfosites geanalyseerd door middel van doelgerichte MS. Zoals al eerder laten zien, observeren wij dat DIA minder gevoelig is vergeleken met SRM. Echter, DIA vertoont meer flexibiliteit, wat post-analyse selectie van alternatieve fosfopeptiden mogelijk maakt. Tevens laten we de toegevoegde waarde van MRM zien door middel van de kwantificatie van twee slecht te bereiken fosfosites. In tegenstelling tot shotgun proteomics, waar van de meerderheid van de geobserveerde fosfosites de functie onbekend is, wordt SRM gebruikt om sites te bestuderen waarvan de relevantie

bekend is, zoals kinase activerende of remmende sites, of fosfosites die een rol spelen in signaal transductie. De bestudeerde sites zijn onder anderen afkomstig van verschillende eiwit kinases (o.a. PKD1, BRAF), het zijn fosforylatie sites die bepalend zijn voor eiwit binding partners (4EBP1, AKT1S1) en fosforylatie sites die eiwit afbraak stimuleren (PIK3C2A). Globaal laten wij bewijs zien dat de EGFRi resistente cellijnen vermeerderd afhankelijk zijn van de MAPK signalering om zo overleving signalering te behouden, terwijl de EGFRi gevoelige cellijn meer mTORC1 laten zien.

Hoofdstuk 5 focust exclusief op de ontwikkeling en de karakterisering van DIA methoden met als doel het analyseren van eiwit fosforylering. De primaire focus is de poging om DIA met elektron-gedreven fragmentatie te combineren en om de effecten daarvan op de data en data analyse te onderzoeken. Een belangrijke obstakel van DIA gecombineerd met elektron-gedreven fragmentatie is de langere cyclus duur, die over het algemeen ongunstig is voor DIA analyses. Door het hoofdstuk heen laten we verschillende manieren zien hoe een langere cyclus duur potentieel omzeild zou kunnen worden. Dit leidt er met name toe dat door het gebruik van een bredere isolatie breedte er een verlies van specificiteit ontstaat. Er is uitvoerig gekeken naar de post-acquisitie data analyse uitgevoerd door mProphet. Het dynamische scorings model is gebruikt om te laten zien hoe tijdens de data analyse, wanneer de kwaliteit van de fragment spectra afneemt, het belang van MS1 en LC karakteristieken, zoals retentie tijd accuraatheid en peptide isotoop verdeling, toeneemt. Daarnaast beschrijft dit werk hoe het gebruik van EThcD de intensiteit van het DIA signaal beïnvloed en leidt tot een algehele vermindering in gevoeligheid doordat meer fragment ionen worden gemaakt van dezelfde hoeveelheid intact peptide. Als laatste zijn synthetische fosfopeptiden met verschillende fosfo-isomeren gebruikt om de potentiele voordelen te laten zien van DIA ten opzichten van shotgun proteomics voor het accuraat kwantificeren van verschillende fosfo-isomeren. Echter hebben we ook laten zien dat de gebreken van shotgun proteomics, betreffende de kwantificatie van fosfo-isomeer analyse, direct overdraagbaar zijn naar DIA als er gebruikt wordt gemaakt van een spectrum bibliotheek gebaseerde DIA analyse.

In **Hoofdstuk 6** geef ik mijn perspectief op de toekomst van de massa spectrometrie en proteomics, en beschrijf ik de belangrijkste uitdagingen die aangepakt moeten worden. In het eerste deel bediscussieer ik mijn voorspelling over de toekomst van DIA, welke in de huidige situatie vooral te lijden heeft van gebreken op het gebied van gevoeligheid tijdens peptide ion isolatie en de noodzaak van spectrum bibliotheek voor de data analyse. Het verbeteren van deze gebreken is cruciaal voor DIA om alomvattend geaccepteerd te worden in de toekomst. Ik bediscussieer actuele en veelbelovende studies die deze richting op proberen te gaan en nog onopgeloste knelpunten geassocieerd met DIA. In het tweede deel licht ik veelvoorkomende problemen toe die te maken hebben met het overdragen van puur kwantitatieve data verkregen door middel van MS-gebaseerde proteomics naar klinisch relevante biologische informatie, nuttig voor diagnose, medicijn ontwikkeling,

enz. Ik bespreek verschillende manieren om op MS gebaseerde kwantitatieve data te combineren met specifieke biologische informatie zoals structuur, eiwit-eiwit interactie en enzymatische activiteit.

