TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Proteomik und Bioanalytik

Mass spectrometry based chemical proteomics for drug selectivity profiling

Dominic F. J. G. G. Helm

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzende:	Univ.-Prof. Dr. D. Langosch
Prüfer der Dissertation:	1. Univ.-Prof. Dr. B. Küster
	2. Univ.-Prof. Dr. S. Sieber

Die Dissertation wurde am 29.04.2015 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 06.07.2015 angenommen.

This work is dedicated to the memory of my mother.

Thank you for your smiles and your love.

TABLE OF CONTENT

ABSTRACT

Mass spectrometry is the standard technology for large scale chemical proteomics. The field benefits to a large extent from new developments in technology, which enable large scale and high throughput chemical proteomics studies. The standard data acquisition strategy in proteomics is the data-dependent acquisition (DDA) approach. In order to circumvent proposed conceptual limitations of DDA, several data-independent acquisitions (DIA) methods have been established, among them the HDMS^e approach on the Waters Q-TOF Synapt G2S. In theory, DIA methods can provide highly reproducible and comprehensive analysis of samples. Based upon these presumptions, the HDMS^e method has been critically evaluated for use in bottom-up proteomics and its performance compared with that of a DDA approach using a LTQ-Orbitrap Elite. After the establishment of reliable database search parameters for DIA data, it could be concluded that the performance of HDMS^e is inferior to that of standard DDA approaches. In consequence, the LTQ-Orbitrap Elite was chosen for the MS read out in a large scale chemical proteomic study.

The performance of quadrupole time-of-flight hybrid mass spectrometers suffers from low duty cycle through the necessary use of orthogonal ion injection systems. In order to improve the performance of the DDA methodology on the Q-TOF instrument, a new ion mobility supported DDA method was developed – HD-DDA. The here described method utilises post-fragmentation ion mobility spectrometry of peptide fragment ions in conjunction with mobility time synchronized orthogonal ion injection leading to a substantially improved duty cycle. HD-DDA enabled the identification of 7,500 human proteins within one day and 8,600 phosphorylation sites within 5h of LC-MS/MS time in bottom-up proteomic experiments. This demonstrated the state-of-the art performance of this novel approach.

The previously mentioned large scale chemical proteomics study aimed to profile all protein kinase inhibitors, approved or currently in clinical development. Protein kinases are key components of signal transduction pathways and are in consequence attractive drug targets for diseases like cancer and inflammation. Comprehensive knowledge about the target space of a kinase inhibitor is crucial for a proper interpretation of its biological effects. The here presented selectivity profiling of ponatinib against 308 protein kinases, an approved chronic myeloid leukemia drug against BCR-Abl, enabled the determination of a set of new highly potent targets, e.g. ZAK and MAPK14. This data helped to reveal the observed inhibitory effect on inflammation processes by ponatinib and lead to a deeper understanding of the mode of action of this drug.

ZUSAMMENFASSUNG

Massenspektrometrie ist die Methode der Wahl für chemische Proteomik. Das Feld profitiert dabei von neuen technologischen Entwicklungen, die chemisch-proteomische Studien im Hochdurchsatz ermöglichen. Die am weitesten verbreitete Strategie für die Datenerfassung in der Massenspektrometrie-basierten Proteomik ist die Daten-abhängige Akquisition (DDA). Mit dem Ziel konzeptionelle Limitierungen des DDA-Ansatzes zu überwinden, wurden zahlreiche Daten-unabhängige Akquisitionsmethoden (DIA) entwickelt, unter anderem HDMS^e. Mittles DIA ist es theoretisch möglich Proben umfassend und mit einem hohen Maß an Reproduzierbarkeit zu analysieren. Mit dieser Premisse wurde die HDMS^e Methode für die Durchführung von bottom-up proteomischen Experimenten evaluiert und zu einem DDA Ansatz auf einer LTQ-Orbitrap Elite verglichen. Nachdem zuverlässige Datenbanksuchparameter für den DIA Ansatz etabliert wurden, konnte festgestellt werden, dass die HDMS^e Methode dem Standard DDA Ansatz unterlegen war. Foglich wurde entschieden die MS Analyse einer großangeleten Studie im Bereich chemischer Proteomik mittels LTQ-Orbitrap Elite durchzuführen.

Die Leistung von Quadrupole-Flugzeit-Massenspektrometern (Q-TOF) grundlegend unter einem ineffektiven Betriebszyklus begründet durch zwingende Verwendung von orthogonalen Ioneninjektionssystemen. Mit dem Ziel die Leistung der vorhandenen DDA Methode auf einem Q-TOF-Instrument zu verbessern, wurde eine neue Ionenmobilität unterstützte DDA Methode entwickelt – HD-DDA. Die hier beschriebene Methode nutzt Ionenmobiliätsauftrennung von Peptidfragmenten in Verbindung mit der Synchronisierung von Ionenmobilität und orthogonaler Ioneninjektion, was zu einem wesentlich effektiveren Betriebszyklus führt. Mit HD-DDA konnten 7.500 menschliche Proteine innerhalb eines Tages und 8.600 Phosphorylierungsstellen innerhalb von 5 Std. LC-MS/MS Zeit in einem bottom-up-Proteomik-Experiment identifiziert werden. Damit war diese neuartige Methode auf Augenhöhe mit dem aktullem Stand der Technik.

Die bereits erwähnte großangelegte Studie im Bereich der chemischen Proteomik hat das Ziel 200 Proteinkinase-Inhibitoren, die schon zugelassen sind oder sich noch in der klinischen Entwicklung befinden, zu untersuchen. Proteinkinasen sind wichtige Bestandteile in Signaltransduktionswegen und sind folglich attraktive Angriffspunkte für Arzneimittel gegen Krebs- oder Entzündungserkrankungen. Das Wissen über die Zielkinasen eines Kinase-Inhibitors, ist von entscheidender Bedeutung für eine möglichst korrekte Interpretation dessen biologischer Wirkungen. Die hier vorgestellte Selektivitätsuntersuchung (mit 308 enthaltenen Proteinkinasen) von Ponatinib, einem zugelassenem BCR-ABL Inhibitor bei chronischer myeloischer Leukaemie ermöglichte die Identifizierung einer Reihe von neuen hochpotenten Zielkinasen, u.a. ZAK und MAPK14. Diese Daten konnten dazu beitragen, die bei Ponatinibanwenung beobachtete hemmende Wirkung auf den Entzündungsprozess zu erklären und führten zu einem tieferen Verständnis über die Wirkungsweise dieses Medikaments.

CHAPTER I

GENERAL INTRODUCTION

MASS SPECTROMETRY BASED PROTEOMICS

PROTEOMICS

When the Human Genome Sequencing Consortium published the map of the human genome in 2004, one interesting observation was the unexpectedly low number of proposed human genes (approx. 20.000)[1]. This indicated that proteins and their modifications are key to the complexity of an organism. The catalogue of proteins in a tissue, biofluid or whole organism comprises its proteome.

The scientific field of proteomics[2] investigates the expression, structure and function of proteins. Whereas the genome can be viewed as the *in vivo* blueprint of an organism, the proteome is the result of the interpretation of the genome at a certain time point and/or under certain conditions, hence representing the biological status of the organism. The complexity of the proteome results from, on the one hand, from influences on the protein translation level such as alternative splicing, processing and post-translational modifications (PTMs)[3] of the protein itself. On the other hand, complexity can arise from more global regulatory levels such as the dynamic of gene expression, subcellular localization, protein-protein interaction, and protein stability[4, 5]. In the past decades, the field of proteomics in combination with genomics and transcriptomics has enabled us to paint an increasingly holistic picture of biological processes. This has been possible due to formidable advances in proteomics over the past 40 years: Starting with the use of two-dimensional gels[6, 7] via the identification and quantification of >10,000 proteins from human cell lines or tissues[8-13] and peaking in the first draft of the human proteome defined by mass spectrometry[14, 15].

The field of proteomics is still greatly influenced by advances and developments of new mass spectrometry technologies, however, with the use of high throughput chemical proteomics screening approaches, the field has already had great impact on clinical research and drug development.

MASS SPECTROMETRY

The fundamental task of a mass spectrometer (MS) is the determination of the mass-to-charge ratio of ions. Thereby a mass spectrometer always consists of three main components: An ion source to charge the analyte molecules. The second part is the mass analyser, where the ions are separated according to their mass-to-charge ratio (*m/z*) before moving to the third part, the detector. Here the abundance of each *m/z* value is recorded. This information is instantaneously translated by the data system into a mass spectrum. In the field of proteomic research two ion ionisation methods are principally applied since they are considered as "soft ionisation" techniques: electrospray ionisation (ESI)[16] and matrix assisted laser desorption/ionisation (MALDI)[17, 18].

IONISATION TECHNIQUES

MATRIX-ASSISTED LASER DESORPTION/IONIZATION (MALDI)

MALDI produces gas phase ions from a solid substrate[17, 18]. Prior to the ionisation process the analyte has to be embedded into a UV-absorbing matrix. The ionisation process is initiated by a short and energetic laser beam. The MALDI ionisation process is not fully mechanistically explained[19], however, the laser energy is predominantly absorbed by the π-electrons of the matrix molecules (commonly small aromatic acids) leading to the desorption of the matrix in conjunction with the nested analyte molecules in a "soft" solid-gas-phase transfer. Thermal decomposition of the analyte molecules is prevented, since the molecules cool as they expand into the vacuum. Due to this, MALDI is suitable for the ionisation of large molecules e.g. intact proteins although MALDI usually produces singly charged ions which can challenge the *m/z* range of a mass spectrometer. MALDI is commonly used for the analysis of low complexity samples.

ELECTROSPRAY IONIZATION (ESI)

ESI provides an soft analyte transportation route from solution into the gas phase[16]. The properties of an ESI ion source enable the on-line coupling of liquid chromatography to a mass spectrometer and made this set-up the most commonly used approach in proteomics[20].

In the initial step of the ionisation process, the analyte solution passes through a thin conductive capillary to which a voltage of several kV is applied. The influence of the electric field on the liquid on the tip of the capillary in conjunction with the surface tension

causing the formation of a Taylor cone. The potential difference between the capillary and the counter electrode pulls the positive ions to the surface of the solvent. Once the electric field overcomes the surface tension, the Taylor cone emits a solvent jet. The tip of the jet is inherently unstable and collapses into charged droplets. Solvent evaporates from these droplets until the charge density on the surface of the droplets reaches the Rayleigh limit. Once this limit is passed, the Coulombic repulsion force is greater than the surface tension causing the droplet to undergo fission into smaller droplets, the so called Coulombic explosion (Fig. 1).

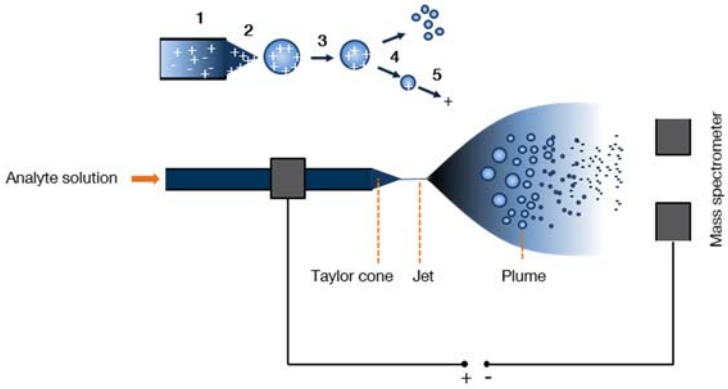


Figure 1 Electrospray ionisation process (1) Electrophoretic separation of anions and cations, (2) Formation of the Taylor cone, (3) Solvent evaporation, (4) Coulombic explosion and formation of single solvated analyte ions (CRM or IEM) and (5) declustering of solvated ions (adapted from Hahne, 2012[21]).

This procedure is repeated until only a single analyte ion remains, according to the charge residue model (CRM)[22]. An alternative explanation to the CRM is proposed by the ion emission model (IEM)[23]. Here the ions are constantly ejected from the droplet, picking up their charge from the surface in order to conserve the field strength on the droplet's surface. For unfolded intact proteins a further mechanism, the chain ejection model (CEM) is proposed[24]. The different models are summarised in Fig. 2.

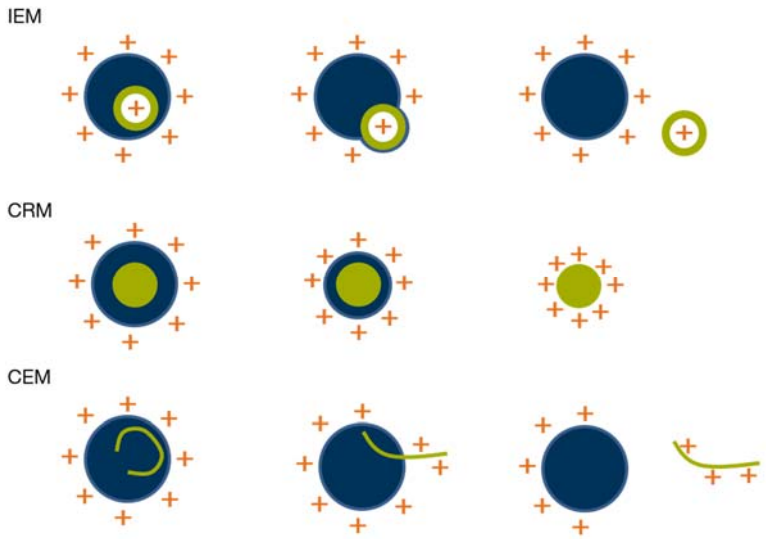


Figure 2 Electrospray mechanisms (IEM) Ions are constantly ejected from the droplet, picking up their charge from the surface in order to conserve the field strength on the droplet's surface (CRM) Solvent constantly evaporates from the droplets until a single analyte ion remains (CEM) A unfolded protein is ejected from the droplet carrying its charges.(adapted from Konermann et al.,2013[24])

Different to MALDI, ESI principally generates multiply charged ions. The efficiency of the ESI process can be optimised by minimising emitter tip dimensions and by the use of low flow rates in the range of several hundred nano-litres per minute - nanoESI[25]. Further, the addition of low percentages of DMSO to the analyte solution causes an extra enhancement in ionisation efficiency[26]. Increases in efficiency of nanoESI are accompanied by a higher analytical sensitivity of the instrumental set-up, an effect which is attributed to the several orders of magnitude resulting in smaller droplets carrying a lower number of analyte molecules emitted from the Taylor cone[25]. It was argued that a fully optimised nanoESI system can achieve an ionisation efficiency of almost 100%.

ION MOBILITY MASS SPECTROMETRY

KEY CONCEPTS OF ION MOBILITY SPECTROMETRY

Ion mobility spectrometry (IMS) is a gas phase ion separation technique. The separation is based on the inherent properties of an ion such are mass (m), charge (z) and the mobility of an ion in the gas phase (K).

In conventional IMS the drift time metric is the recorded time required for an ion to travel a set distance through an inert buffer gas (e.g. N_2) under a low electric field. This process is also described as directed diffusion[27]. During this process, two complementary forces influence the velocity (v_d) of an ion; the ion is accelerated through the drift tube by the electric field whilst it is decelerated by collisions with buffer gas molecules. The mobility of an ion (K) can be defined by the velocity (v_d) and the electric field (E)[28] (Form. 1).

$$v_d = KE \qquad \text{Formula 1}$$

The simplest configuration of an instrument for determining an ion's mobility is composed out of a buffer gas filled tube within a vacuum system plus an applied electric field across the length of the tube[29]. The ions enter the tube and drift through the buffer gas constantly pulled by the electric field. The mobility of an ion (K) becomes directly proportional to the electric field and to the ion's collision cross section (Ω, Form. 2).

$$K_0 = \left(\frac{3ze}{16N}\right)\sqrt{\left(\frac{2\pi}{\mu k_B T}\right)}\left(\frac{1}{\Omega}\right) \qquad \text{Formula 2}$$

Here K_0 is the reduced mobility (measured mobility K, standardised for 273.15 K and 1 bar), z is the charge of the ion, e is the elementary charge, N the gas number density, μ the reduced mass of the ion-neutral pair, k_B the Boltzmann's constant and T is the temperature of the gas[28, 30]. The collision cross section (CCS, Ω) of an ion is directly related to its physical shape in the gas phase and can provide useful information e.g. in the field of structural biology. The CCS can be directly related to crystal structures or computational models[31-35]. In proteomics, two major types of ion mobility devices are currently commercially available and in use: the Travelling Wave Ion Mobility Spectrometry (TWIMS) and the Field Asymmetric waveform Ion Mobility Spectrometry (FAIMS).