List of Publications

Schmidlin, T., Garrigues, L., Lane, C. S., Mulder, T. C., van Doorn, S., Post, H., de Graaf, E. L., Lemeer, S., Heck, A. J., Altelaar, A. F., Assessment of SRM, MRM^3, and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer. Proteomics 2016, 16, 2193-2205

Schmidlin, T., Boender, A. J., Frese, C. K., Heck, A. J., Adan, R. A., Altelaar, A. F., Diet-induced neuropeptide expression: feasibility of quantifying extended and highly charged endogenous peptide sequences by selected reaction monitoring. Anal Chem 2015, 87, 9966-9973.

Daskalov, A., Gantner, M., Wälti, M. A., **Schmidlin, T.**, Chi, C. N., Wasmer, C., Schütz, A., Ceschin, J., Clavé, C., Cescau, S., Meier, B., Riek, R., Saupe, S. J., Contribution of specific residues of the β-solenoid fold to HET-s prion function, amyloid structure and stability. PLoS Pathog 2014, 10, e1004158.

In Preparation:

Schmidlin, T., van Gelder, C. A., Rontogianni, S., van den Eshof, B., Heck, A. J., van de Biggelaar, M., Altelaar, A. F., Targeted Quantification of t-Loop Phosphorylations to Monitor Kinase Activity.

Gautier, V. M., Raaijmakers, L. M., Maddalo, G., Kemper, K., Rontogianni, S., **Schmidlin, T.**, Hoekman, L., Bleijerveld, O. B., Zaal, E., Vargas Diaz, D., Fitzpatrick, M. A., Krijgsman, O., Berkers, C., Heck, A. J., Peeper, D. S., Altelaar, A. F., Integrative-omics reveal a molecular network sustaining BRAF mutant melanoma and harboring CIP2A, CDK4/6 and ASCT2 as targetable vulnerabilities.

Vlaming, H., McLean, C. M., Korthout, T., Alemdehy, M. F., Hendriks, S., Lancini, C., Palit, S., Klarenbeek, S., Molenaar, T. M., Hoekman, L., **Schmidlin, T.**, van Welsem, T., Dannenberg, J. H., Heinz Jacobs, H., van Leeuwen, F., Evolutionarily-conserved HDAC-DOT1 crosstalk generates DOT1L-dependent thymic lymphomas in HDAC1-deficient mice.

Curriculum Vitae

I was born on August 18th, 1988 in Basel, Switzerland and grew up in Bottmingen. After finishing highschool in December 2007 I moved to Zurich to start studying Biology at the *ETH Zurich*. I obtained my BSc in Biology in 2011 and decided to follow up with a Master at the same university. A practical course about plant proteomics in the lab of Katja Baerenfaller primed my interest for proteomics and analytical chemistry and I decided to deepen my knowledge in these areas throughout my MSc internships. I joined the lab of Ruedi Aebersold for a total of nine months, working on investigating kinase substrate relationships in yeast. The project entailed *in vitro* kinase reactions, shotgun proteomics, and target validation by SRM. Two shorter internships brought me to the *Laboratory of Physical Chemistry* at *ETH Zurich* and the *Proteome Exploration Laboratory* at the *California Institute of Technology.* At the *Laboratory of Physical Chemistry,* under the supervision of Roland Riek, I investigated amyloid fibril formation and stability by various analytical technologies such as UV/Vis, circular dichronism, dynamic light scattering and NMR. At the *Proteome Exploration Laboratory* under the supervision of Sonja Hess, I gained additional experience in the field of MS-based proteomics. In 2013 I graduated from *ETH Zurich* with an MSc in Biology, and shortly after joined the *Biomolecular Mass Spectrometry & Proteomics Group* at *Utrecht University*. With the goal in mind to get a well-founded knowledge in biochemical and analytical methodology – especially liquid chromatography and mass-spectrometry – I started my PhD in November 2013 under the joined supervision of Maarten Altelaar and Albert Heck. Having the opportunity to engage in cutting edge biology and LC-MS method development throughout the last 4 years enabled me to substantially grow as a scientist. The results of this work are described in this thesis.

Acknowledgments

Maarten, thanks for the supervision throughout the last years. I truly appreciated your style of supervising PhD students, which involves lots of freedom. You had many ideas about potentially interesting projects, which I could choose from. You promoted my independence as a researcher, whilst always being open to give guidance and help if I asked for it. This way of hands-off supervision was a perfect fit for me. Thanks also for enabling me to perform mainly projects in the area of LC-MS method development, whilst originally being hired to investigate drug resistance in melanoma. I truly enjoyed working with you throughout these years.

Albert, thanks for giving me the opportunity to obtain a PhD in your group. Despite being primarily a student of Maarten, I really appreciated your continued interest in my research. Your inputs and suggestions were always very valuable and helped me a lot. With your passion for research you were able to build up a world-renowned knowledge center in protein mass spectrometry here in Utrecht, and I am grateful for being given the opportunity to perform my PhD in such a place.

I still remember the first few weeks in the lab. I guess it is challenging for every new PhD student to find their own way and acquire all the required knowledge in a group of more than 60 people. I am very grateful to the people who took the time to introduce me to various aspects of the lab and the world of proteomics. First and foremost, thanks to Erik, who introduced me to the TSQ and the Proxeon, which I used for almost all of my research during the PhD. Besides that, he also showed me how he dealt with SRM data, which in combination with my background obtained during my masters largely defined my way of working to date. Nadine, you were once assigned my official lab supervisor – and you once complained that this does not involve enough work for you. Nevertheless, you soon became my first go-to person for MS related questions, thanks a lot for all these insights and discussions.

Over the years I was involved in many research projects, some of which are part of this thesis, some of which are not. All of them involved various collaboration partners without whom none of the projects would have been successful. Luc, you were key for the success of our method comparison paper (Chapter 4 of this thesis), which had to be finalized from idea to manuscript within only 3 months. Without your dedicated contribution this would have never been possible. It was a great experience to work together with you, I think we really worked well as a team and I learned a lot from you!

As I have never performed cell culture myself, I have often been dependent on obtaining cell pellet samples from other people. Celine, thanks for regularly providing me with PC9 cell pellets so that I could carry out all my method development ideas. A big thank you also goes to Charlotte and Matina, who got more involved in my kinome project by performing the biological experiments such as Jurkat cell culture, viability assays, TNFα stimulation

and Western blot validation. Thanks to Maartje and Bart who enabled me to test my kinase methods on primary cells. Furthermore, thanks to Juan for collaboration on the SWATH experiments and Violette for the collaboration on measuring mTOR signaling in melanoma cells. Thanks to Arjen, Roger and Christian for collaboration on the neuropeptide project and thanks to Cathy for collaborating on the effect of MRM^3 on phosphopeptide detection. Thanks a lot also to Wei for teaching me membrane enrichment protocols and to Jing for jointly taking care of the TSQ and mainly the Proxeon G.

Besides my supervisors Maarten and Albert many people have given me crucial inputs and ideas for my projects, some of them on a regular basis. These were mainly Simone, Wei and Kelly for general experimental design, and Thin Thin and Harm for the more technical aspects. These inputs were really appreciated.

Thanks a lot to Corine, without whom the group wouldn't be able to exist. It was a great relief to get your help for all the administrative processes involved in the process of moving to the Netherlands, organizing two contract extensions and the final administration process required for submitting my thesis. Thanks to Geert for taking care of all the ICT part in the lab, which in my case involved helping me with various MS/LC/PC connection problems and once getting rid of a rather nasty virus infection on my working PC. Thanks to Mirjam for taking care of various Lab related issues, including (but not limited to) various HpH experiments and express orders over the years (and obviously also thanks for the many joint coffee breaks). Thanks to Arjan for various technical insights/discussions about MS, vacuum pumps, ion sources, columns etc. Thanks also to the other technicians for their various contributions to the success of the lab: Soenita, Jamila, Anja, Dominique. And thanks to my various office mates over the years, Nathalie, Thin Thin, Patrick, Andrey, Domenico, Anna, Dominique and Federica and all the other current and former lab members.