FIELD ASYMMETRIC WAVEFORM ION MOBILITY SPECTROMETRY

Field asymmetric waveform ion mobility spectrometry (FAIMS) was developed in the beginning of the 1980's in the former USSR[36]. FAIMS devices coupled to ESI-MS instruments are set up in the front of the instrument[30, 37]. During FAIMS, the applied electric field (here: dispersion field, DF) is vertical to the travel direction of the ions. The DF is applied across the space between two electrodes. The ions traverse this space driven by the gas flow. The applied DF attracts the ions towards one of the two electrodes. In order to avoid an ion-electrode contact and to ensure continuing ion flow, a compensation field (CF) is applied. Through the CF it is possible to use the FAIMS device as a mass filter, since every CF is corresponding to an ion population with a specific mobility, by the cost of sensitivity compared to other IMS set-ups. However, during a standard experiment the CF is altering through a certain range to allow a set of ions to pass. The resolution of a FAIMS device can reach almost 500[38]. FAIMS recently was applied in several studies for the separation of complex peptide mixtures[39, 40] or for the resolution of positional isomers of peptides carrying different modifications[41-45].

TRAVELLING WAVE ION MOBILITY

The basic assembly of a travelling wave ion mobility device is a stacked ring ion guide (SRIG), as illustrated in Fig. 3A.

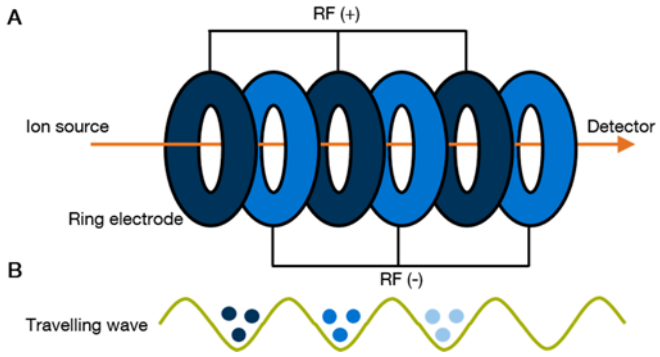


Figure 3 Stacked ring ion guide (SRIG) with travelling wave potential (A) Schematic illustration of a SRIG with applied RF voltage. (B) Travelling wave and the "surfing" of the ions. (adapted from Giles et al.,2004[46])

A SRIG consists of sequentially arranged ring electrodes with opposite phases of radio-frequency (RF) applied to neighbouring rings[46]. Using a RF-only device like SRIG in a mass spectrometer is limited by prolonged travelling times of the ions throughout the ion mobility cell[47]. One approach to enhance the travelling speed of the ions in a SRIG is the use of travelling voltage wave[46, 48]. The travelling voltage wave is superimposed on the RF voltage and allows the ions to surf the voltage wave (Fig. 3B). The wave consists of a direct current (DC) potential applied for a given time to an electrode before it switches

further to an adjacent electrode and moving along the SRIG. The ions are moved by the DC wave through the IMS device and are retarded by the buffer gas, enabling the ion mobility separation[48, 49]. Ions with a high mobility can easily keep up with the travelling wave whilst low mobility ions get passed by the waves more often[46]. Travelling wave ion mobility separation (TWIMS) devices reach resolutions of 45 and are one key part of the Synapt line-up of mass spectrometers from Waters Corporation.

MASS ANALYSIS

QUADRUPOLE

A quadrupole mass analyser (Fig. 4) consists of four parallel rods placed around a common axis. A RF voltage is applied to the opposing electrically paired rods and combined with a DC voltage in order to create the electric field in the quadrupole. The quickly changing RF induces an oscillation motion of the ions whilst the DC causes an *m/z* dependent deflection. This combination leads to three possible trajectories depending on the *m/z* value of the ion. First, the amplitude of the RF induced oscillating movement is too high, leading to the loss of the ion. Second, the ions can collide with a rod by deflection. Third, the ion is on a stable trajectory throughout the quadrupole. Hence, the quadrupole can be used for selection of a distinct *m/z* value or, by rapid variation of the amplitude of the electric field, guiding ions of a broad *m/z* range through it.

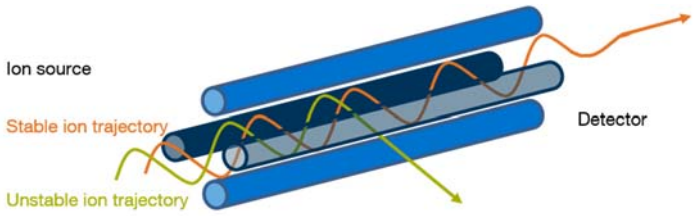


Figure 4 Quadrupole mass analyser The quadrupole consists of four parallel rods imposing a quadrupole field. This can either cause stable (orange) or unstable ion trajectories (green).

TWO- AND THREE-DIMENSIONAL ION TRAPS

A two dimensional trap (linear ion trap) is in principle composed of a quadrupole furnished with two additional electrodes on both ends, enabling the trapping of ions inside the device for a certain duration[50, 51]. Variation of the applied voltages induces an *m/z* dependent ejection of ions towards the surrounding detector. A three dimensional ion trap follows the same idea, however the linear rods have been replaced by a ring electrode and the applied electric field is inducing a parabolic potential for keeping the

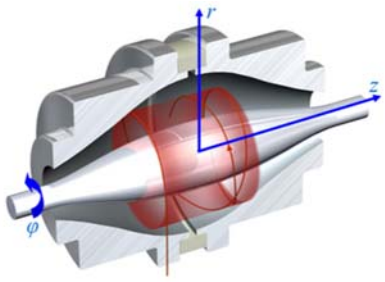
ions inside the trap. Similar to the quadrupole the 2D/3D traps can be used to filter ions for a distinct *m/z* value and trap them. This can be achieved by ejecting all ions other than those of the desired *m/z* value. The advantage of linear ion traps (2D) is the higher capacity for ion storage. This leads to an improved sensitivity and dynamic range compared to 3D traps[52, 53].

FOURIER TRANSFORM ION CYCLOTRON

The mass detection in a Fourier transform ion cyclotron mass spectrometer (FT-ICR) is performed using a strong homogeneous magnetic field (e.g. 15 Tesla) in which the ions are cycled[54]. The frequency is inversely proportional to the *m/z* of the ion and directly proportional to the strength of the magnetic field. A charged cycling ion in a magnetic field is inducing an image current in the detector plates. For multiple ions cycling through the magnetic field, the image current is deconvoluted by Fourier transformation into a mass spectrum. When the ions enter the ICR cell the oscillation starts in the centre of the cell. The ions are moved by RF pulses to higher and larger orbits bringing them closer to the detector plates. Therefore the applied RF pulse and the cyclotron frequency of the ion need to match ("resonance extinction"). Similar to the 2D/3D traps and the quadrupole it is possible to selectively excite an m/z value or do this for a broader mass range.

ORBITRAP

The Orbitrap[55, 56] mass analyser (Fig. 5) consists of an outer barrel shaped and an inner spindle shaped electrode. The created electric field of both electrodes traps the ions inside the analyser. The ions oscillate between both electrodes in the x-, y- and z-axis. The oscillation around the z-axis is here dependent on the *m/z* value of the ion and recorded as image current. The image current can be transformed in a mass spectrum by FT. The principle behind the Orbitrap is the Kingdon trap from 1923[57]. Unfortunately, it took until the invention of the C-trap[58] in 2006, for controlled ion injection, until this analyser type could be used. Currently, the Orbitrap is the most widespread mass analyser in proteomics since it combines very high mass accuracy, high resolving power with high acquisition speed[59-61].



$$\omega_z = \sqrt{\frac{k}{m/z}}$$

Figure 5 Orbitrap mass analyser[62] Ion rotate around the inner spindle electrode (ω_φ) and show radial oscillation (ω_r) and axial oscillation (ω_z). The axial oscillation (ω_z) is only dependent on the *m/z* value.

TIME-OF-FLIGHT

The concept of the time-of-flight (TOF) mass analyser can be illustrated by the example of a football player kicking two unevenly heavy balls against a wall (in a vacuum, Fig. 6A). Assuming both balls are kicked with the same force, their time until they hit the wall only depends on their mass. In a TOF analyser ions (ball) are accelerated by an electric field (football player) towards a detector (wall). After passing the electric field (U) the ions possess the same kinetic energy E_{kin} (Formula 3) and enter the field free vacuum tube.

$$E_U = E_{kin} = Uzq = \frac{1}{2}mv^2 \qquad \text{Formula 3}$$

The velocity v can be defined as the time-of-flight t the ion needs to overcome the field-free region in relation to the length L of the drift distance. Further Formula 3 can be transformed to *m/z* (Form. 4):

$$\frac{m}{z} = \frac{2qU}{L^2}t^2 \qquad \text{Formula 4}$$

Summarizing all the set-up based constants (to C) in Formula 4, this shows that t only depends on the mass of the analyte ion (Form. 5).

$$t = C\sqrt{\frac{m}{z}} \qquad \text{Formula 5}$$

In consequence, differences in the kinetic energy of isobaric ions directly translates into a loss in resolution. For resolution enhancement two main strategies exist for TOF instruments, the delayed pulse extraction and the use of reflectrons.

During delayed extraction[63] ions are allowed to travel (for ns to µs) into a field free region with ions having higher E_{kin} travelling more far. The extraction pulse is then applied and induces more energy to ions that had less E_{kin}. Delayed extraction is used in combination with MALDI sources[64, 65].

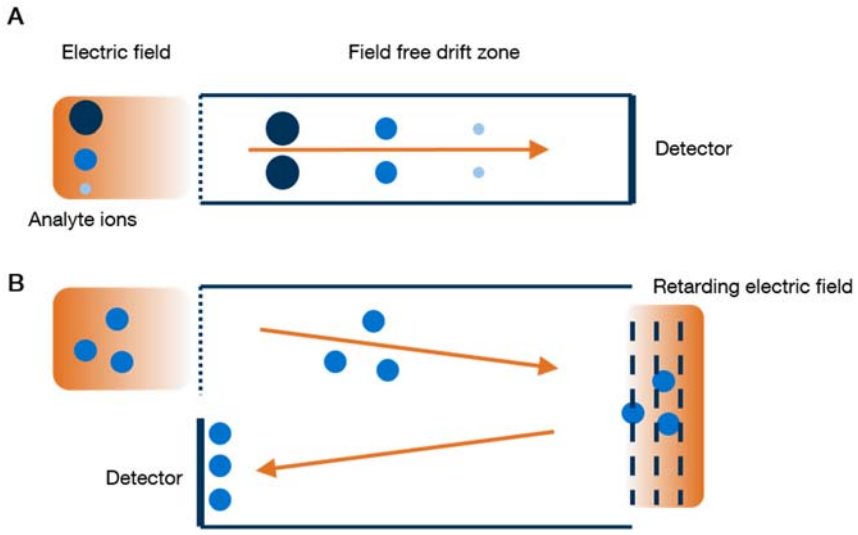


Figure 6 Illustration of the TOF mass analyser and a reflectron (A) Representation of a TOF mass analyser. After the acceleration of the ions in an electric field, the ions enter a field free drift tube (drift zone). Since all ions are equipped with the same potential energy, the velocity only depends on the *m/z* value of an ion (adapted from Maier, 2014[66]). (B) A reflectron consists of a series of ring electrodes leading to the creating a constant retarding electric field. By this the differences in initial kinetic energy of ions of identical *m/z* are corrected.

A reflectron (Fig. 6B) consists of a series of ring electrodes leading to the creating of a cavity shaped constant electric field at the end of the flight tube. Ions entering the electric field are decelerated and accelerated back into the flight tube towards the detector. Ions having a higher kinetic energy will travel farer into the reflectron than those with less initial kinetic energy. In this way the difference in initial kinetic energy of ions of identical *m/z* is corrected.

THE DIFFERENT MASS ANALYSERS – A SYNOPSIS

For mass analysis many different types of analysers have been developed over time. Each of it has its distinct characteristics in terms of resolution, mass accuracy, mass/charge range, scan speed, dynamic range and sensitivity (Table 1). The mass analysers used in proteomics, and introduced in this work can be assigned two major groups: the ion beam (Quardrupole and TOF) and the trapping (2D/3D traps, Orbitrap and FT-ICR) mass analysers. The two most widespread mass analysers in discovery proteomics are the TOF and the Orbitrap analyser. The latter has become the *de facto* workhorse in proteomics[62, 67].

Table 1 Performance characteristics of mass analysers (adapted from Han et al. ,2008[68] and Yates et al., 2009 [69])

	Quadrupole	2D/3D ion trap	FTICR	Orbitrap	TOF
Resolution	200-2,000	20-20,000	> 750,000	> 200,000	> 30,000
Mass accuracy	100-1,000 ppm	100-1,000 ppm	1-2 ppm	1-2 ppm	< 5 ppm
m/z range	< 4,000	< 4,000	< 2,000	< 2,000	> 500,000
Scan speed	1 Hz	5-50 Hz	5-50 Hz	5-50 Hz	MHz
Dynamic range	$1:10^4$	$1:10^3$	$1:10^3$	$1:10^4$	$1:10^3$
Sensitivity	+/++	++/+++	+++	+++	+++

One main feature of a mass analyser is the resolution (Fig. 7). Higher resolution allows distinction between masses that only possess incremental differences[70].

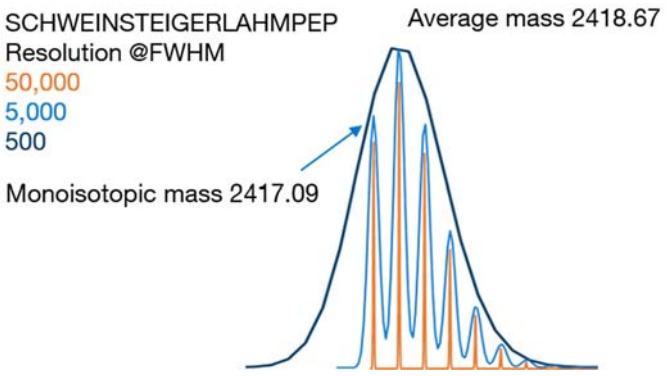


Figure 7 Resolution calculated peak shapes for a peptide at different resolutions Calculated with Xcalibur, adapted Maier, 2014[66]

High resolution mass spectrometers are able to resolve the single isotopic peaks of an *m/z*-value, which enables the determination of the charge state and the calculation of the monoisotopic mass. Although, almost all peptide precursors, relevant in bottom-up proteomics, can be resolved already at 15,000 resolution, a high mass resolution is beneficial since it directly translates into higher mass accuracy[71] and improves data base searching[72]. For TOF instruments, the resolution is independent of the detection time. In contrast, the Orbitrap reaches higher resolution with higher scanning time. Usually, the available resolution power of TOF instruments is several times lower than the highest available resolution on an Orbitrap instrument[67].

Another important feature of mass analysers is the scan speed and the involved sensitivity. Currently, mass spectrometers have limits of detection (LOD) in the amol range. Thus they would be able to detect a sugar cube (3 g) in the Bodensee (4.8e13 l) from just 1 ml sample. Throughout the very high complexity of bottom-up proteomic samples, vendors have developed consistently faster scanning instruments[11, 60, 73-76]. This improvement allows ever deeper proteomic coverage in even smaller amounts of time. Due to its ability to even detect a single ion, the TOF mass analyser is, in theory, the fastest scanning and most sensitive detector. The FT based analysers, in comparison, require several charges until an image current can be detected. However, in reality this advantages of the TOF analyser are compromised by its susceptibility to noise deriving from chemical background, recorded with the same sensitivity[77]. This demands a higher ion flux on the detector by e.g. longer scan times, counteracting the inherent beneficial of this analyser type. On the other hand, the ion detection in the Orbitrap is FT based; the requirements for peak detection (e.g. stability during the analysis) almost by default exclude the detection of noise and allowing lower detection limits for FT instruments compared to TOFs[67].

HYBRID MASS SPECTROMETERS

A hybrid mass spectrometer is a combination of two or more mass analysers in order to combine the strengths of different techniques. A popular combination in quantitative proteomics is the triple quadrupole (QQQ) allowing an ultra-high sensitivity for quantification (low amol range[78, 79]) but lacking resolution. For tandem-mass spectrometry (discussed in the next section) it is beneficial to combine fast technology for mass selection (quadrupole) or storing devices (linear ion traps) for high sensitivity with high resolution mass analysers (TOF and Orbitrap). During the last years various combinations of this different technology pools have led to a huge variety of different hybrid mass spectrometers.

HYBRID ORBITRAP INSTRUMENTS

The first generations of Orbitrap based hybrid instruments has been coupled to linear ion traps (LTQ). This combined the ultra-high resolution of the Orbitrap with the scan speed and the sensitivity of the ion trap (Fig. 8A)[58, 60, 80]. An additional advantage of this marriage is the possibility of parallelisation: While the Orbitrap is acquiring a high resolution MS1 scan the linear ion trap simultaneously is carrying out the precursor selection for MS2 and the fragmentation. The described architecture of an LTQ was further enhanced by the

implementation of a dual-pressure cell design[81]. The fragmentation capabilities were improved by the implementation of a multipole collision cell for beam-type collision induced dissociation (also higher energy collision induced dissociation, HCD)[82] with Orbitrap readout for MS2. Enabling high resolution data in MS1 as well as in MS2[83, 84].

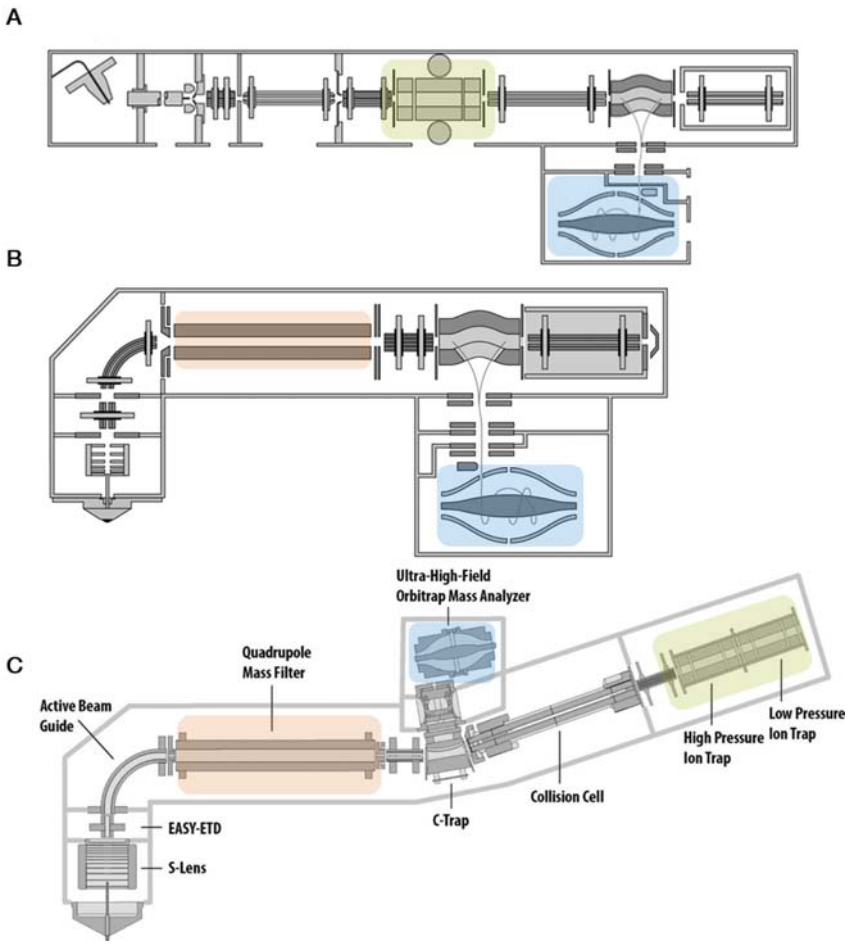


Figure 8 Hybrid Orbitrap systems Schematic pictures of different hybrid Orbitrap systems. Highlighted are the Orbitrap (blue), the linear-ion trap (LTQ, green) and the quadrupole (orange) (A) LTQ-Orbitrap Elite[60] (B) Quadrupole Orbitrap Q-Exaktive[85] (C) The quadrupole-Orbitrap-ion trap system Orbitrap Fusion.

Since HCD was promoted to perform the read-out of both MS1 and MS2 in the Orbitrap, the logical step in the evolution of Orbitrap instruments was launched in 2011[85]. A combination of Orbitrap with a quadrupole (Q-Exaktive, Fig. 8B). This set-up allowed even higher acquisition speeds through the isolation speed and selectivity of the quadrupole. However, this set up was lacking the high sensitivity of the LTQ and was not allowing parallelisation. For the purpose of combining all possible features into one instrument the Orbitrap Fusion (Fig. 8C) was released[73]. The Fusion is a quadrupole-Orbitrap-ion trap-hybrid and allowing a huge set of different acquisition schemes and a fast performance of >25 scans/s.

QUADRUPOLE TOF INSTRUMENTS – SYNAPT G2SI

In the last decade, quadrupole TOF (Q-TOF) instruments have become one of the most widespread platforms in proteomics. The instrument primarily used in this work was the Synapt G2Si from Waters Corporation. Besides the quadrupole and the orthogonal acceleration TOF it also features a travelling wave ion mobility device for gas phase separation. The fundamental geometry is based on the Synapt G2[49] but the G2Si features higher resolution and sensitivity (Fig. 9).

The gain in sensitivity is achieved by an orifice with a higher diameter for increased ion flux into the instrument plus the incorporation of a Stepwave ion guide for increased signal to noise ratio.

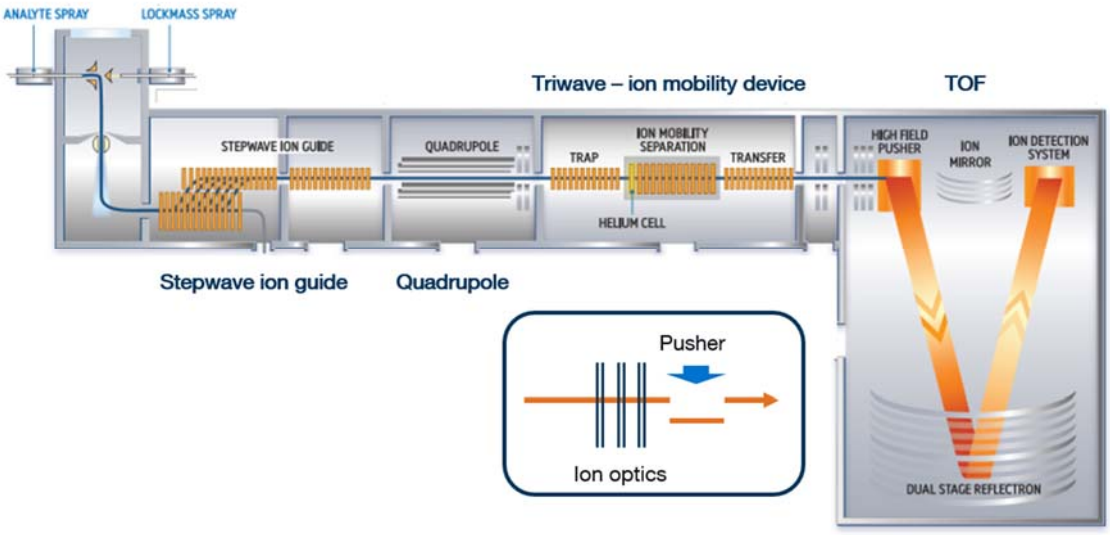


Figure 9 Waters Synapt G2S Schematic illustration of the Waters Synapt G2Si. The insert highlights the process of orthogonal acceleration TOF procedure.

The ions originate from the ESI ion source, consisting of a Z-sprayer architecture for the analyte sprayer to minimize background signal and of a lockmass sprayer for calibration. A baffle allows to switch, manual or automated during an acquisition, between both sprayers. The ions enter the instrument and are focussed by a SRIG and additionally separated from neutral background before arriving at the quadrupole. After the quadrupole the Triwave ion mobility device is incorporated.

In this compartment fragmentation for tandem mass spectrometry takes place. Two regions assembled of stacked ring ion guides can be used for collision-induced-dissociation (CID) fragmentation. CID can be performed either in the trapping region or in the transfer region of the Triwave. Thus it is possible to fragment the selected precursors

and separate the fragments by ion mobility or to separate the precursors by ion mobility and fragment afterwards.

After the ion mobility device the ions enter the orthogonal acceleration TOF (Fig. 9). In order to couple a TOF with an ESI source, the continuous ion beam must be converted into a pulsed beam. This is achieved by a high field orthogonal acceleration pusher. This pusher is in consequence determining the start for the time measurement in the TOF analyser. The ion path in this TOF can be manipulated by a series of reflectrons to achieve different resolutions (10,000 - 40,000) with the eventual cost of a loss in sensitivity.

PROTEIN IDENTIFICATION BY MASS SPECTROMETRY

Mass spectrometry has become the standard technique for holistic and comprehensive protein and peptide identification[2, 86-88]. Besides, it has proven its merits in protein quantification[89] and for the identification of post translational modifications[3, 90].

Top-down and bottom-up are the two main MS based proteomic strategies of today. The top-down approach aims to completely identify and characterize intact proteins. Beside certain challenges on the level of protein fractionation[91] and data analysis[92], this approach is currently able to identify over >1,000 proteins and >3,000 modification states (proteomfrome) in one experiment[93].

Bottom-up proteomics involves the proteolytic cleavage of proteins into peptides prior to the analysis and in consequence creates, out of an already complex protein mixture, an even more complex peptide mixture with >100,000 possible peptide species[94]. However, bottom-up proteomics approaches are able to identify >5,000 proteins in under 2 h[73, 74]. Hence, bottom-up proteomics provides a deeper proteome coverage and was exclusively used during this thesis.

THE BOTTOM UP PROTEOMICS WORKFLOW

The standard bottom-up proteomic workflow (also shotgun proteomics, Fig. 10) includes several steps before the sample is analysed by mass spectrometry. It starts with the protein extraction from the sample (e.g. from tissues or cell lines). At this stage it is possible to incorporate e.g. an affinity purification step for the enrichment of a (low abundant) protein subpopulation e.g. kinases or membrane proteins. In any case the proteins are digested by a sequence specific protease. The standard enzyme is trypsin, cleaving after the C-terminus of the basic amino acids lysine and arginine[95]. Thus every tryptic peptide contains at least one basic amino acid residue resulting in a beneficial effect on ionisation and charge stabilisation[96]. The length of the created peptides is on average 10 amino acids and is well suited for peptide sequencing. In recent approaches, the benefit of using complementary enzymes like LysC, AspN and GluC for increased sequence coverage has been shown[97-100].

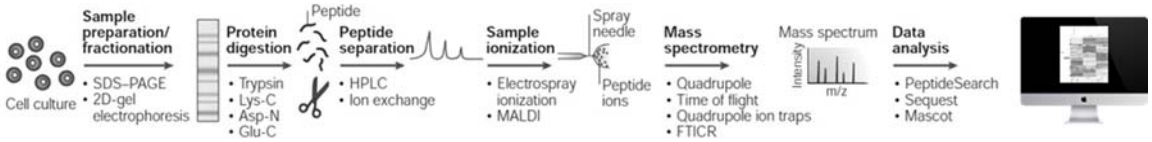


Figure 10 Bottom-up proteomic workflow Schematic overview over a typical mass spectrometry based bottom-up proteomic workflow The different steps include sample preparation, protein digestion, peptide separation, ionisation and adjacent MS and data analysis. Different protocols and procedure for each of these steps are available. (adapted from Steen et al.,2004[101])

Protein digestion can be performed in several ways. Still widely used is the one dimensional gel electrophoresis followed by in gel digestion[102], further protocols for in-solution digestion or filter-aided-sample-preparation (FASP) are available[103].

CHROMATOGRAPHIC SEPARATION OF PEPTIDES

During this workflow a highly complex peptide mixture is created which outruns the analysis capabilities of mass spectrometers. Therefore making upfront separation techniques attractive to adapt to instrument parameters like detection limit, dynamic range and acquisition speed[104]. Liquid chromatography using ion-pair reversed phase C_{18} material is the most common technique and separates the peptides according to their hydrophobicity. The peptides are dissolved under aqueous acidic conditions and the protonated peptides are retained on the hydrophobic column material. Elution occurs when the hydrophobicity of the solvent matches the peptides' hydrophobicity. For the peptide this is dependent on its amino acid composition[105]. For the LC solvent, a gradient of a more hydrophobic solvent (acetonitrile) is applied and thus changing the

hydrophobicity of the mobile phase over the time. The capacity and the resolution of chromatographic systems is thereby increasing by the use of long narrow columns and small particle sizes[69, 106-108]. The commonly used solvents (water and acetonitrile) are MS compatible and therefore allow a directly coupling of this set-up to an ESI source of a mass spectrometer (online LC-MS/MS). Typical nanoflow operates in the range of 100 to 500 nL/min using nano-LC columns of 50 to 100 µm inner diameter (particle size of 1 to 5 µm). Additionally, several attempts have been made to use very long columns (4 m)[109-111] also in combination with ultra-low flowrates of 20 nl/min[110, 111] resulting in zmol sensitivity.

In order to achieve even deeper proteome coverage in large scale proteomic studies, the use of further orthogonal separation steps is often applied. Routinely used approaches are isoelectric focusing (IEF), strong anion/cation exchange chromatography (SAX/SCX) or hydrophilic interaction chromatography (HILIC). Furthermore, it is also possible to enrich for a specific subset of peptides e.g. phosphorylated peptides. This has recently been proven to achieve high reproducibility and a high level of comprehensiveness[112].

Following this workflow, the peptide/protein identification is carried out using tandem mass spectrometry and database searching.

PROTEIN IDENTIFICATION BY PEPTIDE IDENTIFICATION

Tandem mass spectrometry consists of two iterative procedures which enable the structural characterisation of peptides. In the first step the intact *m/z* value of a peptide is determined (precursor, MS1) and in the second step it is selected for adjacent fragmentation leading to a tandem mass spectrum. Conceptual tandem mass spectrometry can be performed by “tandem in time” or by the “tandem in space”. The notation refers to the fact that mass selection and recording of the tandem mass spectrum is performed in two separate mass analysers within the same instrument (tandem in time, ion traps) or consecutively using the same analyser (tandem in space, QQQ, Q-TOF, LTQ-Orbitrap).

FRAGMENTATION

For generating a tandem mass spectrum several fragmentation techniques have been explored. Mostly used in proteomics research are the collision induced dissociation (CID)[113], the high energy C-trap dissociation (HCD)[82] and the electron transfer dissociation (ETD)[114]. Depending on the fragmentation technique, different peptide fragments are created and classified according to the nomenclature introduced by Roepstorff and Fohlmann[115] and modified by Johnson et a[116]. Accordingly the fragments containing the N-terminal side of the peptide are a-, b-, or c-ions and C-terminal fragments are called x-, y-, z- ions (Fig. 11).