I am very happy that beside work-related interactions our lab also had a well-developed social life. Many work colleagues became good friends over the years. Celine, we actually became friends before we became workmates when we met at the Brixen Summer School. The friendship we built up there while drinking wine, hiking, swimming in the pool and discovering the city center was holding firm. Sitting in the office next door it needed just a few steps to pass by for a chat or a coffee. You became my favorite coffee buddy throughout my PhD, thanks for all these nice little breaks over the years – and obviously the nice lasagna you once cooked for me. I still owe you a revenge. Fan, you too became a very important friend of mine over the years. Since I usually worked in the lab during the day and you during the night, we did not often see each other in the lab. Nonetheless, we shared many unforgettable moments during our spare time. Skiing down a black slope in Engelberg with stones peeking through the snow layer everywhere, hiking up the Koenigstuhl in Heidelberg, eating together Schlachtplatte in Darmstadt and pregnant fish in Den Haag are just a few of the highlights over the last years. Philip, you became my opera buddy over the years. I really enjoyed our joined excursions to Amsterdam or

the Tivoli. Since you left to Berlin a year ago, my classical music trips became much more boring. Michiel you were my concert buddy for my other kind of music interest, heavy metal. It was always great fun to go to a nice concert with you. It is not that often that I meet someone who shares my taste of metal music, I guess so far you came closest. It was never hard to convince you to join for a nice gig – the exception being Swans. But even there it finally worked out. Dominique, you were not only my office mate but also became a good friend. Thanks a lot for all our joined activities like chasing wild boars, swimming and hiking. Thanks also for occasionally playing the piano for me at the central station – I promised you to give you a personal concert once myself, a promise I did not keep yet, but I hope it is still going to happen before I leave the Netherlands. Franziska, we were friends before we became workmates. We shared a lot of memorable moments over the last years, an afternoon in the pool in San Antonio, discovering the Alamo, some hikes along the single in Utrecht and most importantly our very successful cinema evening. I really enjoy working in the same lab with you again more than four years later and I am looking forward to dive deeper into DIA together with you in the upcoming months. Angela, sadly you were here just for one summer, but what a time that was. You soon became my favorite Italian lab mate (sorry to all the other Italians) and our coffee breaks and thumbs-up facebook chats were always funny.

Big thanks also to the Friday borrel enthusiasts. Sometimes after a rough week, drinking some (too many) beers on a Friday together with my workmates was exactly what I needed. Thanks to everyone who cherished that tradition – especially the die-hard participants without whom that nice tradition would have likely died within the last few years: Arjan, Clement, Michiel, Philip, Anja, Max, Wei and Vojta. Big thanks also to the table football die-hards, it has always been great fun. Especially two people stand out in this respect, Sara (I always liked playing against you – it was great fun, especially when you were losing) and Piero (you were probably one of my hardest challenges over the years). But also thanks to all other regular or occasional participants: Vojta, Ele, Erik, Nicolas, Sem, Oleg, Max, Maarten, Andrea, Fabio.

When moving abroad to pursue a PhD, not only does the work environment change drastically, also the private live gets affected substantially. As much as I cherish the importance of a nice social life at the workplace, it is equally important to me to have a thriving social network outside the lab. Over the last four years, I had the pleasure to meet lots of nice people, many of which became very close friends. Thanks a lot to every one of you guys (in no particular order): Christian, Debi, Alex & Sonja, Desiree, Michael & Johanna, Simone, Nathalie, Arie & Ina, Hagar & Harm, Sarah, Dina.

Last but not least thanks to Barbara for being here for the last almost two years. I am so glad you joined our lab – it would not have been the same without you! And obviously a very big thank you to my parents, who supported me in all aspects required for being an expat and being a PhD student.