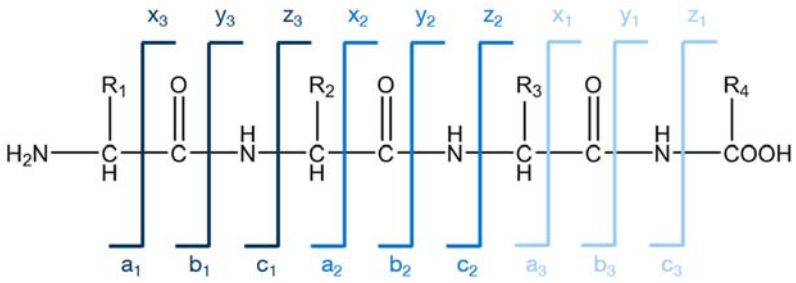


Figure 11 Peptide fragmentation nomenclature Fragment ions containing the N-terminus are called a, b and c ions, while x, y, and z ions contain the C-terminus. The fragment ions are named according to the backbone bond cleavage. The mass difference between two adjacent fragments corresponds to the residue mass of an amino acid. Chemical bond angels are unconsidered. (adapted from Roepstorff et al.,1984[115, 116])

CID and HCD use an inert gas for fragmentation. During the collisions of the intact peptide with the gas molecules (e.g. He, Ar, N_2) the acquired vibrational energy leads to the breakage of chemical bonds. Mostly the peptide bond breaks during CID/HCD creating b- and y- ions, thus makes both suitable techniques for peptide sequencing. The difference between CID and HCD is the amount of energy that is applied during the fragmentation. During “low-energy” CID (trap-type CID) in ion traps the fragmentation in the isolated peptide species is induced by an RF field leading to a higher kinetic energy and the induction of multiple collisions with the gas molecules. One drawback of CID in ion traps is the low mass cut-off caused by an inefficient recovery of ions with a mass of less than ~30% of the precursor[117]. This is a substantial limitation of this kind of fragmentation, since it cannot be used for peptides carrying isobaric mass tags. Further, the identification of labile PTMs (e.g. phosphorylation) is quite limited since during the thermodynamically CID fragmentation procedure the weakest bond is cleaved first. During “High-energy” CID (beam-type CID, HCD), as it is used in combination with Orbitraps or TOF analysers, the precursors are fragmented in a dedicated collision cell (e.g. HCD cell or one of the two SRIGs in the Triwave) by applying a higher kinetic energy.

ETD uses an electron donor for fragmentation induction. Due to chemical ionisation, the donor is becoming a free radical anion and by reaction with the positive charged peptide ions leads to a charge reduced species with an unpaired electron. The electron configuration is highly unstable and undergoes immediate decay by N-C_α bond breakage, resulting in c- and z-ions[114]. In order to increase the information content of ETD spectra and to overcome certain shortcomings of this approach a combinatorial strategy has been developed, making use of ETD and HCD fragmentation- EThcD[118].

DATA ACQUISITION

During a LC-MS/MS experiment the selection of peaks that are chosen for adjacent fragmentation is a fundamental step. The most common is a data dependent acquisition (DDA) approach, selecting the peaks primarily according to their abundance but also considering the charge state ($z \geq 2$) from a MS1 scan. A MS1 scan contains all, at this time present, *m/z* values. The MS1 scan (full scan) is than followed by a number of MS2 scans (fragment spectra) of the selected precursors. In order to avoid resampling of earlier selected precursors, a dynamic exclusion list is used. Additional it is possible to use an inclusion list, which forces the instrument to select a desired peak for fragmentation regardless of its abundance in the MS1 spectrum to ensure the identification. The result of such an acquisition is a list of picked precursor *m/z* values with a descending fragment ion series, which is further used by database search algorithm. In targeted proteomics the MS is given a predefined list of peptide precursor masses corresponding to proteins that are of special interest. The MS focusses here only on the acquisition of the MS2 spectra for the predefined masses – Multi reaction monitoring (MRM)[89, 119] or parallel-reaction-monitoring (PRM)[78]. This barriers the advantage that regardless of the presence of a precursor above the noise, a MS2 is generated. For this approach of importance is the validation of targeted masses and its transitions (recorded fragment ions) prior to the experiment, nevertheless these approaches provides extensive sensitivity. Conceptual limitations of DDA might be a stochastic and irreproducible selection of precursor ions for fragmentation compromising the reproducibility[120] or undersampling[94]. Thus different data-independent acquisitions (DIA) methods have been introduced. In DIA methods, no precursor selection is performed, rather all present precursors of a distinct mass window are co-fragmented. So far almost all mass spectrometer vendors have developed DIA methods – e.g. All-ion fragmentation (AIF, Thermo Scientific)[75], SWATH (ABISciex)[121] and MS^e (Waters Corporation)[122]. The precursor list with the corresponding peptides is regenerated *in silico.*

DATABASE SEARCHING

In bottom-up proteomics proteins are identified based on peptide identifications from tandem mass spectra searched against primary sequence databases. For this purpose a set of search engines like e.g. Mascot[123] or Andromeda[124] have been developed. In principle all these search engines work in a similar fashion (Fig. 12A): The acquired tandem mass spectra are searched against an *in silico* tandem mass spectrum database derived from all possible peptides of the containing protein sequences. Peptides are thereby identified due the correlation of the theoretical tandem mass spectrum to the experimentally acquired one. For the generation of the theoretical spectra, certain (user defined) criteria are considered like intact peptide mass, enzyme specificity, allowed number of missed cleavages, possible modifications or used fragmentation technique. Every peptide-spectrum-match (PSM) receives a score, reflecting the similarity of the theoretical to the experimental data. Here the score is calculated from several metrics, such as length of consecutive ion series, overlap of ion series (e.g. b-/y-ions), number of matching fragments, precursor mass error, fragment mass error, etc. Therefore database search is profiting from high resolution and high mass accuracy experimental data. Regardless slight differences in the scoring schemes of different search algorithms[125], in the end the presence of proteins in a sample are than correlated to the identified peptides.

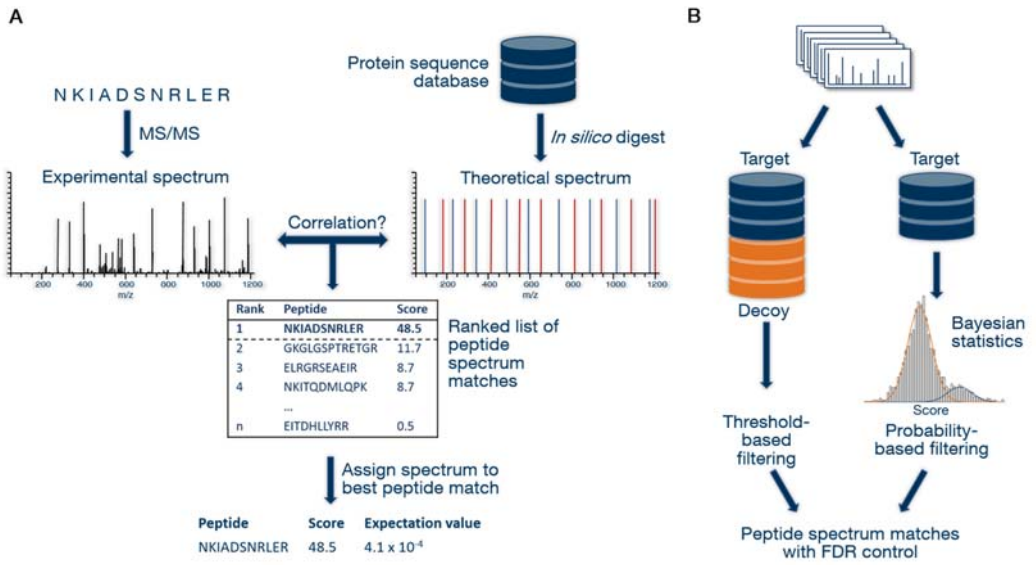


Figure 12 Database searching (A) Peptide and protein identification by tandem mass spectra is performed by comparison of experimental and theoretical spectra. A score for the peptide-spectrum-match is indicating the similarity of the theoretical to the experimental data. (B) The target-decoy or the Bayesian statistics approach are used for controlling the FDR. (adapted from Hahne, 2012[21] and Nesvizhskii et al.,2004[126])

One problem of all scoring based approaches is the identification of false-positive-hits on peptide, and corresponding on the protein, level. The reasons for this can be various like low signal-to-noise spectra, chimeric spectra, incorrectly assigned precursor charge/intact peptide mass[127]. A recent study showed that a significant amount of these false positive hits consist of no more than 7 amino acids but still obtains reasonable scores[14]. In order to obtain an estimation for the amount of false-positive hits in a dataset a global and a local approach can be used. For the calculation of a global false-discovery-rate (Fig. 12B) (FDR) a target-decoy approach is used. For this purpose a normal "target" sequence database is used plus a randomised or inverted "decoy" database. All matches in the decoy database are by definition false-positive-hits and, by definition, wrong. The ratio between the target hits and the decoy hits at a certain score cut-off is referred as the FDR. The Bayesian statistics approach, on the other hand, results in a local peptide probability based on score distribution models of correct and false hits.

Recently it has been shown that the FDR approach on the global scale is affected by diminishing returns in very large studies, e.g. containing almost the complete human proteome[14]. The underlining issue is that incorporating very large datasets simultaneously large numbers of decoy hits accumulate. Thus leading to a loss of true protein hits while using a constant protein FDR (e.g. 1%), which speaks against pure logic. Current approaches to circumvent this issue are the use of very stringent criteria on the PSM level[128] or the development of a new "picked" protein FDR approach (Savitski et al. 2015, manuscript in revision).

QUANTITATIVE PROTEOMICS

In the early stages of mass spectrometry based proteomics, only the presence of a protein could be monitored. In recent years the quantification of thousands of proteins and PTMs has become daily routine in proteomics[89]. Today's scientists can choose from a toolbox rich in different techniques for relative and absolute quantification[87, 129, 130] (Fig. 13).

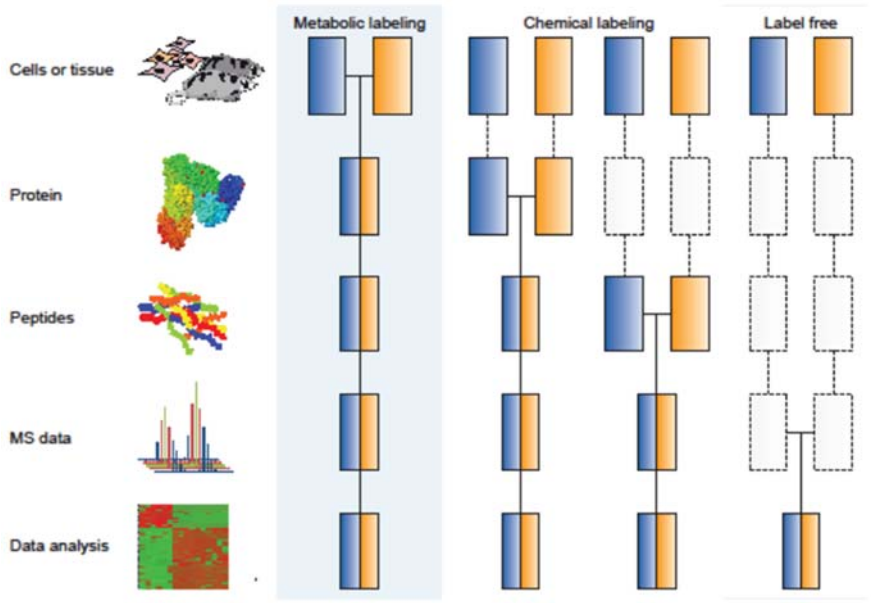


Figure 13 Quantitative mass spectrometry The different coloured boxes (blue and yellow) represent two biological experiments. Dashed lines indicate possible levels of experimental variation due to sample handling or analysis. (adapted and modified from Bantscheff et al., 2007[131])

The workflows can be discriminated by the use of a stable isotope label or if the quantification is performed label-free. During the quantification using stable isotopes, a mass shift is introduced into the different labelled peptides. So the identification of the sample's affiliation is possible and quantitative information is provided. Whereas in label-free approaches samples are quantified according to the signal intensity (or number of acquired spectra) in the mass spectrometer during individual measurements.

The labelling using stable isotopes rests on the stable isotope dilution theory. This theory claims that the physicochemical properties of the labelled and the native version of the peptide are identical and show similar mass spectrometric response. Due to the fact that mass spectrometers are able to recognise the introduced mass shift through the isotopic label, samples can be combined during the workflow and relative quantification is performed by the relative signal response. The sample merging can be performed at

different stages of the proteomic workflow, depending on how early the label is introduced. The earliest stage the label can be introduced is by metabolic labelling (e.g. SILAC[132]). Later labelling is possible by chemically (e.g. iTRAQ[133], TMT[134], dimethyl[135]) or enzymatically (e.g. O^{18} labelling[136]) labelling of proteins or peptides.

Different to metabolic labelling, which requires a functional protein biosynthesis machinery, chemical labelling can be applied to a wide range of samples and thus became fairly popular. These methods, e.g. the tandem mass tag (TMT) and the isobaric tags for absolute and relative quantification (iTRAQ), rely on N-hydroxysuccinimide chemistry and target primary amines of the peptide/protein N-term or the ε-amino group of the lysine.

The chemical design of the e.g. TMT reagents is done in a way that on the peptide level (MS1) the mass shift for all peptides is identical, however, the fragmentation pattern of each label differs and produces different reporter ions in the MS2 for quantification. The benefit of this MS2 based approach is the possibility of extensive multiplexing[137, 138] by e.g. using 10-plex TMT[139] during a single MS analysis and in the same time preventing increased MS1 complexity. Highly selective precursor selection is crucial to avoid interference of reporter ions leading to ratio compression[140], however, several approaches have been developed to overcome this limitation by using multi-stage fragmentation approaches[12, 141, 142] or making use of gas phase purifications[143] e.g. by the use of travelling wave ion mobility[144]. Dimethyl labelling is a further chemical labelling approach, here the amine groups are reacted with (heavy) deuterated formaldehyde or without (light). For relative quantification the introduced 4 Da mass shift is used for MS1 based quantification via signal integration.

Label-free quantification, on the other hand, offers the possibility to compare (in theory) an infinite number of experiments. It can be performed either intensity-based or by spectral counting. Spectral counting relies on the experimentally observed positive correlation between protein amount in the sample and number of matched spectra. However, this has the downside that there is also a correlation between protein size, number of peptides and consecutively number of possible spectrum matches. Quantitative information in the intensity based approaches is derived from extracted ion chromatograms (XIC) of the identified peptides and integration of the area-under-the-curve[145]. Using the integrated peptide signal between different experiments results in a relative protein/peptide quantification. In addition several tools have been implemented to improve the extraction of the XICs and to improve sample coverage, e.g. by match-

between-run option[146-148] or even lead to an absolute protein quantification (iBAQ[149, 150]). Label-free quantification can not only be performed by MS1 peak integration, like it is done in classical DDA experiments. Label-free quantification can also be performed by integrating the signal intensities of fragment ions in the MS2 spectrum. Especially during multi-reaction-monitoring (MRM)[119] or parallel-reaction-monitoring (PRM)[78] approaches. Here no MS1 information is available. The main differences between label-free quantification by classical DDA (MS1) or MRM (MS2) is that for DDA no list needs to be defined beforehand and the MRM approaches provides extensive sensitivity. The later point originates from the fact that the signal for the predefined peptide is collected even though the precursor is not detected above the noise in MS1. Quantification on the MS2 level may also be a useful workaround for MS systems (e.g. Q-TOFs) that are prone to detector saturation[151]. Nevertheless, the aforementioned possible combination of an unrestricted number of samples in label-free quantification involves extensive MS analysis time.

CHEMICAL PROTEOMICS

FOR DRUG SELECTIVITY PROFILING

CHEMICAL PROTEOMICS

The field of proteomics greatly benefited from the ability to quantify >10,000 proteins over a broad dynamic range for single cell line[8, 9], over a big variety of different cell lines[13] or even the whole human proteome[14]. However, the protein expression level is just one level of regulation. Protein kinases, for example, regulate cellular processes via protein phosphorylation, but the introduced modification, as well as, the protein kinase itself are of low abundance. Chemical proteomics is a highly interdisciplinary approach, combining organic chemistry, cell biology with mass spectrometry to provide probes for the specific enrichment of sub-proteomes. Thereby preserving close to *in vivo* conditions. The merits of different chemical proteomic strategies[152-155] have already been shown for several subsets of functional related protein targets, like kinases[129, 156] or histone-deacetylases[157]. Experimental strategies (Fig. 14) can be grouped into activity- and affinity based chemical proteomics[153]. Either the activity of proteins e.g. enzymes (activity-based protein profiling, ABPP) or immobilized bioactive molecules are used for a specific the enrichment (compound-centric chemical proteomics, CCCP).

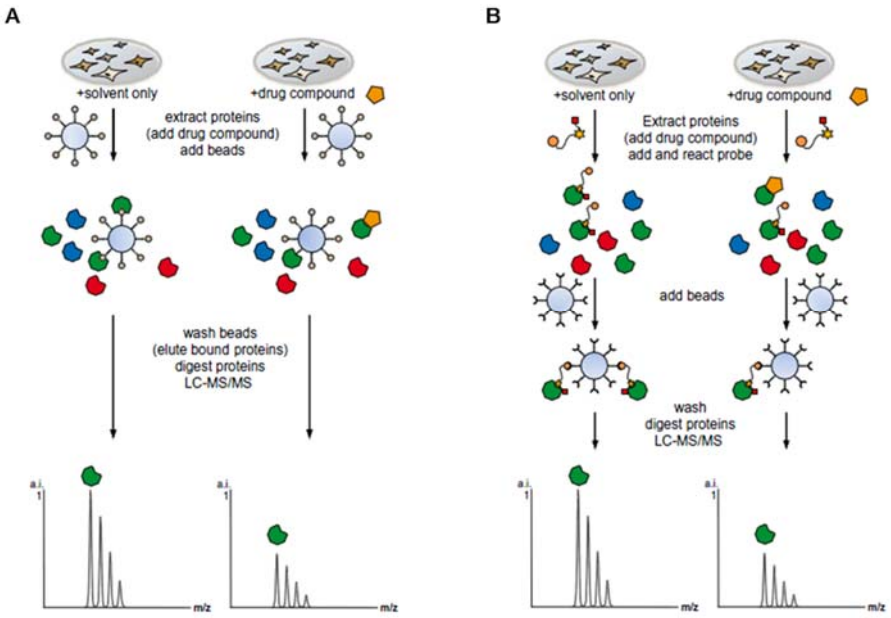


Figure 14 Chemical proteomics workflows (A) CCCP: The immobilized drug of interest is incubated with a protein extract for enrichment of the desired protein family. Adding free drug is inhibiting the binding of a protein to the matrix and thus reducing the abundance of the captured protein. (B) ABPP: Essential is a bifunctional probe is specifically targeting the active site of an enzyme family. The probe is covalently binding to its target. The purification of the probe and its bond target are purified in a subsequent step. The quantification is in both cases performed by mass spectrometry. (adapted from Bantscheff et al., 2012[154])

In ABPP a bifunctional chemical probe is used that covalently binds to the active site of its enzymatic target class. The reactive group is responsible for the irreversible binding of the active site of the target class (ideal all members), and the reporter tag is used for the enrichment and identification of the targeted proteins[158, 159].

CCCP strategies in affinity based chemical proteomics use active small molecules that are immobilised on a matrix to enrich their binding partners[154, 160]. Commonly used compounds for this approach are drugs or drug candidates from medicinal chemistry, which bind to their designated target in a non-covalent fashion but with a high affinity. The chosen compound is coupled to an inert matrix ensuring the retention of its activity during this functionalization procedure[129, 160, 161]. Afterwards the drug-matrix is incubated with cell lysate for the enrichment and subsequently analysed by LC-MS/MS. Affinity based approaches allow, in contrast to ABPP strategies, the identification of binding and interaction partners without the necessity of enzymatic activity. The CCCP strategy has already provided deeper insight into the target spectrum of several clinical drugs designed against protein kinases[129, 161-164], or histone deacetylases[157]. Furthermore, this approach also allowed the investigation of the target class in the context of diseases, like head and neck cancer[165] or leukaemia[166].

KINASES IN CANCER

Cancer covers a group of diseases, all feature the dysfunction of several cellular functions, which lead to, initially local then systemic, unorganised cell proliferation[167]. For the year 2015 the American Cancer Society projects approx. 1.6 million new cancer cases and almost 600,000 cancer related deaths[168]. On a global scale over 14 million new cases and over 8 million cancer related deaths occurred in the year 2012, with lung cancer being still number one cause of death for male[169]. Consequently huge effort has been dedicated for fundamental cancer research as well as for cancer drug development[170].

Protein kinases can be considered as one of the key regulators in cellular signalling and their malfunction could be associated with several human diseases, like inflammation, diabetes and in particular cancer[171-173]. The human genome contains 518 genes coding for putative protein kinases. This represents one of the largest gene families[174]. Based on sequence homologies in the catalytic domain, kinases can be subdivided into 7 families (ACG, CAMK, CK1, CMGK, STE, TK and TKL, Fig. 15A). Kinases are enzymes that transfer a phosphate group (usually γ-phosphate of ATP) either to serine and threonine or tyrosine residues of their substrate. This has influence on the activity, the cellular localisation or the interaction spectrum of the substrate. The antagonist of the kinase is the phosphatase, which is able to hydrolyse the phospho-ester bond.

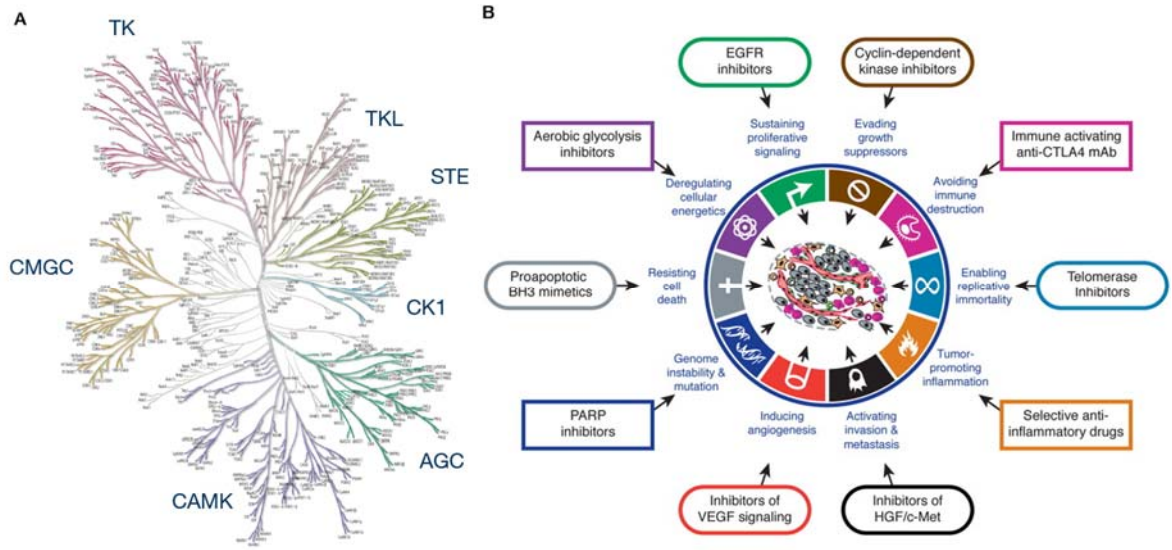


Figure 15 Protein kinases and cancer (A) Kinometree showing the 7 mayor groups of protein kinases, sorting according to genetic homology.(© Cell Signaling Technology) (B) The hallmarks of cancer and the therapeutic targeting approaches. (adapted from Hanahan et al.,2011[175])

Several impacts on different biological processes can lead to a pathological interference of cell's communication system. The hallmarks of cancer[175, 176] (Fig. 15B) summarise a set of characteristics a normal cell needs to acquire for becoming a cancer cell.

Protein kinases are primary regulators of all hallmarks of cancer and consequently very attractive targets for cancer therapy. A prominent example for an (singular) event altering the function of a protein kinase is the Philadelphia chromosome translocation in chronic myelogenous leukemia (CML). The gene fusion between the bcr and the abl gene result in a permanent activation of the tyrosine kinase ABL, awarding the transformed cell with the ability of uncontrolled cell division[177]. Imatinib (Gleevec®, Novartis) was in 2001[178] the first FDA approved small molecule kinase inhibitor drug designed for a distinct target in cancer therapy and is often referred to as the paradigm for targeted cancer therapy[179].

SMALL MOLECULE KINASE INHIBITORS

Protein kinases are structurally quite homogenous and the folding is mainly composed out of a ß-sheet lobe and another lobe manly build from α-helices. The two lobes accommodate the active ATP binding site[174, 180]. Further, the flexible activation loop is marked by a conserved DFG motif and exist also for all protein kinases[181-183]. In general two positions are characteristic for the activation loop – "DFG-in" and "DFG-out" – which can be translated in the status of the kinase as "active" or "in-active" respectively.

Small molecule inhibitors are mostly designed to imitate ATP binding in the active site of the kinase[184]. In consequence the designed small molecule inhibitors are competing with the co-substrate ATP and thus inhibiting the activity of the kinase. Due to the aforementioned structural homology of protein kinases, the designed small molecules often lack specificity and target multiple kinases at the same time. Although this can may be referred as "chemists failure"[156] to develop selective molecules, this broadband kinase inhibitors can have beneficial effects in cancer therapy. Exactly because of the synergistic effects caused by blocking several pathways at the same time. As mentioned before, the BCR-Abl inhibitor Imatinib was first approved for the use in CML, but due to its "un-specificity" and coexisting potent inhibition of c-KIT and PDGFR it was also approved for the treatment of gastrointestinal tumours (GIST)[182].

At the end of 2014, about 30 small molecules kinase inhibitors are currently approved by the FDA and several hundred (approx. 200) are in clinical trials. Illuminating the whole target spectrum of a kinase inhibitor is of huge importance for understanding biological effects as well for finding new indications for a drug.

SELECTIVITY PROFILING OF KINASE INHIBITORS

More than 200 small molecule kinase inhibitors are in clinical trials, of which about 30 have been approved for use in humans so far. Because of their chemical structure, kinase inhibitors may not only act on its designated target, but may also target other proteins. This promiscuity establishes both challenges and opportunities.

Consequently it is of great importance to deconvolute the target space of a kinase inhibitor. Drug discovery is still experimentally driven by phenotypic screens[185-187] or by screening compounds against panels of protein kinases[188-191]. Last-mentioned kinase assays often only rely on the artificial expression of the kinase domain only. Albeit powerful and generally used, these assays do not detect many additional factors that influence kinase activity in cells (e.g. regulatory domains, interacting proteins, the proteins' conformational state and/or its post translational modification (PTMs) status)[152, 188, 192]. Chemical proteomics in combination with quantitative mass spectrometry attempts to give an unbiased selectivity profile of a drug[89, 152, 153, 193].

One out of many available approaches[193-196] for a drug-centric profiling is the Kinobeads technology[129, 156]. Kinobeads are a mixture of several immobilized ATP-competitive kinase inhibitors with a broad target range and have been introduced in 2007[129] and further

developed over the years[156, 197-199]. The current version of the Kinobeads (Kinobeads γ) is able to bind over 350 protein kinases as well as over 2000 additional ATP and nucleotide binding proteins from different cell lines[13, 156]. Kinobeads can be used for either the profiling of kinase abundance from different biological sources[13, 165] or determining the selectivity profile of a small molecule kinase inhibitor against over 250 native protein kinases in a single experiment[156].

For a standard profiling approach a concentration range of drug is added to a cell lysate mixture of 4 different cancer cell lines, providing improved kinome coverage, and subsequent kinase enrichment is performed. The ATP-competitive drug in the lysate is competing with the immobilized inhibitors on the affinity matrix for binding the kinases. Targets of the free drug lose their ability to bind to the matrix and are competed in a dose dependent manner. On the other hand kinases that are not targets of the drug are unaffected.

Kinobead selectivity screenings can be performed in high-throughput and engaged a public-private partnership to screen thousands of advanced drug molecules from the pharmaceutical industry with the purpose to unlock the current state of the drug-able kinome. This will in the future allow the detection of new chemical probes or give a starting point for medicinal chemistry[170].

OBJECTIVE AND OUTLINE

MS has become the standard technology for high-throughput proteomic studies. The field is largely driven by advances in technology as well as by the development of high throughput assays for sample generation. In this context the primarily objective of this thesis was the evaluation and development of new mass spectrometry based technologies for the usage in a large scale chemical proteomics assays with the goal of screening 200 kinase inhibitors currently in clinical studies. Therefore, in chapter II a data-independent acquisition strategy (HDMS^e), exclusively available on the Synapt Q-TOF instruments from Waters Corporation was evaluated for the use in high throughput bottom-up proteomics and especially for the use in the arforementioned large scale chemical proteomics study. During this evaluation the proprietary software for processing the DIA data was evaluated as well as the underlying hardware performance of the Q-TOF instrument. The obtained results have been compared to a standard DDA approach on a LTQ-Orbitrap Elite. It could be seen that despite conceptual advantages of the DIA method, the DDA approach still provided superior performance and was hereon used for the large scale chemical proteomics assay.

One limitation of orthogonal acceleration TOF instruments is the fairly low duty cycle, caused by the necessary bending of the beam by the orthogonal acceleration pusher. For this reason Chapter III was engaged with the development of a complete new way of performing DDA on a Q-TOF instrument by the use of travelling wave ion mobility, termed HD-DDA. The HD-DDA approach was evaluated for the use in bottom-up proteomics and showed state-of-the-art performance. This method is today available on all Synapt G2Si instruments.

Chapter IV describes the chemical proteomics characterisation of the CML drug ponatinib, which was performed during the prior mentioned large scale screening of clinical kinase inhibitors. Ponatinib is a FDA approved drug and raised attention since it was pulled of the market for some time due to the occurrence of severe side-effects. The here presented selectivity profiling of ponatinib did not lead to a direct explanation of the occurring toxicity. Nevertheless, it was possible to determine a set of new highly potent targets, e.g ZAK and MAPK14, of these drug. This helped to reveal a previously observed inhibitory effect on inflammation processes by ponatinib.

ABBREVIATIONS

ABPP	Activity-based protein profiling
AIF	All ion fragmentation
ACG, CAMK, CK1, CMGK, STE, TK, TKL	The 7 mayor kinase families
ATP	Adenosine triphosphate
CCCP	Compound-centric chemical proteomics
CCS	Collision cross section
CEM	Chain ejection model
CF	Compensation field
CID	Collision induced dissociation
CML	Chronic myeloid leukemia
CRM	Charge residue model
DB	Database
DC	Direct current
DDA	Data dependent acquisition
DF	Dispersion field
DFG motive	Single amino acid code – Aspartic acid, Phenylalanine, Glycine motive
DIA	Data independent acquisition
DMSO	Dimethyl-sulfoxide
ESI	Electrospray ionisation
ETD	Electron transfer dissociation
FAIMS	Field asymmetric waveform ion mobility separation
FASP	Filter-aided sample preparation
FDR	False discovery rate
FT	Fourier transformation
GIST	Gastrointestinal tumour
HCD	High energy C-trap dissociation
HILIC	Hydrophilic interaction chromatography
HDMSe	Data independent acquisition using ion mobility separation (Waters, Corporation)
ICR	Ion cyclotron mass spectrometer
IEM	Ion evaporation model
IMS	Ion mobility separation
iTRAQ	Isobaric tag for relative and absolute quantification
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography coupled with tandem mass spectrometry
LOD	Limit of detection
LTQ	Linear trap quadrupole

m/z	Mass to charge ratio
MALDI	Matrix-assisted laser desorption/ionisation
MS	Mass spectrometer
MS1	Mass spectrum with intact precursors
MS2	Fragment mass spectrum
MSe	Data independent acquisition (Waters, Corporation)
MRM	Multi reaction monitoring
oaTOF	Orthogonal acceleration TOF
PRM	Parallel-reaction monitoring
PTM	Post translational modification
Q	Quadrupole
Q-TOF	Quadrupole Time-of-flight mass spectrometer
RF	Radio frequency
SAX	Strong anion exchange chromatography
SCX	Strong cation exchange chromatography
SILAC	Stable isotope labelling with amino acids in cell culture
SRIG	Staged ring ion guide
TMT	Tandem mass tag
TOF	Time-of-flight mass analyser
XIC	Extracted ion chromatogram

REFERENCES

1. International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
2. Wasinger, V.C. et al. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* **16**, 1090-1094 (1995).
3. Mann, M. & Jensen, O.N. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255-261 (2003).
4. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291-294 (1998).
5. Blackstock, W.P. & Weir, M.P. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol* **17**, 121-127 (1999).
6. O'Farrell, P.H. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-4021 (1975).
7. Klose, J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231-243 (1975).
8. Nagaraj, N. et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* **7**, 548 (2011).
9. Beck, M. et al. The quantitative proteome of a human cell line. *Mol Syst Biol* **7**, 549 (2011).
10. Wang, H. et al. Systematic optimization of long gradient chromatography mass spectrometry for deep analysis of brain proteome. *J Proteome Res* **14**, 829-838 (2015).
11. Helm, D. et al. Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol Cell Proteomics* **13**, 3709-3715 (2014).
12. McAlister, G.C. et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem* **86**, 7150-7158 (2014).
13. Moghaddas Gholami, A. et al. Global proteome analysis of the NCI-60 cell line panel. *Cell reports* **4**, 609-620 (2013).
14. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587 (2014).
15. Kim, M.S. et al. A draft map of the human proteome. *Nature* **509**, 575-581 (2014).
16. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
17. Tanaka, K. et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time? of? flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **2**, 151-153 (1988).
18. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-2301 (1988).
19. Knochenmuss, R. Ion formation mechanisms in UV-MALDI. *Analyst* **131**, 966-986 (2006).
20. Wilm, M. Principles of electrospray ionization. *Mol Cell Proteomics* **10**, M111 009407 (2011).
21. Hahne, H. (Technische Universität München,Diss., 2012).
22. Mann, M. & Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* **66**, 4390-4399 (1994).
23. Iribarne, J.V. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics* **64**, 2287 (1976).

24. Konermann, L., Ahadi, E., Rodriguez, A.D. & Vahidi, S. Unraveling the mechanism of electrospray ionization. *Anal Chem* **85**, 2-9 (2013).
25. Wilm, M. et al. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466-469 (1996).
26. Hahne, H. et al. DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Methods* **10**, 989-991 (2013).
27. Kanu, A.B., Dwivedi, P., Tam, M., Matz, L. & Hill, H.H., Jr. Ion mobility-mass spectrometry. *J Mass Spectrom* **43**, 1-22 (2008).
28. McDaniel, E.W. & Mason, E.A. Transport properties of ions in gases. (Wiley, 1988).
29. Hogg, A. & Kebarle, P. Mass‐Spectrometric Study of Ions at Near‐Atmospheric Pressure. II. Ammonium Ions Produced by the Alpha Radiolysis of Ammonia and Their Solvation in the Gas Phase by Ammonia and Water Molecules. *The Journal of Chemical Physics* **43**, 449-456 (1965).
30. Jurneczko, E. & Barran, P.E. How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* **136**, 20-28 (2011).
31. Bereszczak, J.Z. et al. Structure, stability and dynamics of norovirus P domain derived protein complexes studied by native mass spectrometry. *J Struct Biol* **177**, 273-282 (2012).
32. Lanucara, F., Holman, S.W., Gray, C.J. & Eyers, C.E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem* **6**, 281-294 (2014).
33. Hall, Z., Politis, A. & Robinson, C.V. Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. *Structure* **20**, 1596-1609 (2012).
34. Politis, A., Park, A.Y., Hall, Z., Ruotolo, B.T. & Robinson, C.V. Integrative modelling coupled with ion mobility mass spectrometry reveals structural features of the clamp loader in complex with single-stranded DNA binding protein. *J Mol Biol* **425**, 4790-4801 (2013).
35. Zhou, M. et al. Ion mobility-mass spectrometry of a rotary ATPase reveals ATP-induced reduction in conformational flexibility. *Nat Chem* **6**, 208-215 (2014).
36. Buryakov, I., Krylov, E., Nazarov, E. & Rasulev, U.K. A new method of separation of multi-atomic ions by mobility at atmospheric pressure using a high-frequency amplitude-asymmetric strong electric field. *International Journal of Mass Spectrometry and Ion Processes* **128**, 143-148 (1993).
37. Purves, R.W. & Guevremont, R. Electrospray ionization high-field asymmetric waveform ion mobility spectrometry-mass spectrometry. *Anal Chem* **71**, 2346-2357 (1999).
38. Shvartsburg, A.A. et al. High-definition differential ion mobility spectrometry with resolving power up to 500. *J Am Soc Mass Spectrom* **24**, 109-114 (2013).
39. Purves, R.W. Enhancement of biological mass spectrometry by using separations based on changes in ion mobility (FAIMS and DMS). *Anal Bioanal Chem* **405**, 35-42 (2013).
40. Creese, A.J., Shimwell, N.J., Larkins, K.P., Heath, J.K. & Cooper, H.J. Probing the complementarity of FAIMS and strong cation exchange chromatography in shotgun proteomics. *J Am Soc Mass Spectrom* **24**, 431-443 (2013).
41. Creese, A.J. & Cooper, H.J. Separation and identification of isomeric glycopeptides by high field asymmetric waveform ion mobility spectrometry. *Anal Chem* **84**, 2597-2601 (2012).
42. Shvartsburg, A.A., Creese, A.J., Smith, R.D. & Cooper, H.J. Separation of a set of peptide sequence isomers using differential ion mobility spectrometry. *Anal Chem* **83**, 6918-6923 (2011).
43. Shvartsburg, A.A., Creese, A.J., Smith, R.D. & Cooper, H.J. Separation of peptide isomers with variant modified sites by high-resolution differential ion mobility spectrometry. *Anal Chem* **82**, 8327-8334 (2010).

44. Xuan, Y., Creese, A.J., Horner, J.A. & Cooper, H.J. High-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled with high-resolution electron transfer dissociation mass spectrometry for the analysis of isobaric phosphopeptides. *Rapid Commun Mass Spectrom* **23**, 1963-1969 (2009).
45. Shvartsburg, A.A., Zheng, Y., Smith, R.D. & Kelleher, N.L. Ion mobility separation of variant histone tails extending to the "middle-down" range. *Anal Chem* **84**, 4271-4276 (2012).
46. Giles, K. et al. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* **18**, 2401-2414 (2004).
47. Gerlich, D. Inhomogeneous rf fields: a versatile tool for the study of processes with slow ions. *State-Selected and State-to-State Ion-Molecule Reaction Dynamics, Part* **1**, 1-176 (2007).
48. Giles, K., Williams, J.P. & Campuzano, I. Enhancements in travelling wave ion mobility resolution. *Rapid Commun Mass Spectrom* **25**, 1559-1566 (2011).
49. Pringle, S.D. et al. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry* **261**, 1-12 (2007).
50. Jonscher, K.R. & Yates, J.R., 3rd The quadrupole ion trap mass spectrometer--a small solution to a big challenge. *Anal Biochem* **244**, 1-15 (1997).
51. March, R.E. Quadrupole ion traps. *Mass Spectrom Rev* **28**, 961-989 (2009).
52. Schwartz, J.C., Syka, J.E.P. & Quarmby, S.T. (
53. Douglas, D.J., Frank, A.J. & Mao, D. Linear ion traps in mass spectrometry. *Mass Spectrom Rev* **24**, 1-29 (2005).
54. Comisarow, M.B. & Marshall, A.G. Fourier transform ion cyclotron resonance spectroscopy. *Chemical physics letters* **25**, 282-283 (1974).
55. Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **72**, 1156-1162 (2000).
56. Hu, Q. et al. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **40**, 430-443 (2005).
57. Kingdon, K.H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review* **21**, 408-418 (1923).
58. Makarov, A. et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* **78**, 2113-2120 (2006).
59. Makarov, A., Denisov, E. & Lange, O. Performance evaluation of a high-field Orbitrap mass analyzer. *J Am Soc Mass Spectrom* **20**, 1391-1396 (2009).
60. Michalski, A. 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* **11**, O111 013698 (2012).
61. Denisov, E., Damoc, E., Lange, O. & Makarov, A. Orbitrap mass spectrometry with resolving powers above 1,000,000. *International Journal of Mass Spectrometry* **325**, 80-85 (2012).
62. Scigelova, M. & Makarov, A. Orbitrap mass analyzer - Overview and applications in proteomics. *Proteomics* **6 Suppl 2**, 16-21 (2006).
63. Wiley, W. & McLaren, I.H. Time‐of‐flight mass spectrometer with improved resolution. *Review of Scientific Instruments* **26**, 1150-1157 (1955).
64. Whittal, R.M. & Li, L. High-resolution matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer. *Anal Chem* **67**, 1950-1954 (1995).
65. Whittal, R.M., Russon, L.M., Weinberger, S.R. & Li, L. Functional wave time-lag focusing matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer: improved mass accuracy. *Anal Chem* **69**, 2147-2153 (1997).
66. Maier, S.K. (München, Technische Universität München, Diss., 2014, 2014).
67. Zubarev, R.A. & Makarov, A. Orbitrap mass spectrometry. *Anal Chem* **85**, 5288-5296 (2013).

68. Han, X., Aslanian, A. & Yates, J.R., 3rd Mass spectrometry for proteomics. *Curr Opin Chem Biol* **12**, 483-490 (2008).
69. Yates, J.R., Ruse, C.I. & Nakorchevsky, A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* **11**, 49-79 (2009).
70. Mallick, P. & Kuster, B. Proteomics: a pragmatic perspective. *Nat Biotechnol* **28**, 695-709 (2010).
71. Pachl, F., Ruprecht, B., Lemeer, S. & Kuster, B. Characterization of a high field Orbitrap mass spectrometer for proteome analysis. *Proteomics* **13**, 2552-2562 (2013).
72. Zubarev, R.A., Håkansson, P. & Sundqvist, B. Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements. *Analytical Chemistry* **68**, 4060-4063 (1996).
73. Hebert, A.S. et al. The One Hour Yeast Proteome. *Mol Cell Proteomics* (2013).
74. Scheltema, R.A. 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* **13**, 3698-3708 (2014).
75. Geiger, T., Cox, J. & Mann, M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol Cell Proteomics* **9**, 2252-2261 (2010).
76. 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* **83**, 5442-5446 (2011).
77. Kast, J., Gentzel, M., Wilm, M. & Richardson, K. Noise filtering techniques for electrospray quadrupole time of flight mass spectra. *J Am Soc Mass Spectrom* **14**, 766-776 (2003).
78. Gallien, S. et al. Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol Cell Proteomics* **11**, 1709-1723 (2012).
79. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212-217 (2006).
80. Olsen, J.V. et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **4**, 2010-2021 (2005).
81. Second, T.P. et al. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal Chem* **81**, 7757-7765 (2009).
82. Olsen, J.V. et al. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **4**, 709-712 (2007).
83. Phanstiel, D.H. et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Methods* **8**, 821-827 (2011).
84. Hahne, H., Moghaddas Gholami, A. & Kuster, B. Discovery of O-GlcNAc-modified proteins in published large-scale proteome data. *Mol Cell Proteomics* **11**, 843-850 (2012).
85. Michalski, A. et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **10**, M111 011015 (2011).
86. Wilkins, M.R. et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/technology* **14**, 61-65 (1996).
87. Wasinger, V.C., Zeng, M. & Yau, Y. Current status and advances in quantitative proteomic mass spectrometry. *International journal of proteomics* **2013** (2013).
88. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
89. Bantscheff, M., Lemeer, S., Savitski, M.M. & Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **404**, 939-965 (2012).
90. Olsen, J.V. & Mann, M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics* **12**, 3444-3452 (2013).
91. Yates, J.R., 3rd & Kelleher, N.L. Top down proteomics. *Anal Chem* **85**, 6151 (2013).

92. Kelleher, N.L. Top-down proteomics. *Anal Chem* **76**, 197A-203A (2004).
93. Tipton, J.D. et al. Analysis of intact protein isoforms by mass spectrometry. *J Biol Chem* **286**, 25451-25458 (2011).
94. 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. *Journal of Proteome Research* **10**, 1785-1793 (2011).
95. Olsen, J.V., Ong, S.E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **3**, 608-614 (2004).
96. Vékey, K. Multiply charged ions. *Mass Spectrometry Reviews* **14**, 195-225 (1995).
97. 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* **13**, 1573-1584 (2014).
98. Choudhary, G., Wu, S.L., Shieh, P. & Hancock, W.S. Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS. *J Proteome Res* **2**, 59-67 (2003).
99. Swaney, D.L., Wenger, C.D. & Coon, J.J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* **9**, 1323-1329 (2010).
100. Tsiatsiani, L. & Heck, A.J. Proteomics beyond trypsin. *FEBS J* (2015).
101. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nature reviews. Molecular cell biology* **5**, 699-711 (2004).
102. Schirle, M., Heurtier, M.A. & Kuster, B. Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics* **2**, 1297-1305 (2003).
103. Wisniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).
104. Delmotte, N., Lasaosa, M., Tholey, A., Heinzle, E. & Huber, C.G. Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. *J Proteome Res* **6**, 4363-4373 (2007).
105. Spicer, V. et al. Sequence-specific retention calculator. A family of peptide retention time prediction algorithms in reversed-phase HPLC: applicability to various chromatographic conditions and columns. *Anal Chem* **79**, 8762-8768 (2007).
106. Kocher, T., Swart, R. & Mechtler, K. Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Anal Chem* **83**, 2699-2704 (2011).
107. Gale, D.C. & Smith, R.D. Small volume and low flow‐rate electrospray lonization mass spectrometry of aqueous samples. *Rapid communications in mass spectrometry* **7**, 1017-1021 (1993).
108. Thakur, S.S. et al. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol Cell Proteomics* **10**, M110 003699 (2011).
109. Horie, K. et al. Hydrophilic interaction chromatography using a meter-scale monolithic silica capillary column for proteomics LC-MS. *Anal Chem* **86**, 3817-3824 (2014).
110. Yue, G., Luo, Q., Zhang, J., Wu, S.L. & Karger, B.L. Ultratrace LC/MS proteomic analysis using 10-microm-i.d. Porous layer open tubular poly(styrene-divinylbenzene) capillary columns. *Anal Chem* **79**, 938-946 (2007).
111. Li, S. et al. An Integrated Platform for Isolation, Processing and Mass Spectrometry-based Proteomic Profiling of Rare Cells in Whole Blood. *Mol Cell Proteomics* (2015).
112. Ruprecht, B. et al. Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. *Mol Cell Proteomics* **14**, 205-215 (2015).
113. Hunt, D.F., Yates, J.R., 3rd, Shabanowitz, J., Winston, S. & Hauer, C.R. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* **83**, 6233-6237 (1986).

114. 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* **101**, 9528-9533 (2004).
115. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **11**, 601 (1984).
116. Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T. & Watson, J.T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal Chem* **59**, 2621-2625 (1987).
117. Cunningham, C., Jr., Glish, G.L. & Burinsky, D.J. High amplitude short time excitation: a method to form and detect low mass product ions in a quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* **17**, 81-84 (2006).
118. Frese, C.K. et al. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J Proteome Res* **12**, 1520-1525 (2013).
119. Domon, B. & Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology* **28**, 710-721 (2010).
120. Geromanos, S.J. et al. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683-1695 (2009).
121. Gillet, L.C. 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* **11**, O111 016717 (2012).
122. Li, G.Z. et al. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696-1719 (2009).
123. Perkins, D.N., Pappin, D.J.C., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).
124. Cox, J. et al. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research* **10**, 1794-1805 (2011).
125. Nesvizhskii, A.I., Vitek, O. & Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **4**, 787-797 (2007).
126. Nesvizhskii, A.I. & Aebersold, R. Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug Discov Today* **9**, 173-181 (2004).
127. Nesvizhskii, A.I. & Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* **4**, 1419-1440 (2005).
128. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
129. Bantscheff, M. et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* **25**, 1035-1044 (2007).
130. Bantscheff, M. & Kuster, B. Quantitative mass spectrometry in proteomics. *Anal Bioanal Chem* **404**, 937-938 (2012).
131. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**, 1017-1031 (2007).
132. Ong, S.E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).
133. Ross, P.L. et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-1169 (2004).
134. Thompson, A. et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).
135. Hsu, J.L., Huang, S.Y., Chow, N.H. & Chen, S.H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* **75**, 6843-6852 (2003).

136. Yao, X., Freas, A., Ramirez, J., Demirev, P.A. & Fenselau, C. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* **73**, 2836-2842 (2001).
137. Werner, T. et al. High-resolution enabled TMT 8-plexing. *Anal Chem* **84**, 7188-7194 (2012).
138. McAlister, G.C. et al. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal Chem* **84**, 7469-7478 (2012).
139. Werner, T. et al. Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal Chem* **86**, 3594-3601 (2014).
140. Savitski, M.M. et al. Measuring and managing ratio compression for accurate iTRAQ/TMT quantification. *J Proteome Res* **12**, 3586-3598 (2013).
141. Pachl, F., Fellenberg, K., Wagner, C. & Kuster, B. Ultra-high intra-spectrum mass accuracy enables unambiguous identification of fragment reporter ions in isobaric multiplexed quantitative proteomics. *Proteomics* **12**, 1328-1332 (2012).
142. Ting, L., Rad, R., Gygi, S.P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods* **8**, 937-940 (2011).
143. Wenger, C.D. et al. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Methods* **8**, 933-935 (2011).
144. Shliaha, P.V. et al. Additional Precursor Purification in Isobaric Mass Tagging Experiments by Traveling Wave Ion Mobility Separation (TWIMS). *J Proteome Res* **13**, 3360-3369 (2014).
145. Higgs, R.E., Knierman, M.D., Gelfanova, V., Butler, J.P. & Hale, J.E. Comprehensive label-free method for the relative quantification of proteins from biological samples. *J Proteome Res* **4**, 1442-1450 (2005).
146. Wang, P. et al. A statistical method for chromatographic alignment of LC-MS data. *Biostatistics* **8**, 357-367 (2007).
147. Strittmatter, E.F., Ferguson, P.L., Tang, K. & Smith, R.D. Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J Am Soc Mass Spectrom* **14**, 980-991 (2003).
148. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
149. Schwanhausser, B. et al. Corrigendum: Global quantification of mammalian gene expression control. *Nature* **495**, 126-127 (2013).
150. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).
151. Daly, C.E., Ng, L.L., Hakimi, A., Willingale, R. & Jones, D.J. Qualitative and quantitative characterization of plasma proteins when incorporating traveling wave ion mobility into a liquid chromatography-mass spectrometry workflow for biomarker discovery: use of product ion quantitation as an alternative data analysis tool for label free quantitation. *Anal Chem* **86**, 1972-1979 (2014).
152. Schirle, M., Bantscheff, M. & Kuster, B. Mass spectrometry-based proteomics in preclinical drug discovery. *Chem Biol* **19**, 72-84 (2012).
153. Rix, U. & Superti-Furga, G. Target profiling of small molecules by chemical proteomics. *Nat Chem Biol* **5**, 616-624 (2009).
154. Bantscheff, M. & Drewes, G. Chemoproteomic approaches to drug target identification and drug profiling. *Bioorg Med Chem* **20**, 1973-1978 (2012).
155. Bantscheff, M., Scholten, A. & Heck, A.J. Revealing promiscuous drug-target interactions by chemical proteomics. *Drug Discov Today* **14**, 1021-1029 (2009).
156. Medard, G. et al. Optimized Chemical Proteomics Assay for Kinase Inhibitor Profiling. *J Proteome Res* (2015).

157. Bantscheff, M. et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotechnol* **29**, 255-265 (2011).
158. Cravatt, B.F., Wright, A.T. & Kozarich, J.W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu Rev Biochem* **77**, 383-414 (2008).
159. Nomura, D.K., Dix, M.M. & Cravatt, B.F. Activity-based protein profiling for biochemical pathway discovery in cancer. *Nat Rev Cancer* **10**, 630-638 (2010).
160. Rix, U., Gridling, M. & Superti-Furga, G. Compound immobilization and drug-affinity chromatography. *Methods Mol Biol* **803**, 25-38 (2012).
161. Rix, U. et al. Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. *Blood* **110**, 4055-4063 (2007).
162. Brehmer, D., Godl, K., Zech, B., Wissing, J. & Daub, H. Proteome-wide identification of cellular targets affected by bisindolylmaleimide-type protein kinase C inhibitors. *Mol Cell Proteomics* **3**, 490-500 (2004).
163. Brehmer, D. et al. Cellular targets of gefitinib. *Cancer Res* **65**, 379-382 (2005).
164. Daub, H. et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol Cell* **31**, 438-448 (2008).
165. Wu, Z. et al. Quantitative chemical proteomics reveals new potential drug targets in head and neck cancer. *Mol Cell Proteomics* **10**, M111 011635 (2011).
166. Kruse, U. et al. Chemoproteomics-based kinome profiling and target deconvolution of clinical multi-kinase inhibitors in primary chronic lymphocytic leukemia cells. *Leukemia* **25**, 89-100 (2011).
167. Matthews, D.J. & Gerritsen, M.E. Targeting protein kinases for cancer therapy. (John Wiley & Sons, 2011).
168. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2015. *CA Cancer J Clin* **65**, 5-29 (2015).
169. Torre, L.A. et al. Global cancer statistics, 2012. *CA Cancer J Clin* **65**, 87-108 (2015).
170. Knapp, S. et al. A public-private partnership to unlock the untargeted kinome. *Nat Chem Biol* **9**, 3-6 (2013).
171. Melnikova, I. & Golden, J. Targeting protein kinases. *Nature reviews. Drug discovery* **3**, 993-994 (2004).
172. Fedorov, O., Muller, S. & Knapp, S. The (un)targeted cancer kinome. *Nat Chem Biol* **6**, 166-169 (2010).
173. Page, T.H., Smolinska, M., Gillespie, J., Urbaniak, A.M. & Foxwell, B.M. Tyrosine kinases and inflammatory signalling. *Current molecular medicine* **9**, 69-85 (2009).
174. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-1934 (2002).
175. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
176. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
177. Kurzrock, R., Kantarjian, H.M., Druker, B.J. & Talpaz, M. Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Ann Intern Med* **138**, 819-830 (2003).
178. Savage, D.G. & Antman, K.H. Imatinib mesylate--a new oral targeted therapy. *N Engl J Med* **346**, 683-693 (2002).
179. Stegmeier, F., Warmuth, M., Sellers, W.R. & Dorsch, M. Targeted cancer therapies in the twenty-first century: lessons from imatinib. *Clin Pharmacol Ther* **87**, 543-552 (2010).
180. Johnson, L.N., Lowe, E.D., Noble, M.E. & Owen, D.J. The Eleventh Datta Lecture. The structural basis for substrate recognition and control by protein kinases. *FEBS Lett* **430**, 1-11 (1998).
181. Baselga, J. Targeting tyrosine kinases in cancer: the second wave. *Science* **312**, 1175-1178 (2006).

182. Grant, S.K. Therapeutic Protein Kinase Inhibitors. *Cellular and Molecular Life Sciences* **66**, 1163-1177 (2008).
183. Zhang, J., Yang, P.L. & Gray, N.S. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* **9**, 28-39 (2009).
184. Liu, Y. & Gray, N.S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* **2**, 358-364 (2006).
185. Moffat, J.G., Rudolph, J. & Bailey, D. Phenotypic screening in cancer drug discovery - past, present and future. *Nature reviews. Drug discovery* **13**, 588-602 (2014).
186. Zheng, W., Thorne, N. & McKew, J.C. Phenotypic screens as a renewed approach for drug discovery. *Drug Discov Today* **18**, 1067-1073 (2013).
187. Lee, J.A., Uhlik, M.T., Moxham, C.M., Tomandl, D. & Sall, D.J. Modern phenotypic drug discovery is a viable, neoclassic pharma strategy. *J Med Chem* **55**, 4527-4538 (2012).
188. Anastassiadis, T., Deacon, S.W., Devarajan, K., Ma, H. & Peterson, J.R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1039-1045 (2011).
189. Karaman, M.W. et al. A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol* **26**, 127-132 (2008).
190. Davis, M.I. et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1046-1051 (2011).
191. O'Hare, T. et al. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell* **16**, 401-412 (2009).
192. Zinn, N., Hopf, C., Drewes, G. & Bantscheff, M. Mass spectrometry approaches to monitor protein-drug interactions. *Methods* **57**, 430-440 (2012).
193. Xiao, Y. & Wang, Y. Global discovery of protein kinases and other nucleotide-binding proteins by mass spectrometry. *Mass Spectrom Rev* (2014).
194. Patricelli, M.P. et al. Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **46**, 350-358 (2007).
195. Duncan, J.S. et al. Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. *Cell* **149**, 307-321 (2012).
196. Zhang, L. et al. Characterization of the novel broad-spectrum kinase inhibitor CTx-0294885 as an affinity reagent for mass spectrometry-based kinome profiling. *J Proteome Res* **12**, 3104-3116 (2013).
197. Pachl, F. et al. Characterization of a chemical affinity probe targeting Akt kinases. *J Proteome Res* **12**, 3792-3800 (2013).
198. Ku, X., Heinzlmeir, S., Helm, D., Medard, G. & Kuster, B. New Affinity Probe Targeting VEGF Receptors for Kinase Inhibitor Selectivity Profiling by Chemical Proteomics. *J Proteome Res* (2014).
199. Ku, X., Heinzlmeir, S., Liu, X., Medard, G. & Kuster, B. A new chemical probe for quantitative proteomic profiling of fibroblast growth factor receptor and its inhibitors. *J Proteomics* **96**, 44-55 (2014).

CHAPTER II

EVALUATION OF DATA INDEPENDENT ACQUISITION

ABSTRACT

Mass spectrometry is the key technology for large scale proteomics. The standard analytical approach for the identification and quantification of peptides/proteins is the coupling of reversed phase liquid chromatography to nano-electrospray ionization tandem mass spectrometry (LC-MS/MS). The commonly used strategy implemented in the operating software uses a data-dependent acquisition (DDA) approach for peak selection prior to fragmentation. In order to circumvent proposed conceptual limitations of DDA several data-independent acquisitions (DIA) methods have been established. Here the approaches MS^e and $HDMS^e$ on the Waters Q-TOF Synapt G2S have been evaluated and compared to the performance of a DDA approach using a LTQ-Orbitrap Elite. After the establishment of reliable database search parameters, it can be concluded that the performance of the G2S is inferior to the Elite and comparable with the LTQ-Orbitrap XL Nevertheless, it is shown that the resolution of co-eluting phospho peptide isomers benefits from additional gas phase separation.

INTRODUCTION

Mass spectrometry is the key technology for the performance of large scale proteomic analysis[1]. The current standard analytical approach for the identification and quantification of peptides/proteins is the coupling of reversed phase liquid chromatography to nano-electrospray ionization tandem mass spectrometry (LC-MS/MS)[2]. Advances in chromatographic and mass spectrometric technology are crucial for the analysis of ever more complex proteomes at ever-decreasing time and sample quantity requirements[3-6]. Despite advances in the available hardware, the commonly used strategy implemented in the operating software uses a data-dependent acquisition (DDA) approach. In DDA the precursors are selected based on the charge state and the intensity prior to serial fragmentation. The literature mentions conceptual limitations for DDA, like a stochastic and irreproducible selection of precursor ions for fragmentation compromising the reproducibility[7], undersampling[8] and prolonged instrument cycle times for deep proteome coverage[7]. In order to circumvent these limitations of DDA, different data-independent acquisitions (DIA) methods have been introduced. DIA methods do not attempt to isolate precursors based on any criteria prior to fragmentation rather co-fragmenting a group of precursors covering a distinct mass range. So far almost all mass spectrometer vendors have developed DIA methods – e.g. All-ion fragmentation (AIF, Thermo Scientific)[9], SWATH (ABISciex)[10] and MS^e (Waters Corporation)[11].

The “parallel fragmentation approach” was introduced by Clemmer and co-workers[12-15]. The group used an in-house LC-IMS-MS instrument architecture to acquire MS1 (precursor ion) and MS2 (fragment ion) data across the complete mass range. During the experiment the instrument is cycling between two scanning modes. A “low energy” mode for acquiring the intact precursor information and a “high energy” mode to record all fragments deriving from the parallel fragmented precursors. As the data are recorded without any ion selection, the precursor and fragments are paired retrospectively. This can be based on the precursors’ and the fragments’ similar profiles in retention time and ion mobility[7, 14]. The MS^e approach by Waters Corporation is based on the principles of the “parallel fragmentation approach”, using the retention time profile for the precursor and fragment correlation[11](Fig. 1). A new generation of Synapt Q-TOF mass spectrometers launched by Waters recently incorporated a TWIMS (travelling wave ion mobility separation) device. This enabled further ion separation based on ion mobility and is still compatible with subsequent all ion measurement.

Compatibility is here achieved due to the suitable differences in timescales between LC (min/s), IMS (low ms range) and TOF (µs) separations. The implementation of TWIMS into the Synapt MS instrument enables the use of MS^e in combination with ion mobility, in a hybrid method called $HDMS^e$ (Fig. 1).

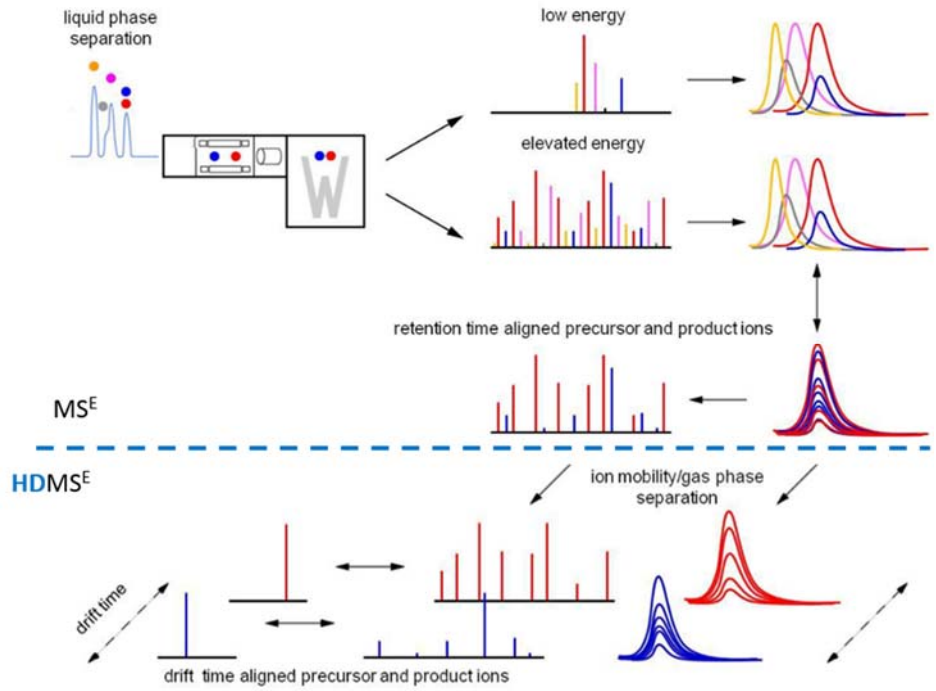


Figure 1 Data independent acquisition Principles underlying the DIA approaches MS^e and $HDMS^e$. During both methods the mass spectrometer records a "low energy" trace containing precursor information and a "high energy" trace containing fragment ion data. After the acquisition the software PLGS matches a precursor ion with its fragments based on a retention time alignment (MS^e) and additional drift time alignment ($HDMS^e$). (adapted from Waters Corporation presentation, personal communication).

ProteinLynx Global Server (PLGS) is the Waters proprietary software tool for processing MS^e and $HDMS^e$ data[11](Fig. 2). Two algorithms, Apex3D and Pep3D, are used by PLGS in order to process the raw data: I) Apex3D is responsible for ion detection, noise subtraction, mass correction and determining the chromatographic elution and ion mobility profiles for each feature. II) Pep3D creates tables of precursor and fragment masses organized in EMRTs (exact mass and retention time). An EMRT represents a peptide of unknown sequence and is created by merging isotopes and charge states of common peptides based on the mass, intensity, retention and, in the case of $HDMS^e$, drift time. At this stage fragment ions can be assigned to more than one precursor.

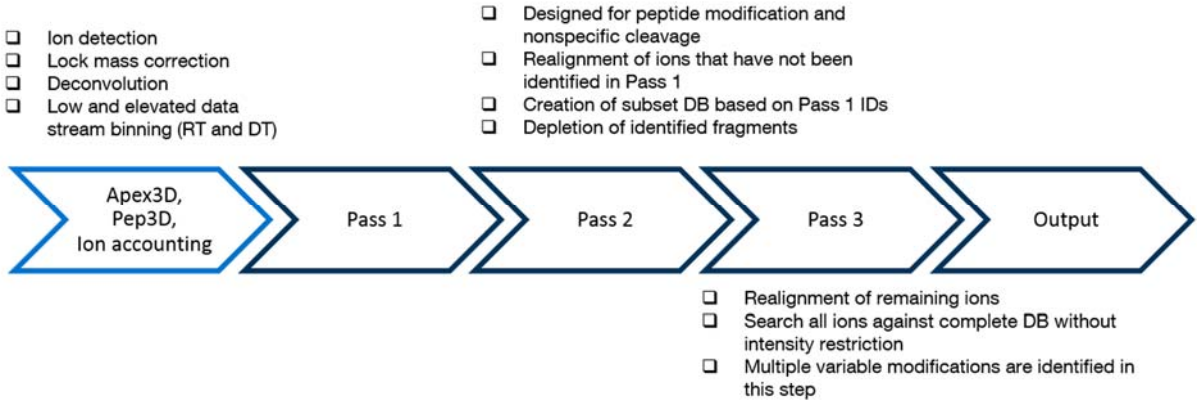


Figure 2 Data processing PLGS Schematic comprehension of the processes involved in the data processing (blue arrow) and the database search (black arrows) in ProteinLynx Global server. Apex3D and Pep3D are used for raw data processing: Apex3D is e.g. peak detection. Pep3D creates EMRTs based on retention time (RT) and drift time (DT) profiles. The Ion accounting algorithm is performing the database search in three stages: I) EMRTs that match completely cleaved tryptic peptides belonging to the most confident protein identifications are iteratively removed from the data. II) A sub database consisting of the proteins identified in pass 1 is the basis of the second pass. Now peptides can contain missed cleavages, variable modifications, neutral losses and in-source fragments. III) For the final iteration of the search, the whole initial database is searched again and multiple modifications and in-source fragments are examined.

Beforehand the database search (by the Ion accounting algorithm), a decoy database is merged with the forward database. The database search is considered as protein centric and can be divided into three stages (passes) (Fig. 2): I) EMRTs that match completely cleaved tryptic peptides belonging to the most confident protein identifications are iteratively removed from the data. Fragment ions (b and y) connected to the sequence of the identified EMRT are removed from the data set. If the specified protein FDR (false discovery rate) is reached the search is halted. Although the peptide spectrum matches are ranked by the strength of spectrum to sequence match, the FDR is exclusively calculated on the protein level. II) A sub database consisting of the proteins identified in pass 1 is the basis of the second pass. Now peptides are allowed to contain missed cleavages, variable modifications, neutral losses and in-source fragments. III) For the final iteration of the search, the whole initial database is searched again and multiple modifications and in-source fragments are examined.

Here the utility of DIA acquisition methods available on a Synapt G2S Q-TOF mass spectrometer for chemical and shot gun proteomics are critically evaluated as well compared to the performance of hybrid LTQ-Orbitrap instruments. Further the capabilities of the PLGS search algorithm for protein identification and PTM assignment are investigated. Last, deeper insight is gained into the applicability of TWIMS for the separation of positional phosphopeptide isomers.

MATERIAL AND METHODS

SAMPLE PREPARATION

PREPARATION OF THE *E.COLI* DIGEST

A commercially available *E.coli* standard (Waters Corp., Manchester, UK) was diluted to the desired concentration using LC grade water with 0.1% formic acid.

PREPARATION OF THE BSA DIGEST

A commercially available protein digest standard of bovine serum albumin (BSA) (Michrom Bioresources, CA, USA) was diluted to the desired concentration using LC grade water with 0.1% formic acid.

PREPARATION OF THE HELA DIGEST

HeLa S3 cervix carcinoma cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) with high glucose (PAA, Pasching, Austria) supplemented with 10% (v/v) fetal bovine serum (FBS, PAA, Pasching, Austria) at 37°C in humidified air and 10% CO 2. Cells were washed with phosphate buffered saline (PBS) and harvested by lysis using 50mM Tris/HCl pH 7.5, 5% Glycerol, 1.5 mM MgCl2, 150 mM NaCl, 0.8% NP-40, 1 mM dithiothreitol and 25 mM NaF with freshly added protease inhibitors and phosphatase inhibitors (5x phosphatase inhibitor cocktail 1, Sigma-Aldrich, Munich, Germany, 5x phosphatase inhibitor cocktail 2, Sigma-Aldrich, Munich, Germany, 1 mM sodium ortho-vanadate and 20 nM Calyculin A, LC Laboratories, Woburn, MA, USA). Protein extracts were clarified by ultracentrifugation for 1 h at 145,000xg at 4°C and protein concentration was determined by the Bradford method. Lysates were diluted in 8 M urea, 0.1 M Tris/HCl followed by protein digestion with trypsin (Promega Corp., Madison, WI, USA) according to the FASP protocol[16]. After overnight digestion, peptides were eluted from the filters with 50 mM TEAB and purified on C 18 StageTips as described[17]. The desired experimental concentration was reached by dilution using water and 0.1% formic acid.

The design and the synthesis of the phospho library for the assessing the potential of HDMS^{e} to identify PTMs is detailed described in Marx et al. (2013)[18]. In short, each of the so called 96 “seed” peptides represents a permutation template for the generation of one of the 96 libraries. The applied permutation scheme incorporates (S, T, Y, pS, pT, pY) at the position of the phospho site (X 0)(in the original peptide from the literature) and the 20 standard amino

acids in the direct vicinity (X -1 and X +1) resulting in libraries of the sizes of 2,400 peptides (n=84) or 120 peptides (n=12) in case the phosphorylation site is either at the peptide N-terminus or directly N-terminal to the C-terminal Lys and Arg residues of the tryptic peptides. The here used library was based on the peptide L-L-L-X-$_{-1}$-X$_0$-X$_1$-E-T-K. The combinatorial libraries were synthesized by standard solid-phase synthesis following the Fmoc strategy on a parallel peptide synthesizer (Intavis, Cologne, Germany). Briefly, the synthesis began with a C-Terminal tryptic amino acid (R, K) and then proceeded sequentially to concatenate single amino acids, except at the permutation site(s), where isokinetic mixtures of amino acids (20 or 6) were incorporated to attempt to create a discrete uniform distribution. The crude synthetic peptide library with the positon A4 on the 96 well plate (containing the whole 96 libraries) was subjected to the analysis with HDMS^e without further purification.

The design and the synthesis for the positional phosphopeptide isomers for evaluating the benefit of further gas phase separation has been previously described [19]. In brief, based on a list of naturally occurring phospho peptides[20], 180 peptides including positional p-site isomers were synthesized individually by solid-phase synthesis on a parallel peptide synthesizer (Intavis, Cologne, Germany) following the standard Fmoc strategy (see Appendix for a list of chosen phospho peptide isomers). Crude peptides used for the HDMS^e experiments without further purification.

The used Kinobead sample was generated as described in Bantscheff et al (2007)[20] and Medard et al. (2015)[21]. In short, lysates of a cell mix (MV4-11,K562, SK-N-BE-2,Colo 205) were diluted with equal volumes of 1x compound pulldown (CP) buffer (50 mM Tris/HCl pH 7.5, 5% glycerol, 1.5 mM MgCl 2, 150 mM NaCl, 25 mM NaF, 1 mM dithiothreitol and freshly added protease inhibitors and phosphatase inhibitors (5x phosphatase inhibitor cocktail1, (Sigma-Aldrich, Munich, Germany), 5x phosphatase inhibitor cocktail 2, (Sigma-Aldrich, Munich, Germany), 1 mM sodium ortho-vanadate and 20 nM Calyculin A, (LC Laboratories, Woburn, MA, USA)). Lysates were diluted to a final protein concentration of 5 mg/ml using 1x CP buffer supplemented with 0.4% NP-40. Kinobeads γ[21] (100 µL settles beads) were incubated with lysates (total of 5 mg of protein) for 1 h at 4°C. Subsequently, beads were washed with 1x CP buffer and collected by centrifugation. Bound proteins were eluted with 2x NuPAGE LDS Sample Buffer (Invitrogen, Darmstadt, Germany) and eluates were reduced and alkylated by 50 mM dithiothreitol and 55 mM iodoacetamide. Samples were then run into a 4–12% NuPAGE gel (Invitrogen, Darmstadt, Germany) for about 0.5 cm to concentrate the sample prior to in-gel tryptic digestion. In-gel trypsin digestion was performed according to standard procedures.

LC-MS CONFIGURATIONS

The peptides were separation was performed using a nanoAcquity UPLC system (Waters, Manchester, UK). The mobile phase A was 0.1% formic acid in LC grade water. The mobile phase B was acetonitrile with 0.1%formic acid. The samples were trapped on Symmetry C_{18} 5 µ m, 180 µm × 20 mm precolumn (Waters, Manchester, UK) and desalted at 5 µ L/min flow for 3 min with 99.9% A. The separation was performed on BEH130 1.7 µm, 75 µm × 250 mm (Waters, Manchester, UK) at 300 nL/min flow rate ,with 1 – 40% B, using gradients of varying length from 15-210 min. After separation, the column was washed with 80% mobile phase B for 5 min and re-equilibrated with 1% mobile phase B for 15 min. The column temperature was maintained at 45° C. 100 fmol/µL [Glu1] - fibrinopeptide B was infused at 800 nL/min for lock mass correction, which was only recorded but not applied during the acquisition and later applied in the software. Mass spectrometric analysis of eluting peptides was performed on Synapt G2S mass spectrometer (Waters, Manchester, UK). The Synapt G2S was operated in a way that low and high energy scans of equal duration of 0.5 s alternated and data are acquired over *m/z*-range from 50-2000. Lock spray was acquired once every30 s for 1 s period. For fragmentation in MS^E mode a collision energy ramp was used in the Trap region of the TriWave from 15 to 45 V across the duration of the high energy scan further a transfer collision of 10 V was applied. For improved ion transfer the collision energy during the low energy scan was set to 4 V/2 V in the trap and transfer region.

Mass spectrometric analysis was performed using a Synapt G2S HDMS mass spectrometer (Waters, Manchester, UK). The MS was operated in “V-mode” with a typical resolution 25,000 FWHM. All analyses were performed in positive ESI mode. The data were post-acquisition lock mass corrected using the doubly charged monoisotopic ion of [Glu1]-Fibrinopeptide B – 785.8427 m/z. The reference signal was acquired every 30 s. The system was operated in DIA mode. For IMS, a wave height of 40 V was applied. Traveling wave velocity was ramped from 1000 m/s to 300 m/s over the full IMS cycle. Wave velocities in the trap and transfer cell were set to 311 m/s and 380 m/s and wave heights to 4 V. One IMS cycle consist of 200 bins (200 orthogonal acceleration pushes) and takes 200 times t_{pp}(pusher period) of time. The here used ion mobility settings in combination with an *m/z*-range from 50-2000 result in one “bin” being 69 µs (t_{pp})[22]. The spectral acquisition time in each mode was 0.5 s. In low-energy MS mode, data were collected at constant CE of 4 eV. A CE ramp from 15 eV to 45 eV and 19 eV to 50 eV during each 0.5 s was used as standard setting for the elevated energy scan in MS^E and $HDMS^E$ mode, respectively. The differences between the CE ramps used in MS^E and $HDMS^E$ are due to the instrument design of the Synapt G2S. When IMS separation is

activated (HDMS^E), the IMS cell of the Triwave region confronted with 1 mbar N_2, which also diffuses into the transfer cell region of the Triwave device, thus lowering the argon concentration, which is used for collision. Since the mass of the N_2 molecules is lower than that of the Ar atoms, a higher CE is required when IMS is activated. The quadrupole mass analyzer was adjusted to a LM resolution of 4.7 and a HM resolution of 15 to ensure optimized ion transmission.

DATA ANALYSIS

DIA MS data were processed and searched using ProteinLynx Global SERVER version 3.1 (PLGS, Waters,Manchester, UK). Processing was performed using lock mass correction, noise reduction thresholds for low energy scan ion, high energy scan ion and peptide intensity combined across charge states and isotopes were optimized for the present detector voltage. Protein identifications were obtained by searching the databases IPI human database (v3.68, HeLa samples and phospho peptide isomers), a Uniprot E.coli K12 database (9.691 entries, E.coli samples) or the customized database IPI human v3.72 including the sequences of all 96 peptide libraries[18] .The precursor and fragment-ion mass tolerances automatically determined. Further, the following criteria were considered: Trypsin as digestion enzyme, up to two missed cleavages allowed, fixed modification was carbamidomethyl cysteine (for the BSA sample carboxymethyl was used) and variable modification was methionine oxidation. For standard searching the vendor recommended settings for peptide and protein identification were used: A minimum of one identified fragment ions for a peptide, minimum of 3 fragments for a protein and 1 peptide for a protein identification. The false discovery rate (FDR) for protein identification was assessed by searching a reversed sequence database and set to 1%. For using the Mascot search engine (v. 02.04.2001) the peak list was exported after PLGS processing and search using the same settings as in PLGS using a precursor mass tolerance of ±25 ppm and a fragment mass tolerance of ±0.1 Da. For FDR filtering the result file was imported in to Scaffold (v. 4.1.1) and set to 1% protein FDR.

The data analysis of the data from the LTQ-Oribtrap Elite was performed using MaxQuant (1.2.7.3) and has been previously described[23].

Data illustration and statistical analysis were performed using GraphPad Prism 5 (v. 5.03).

RESULTS

INTEGRATION OF TRAVELLING WAVE ION MOBILITY FOR DIA IN PROTEOMICS STUDIES

On the G2S two DIA methods are available – MS^e and HDMS^e. In order to elucidate the benefits of incorporating TWIMS into the MS^e workflow 100 ng of an *E.coli* digest were separated via a 60 min LC gradient and either analysed with (HDMS^e) or without (MS^e) additional ion mobility separation. Data were analysed using Waters recommended settings for database search and filtered to 1% Protein FDR.

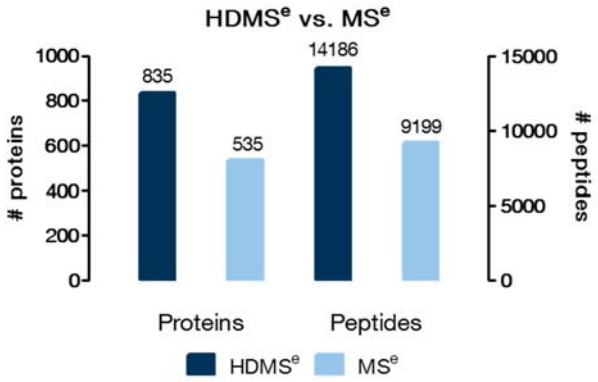


Figure 3 Performance comparison MS^e and HDMS^e A sample of 100 ng *E.coli* digest have been analysed using MS^e (light blue) and HDMS^e (dark blue) during a 60 min LC gradient. The data have been analysed using PLGS standard settings. The increase in performance can be attributed to the increased peak capacity of the system through the incorporation of the IMS device.

A general increase in performance could be observed. In more detail, HDMS^e increased the number of identified proteins from 535 (MS^e) to 835, a gain of 56% (Fig. 3). The number of identified peptides was also improved by 54% (14,186 HDMS^e and 9,199 MS^e), respectively.

Displaying the peptide separation by retention time and drift time from 100 ng *E.coli* obtains Fig. 4A. Here vertically aligned areas represent peptides that had similar retention time in the liquid chromatography but different drift times in the ion mobility.

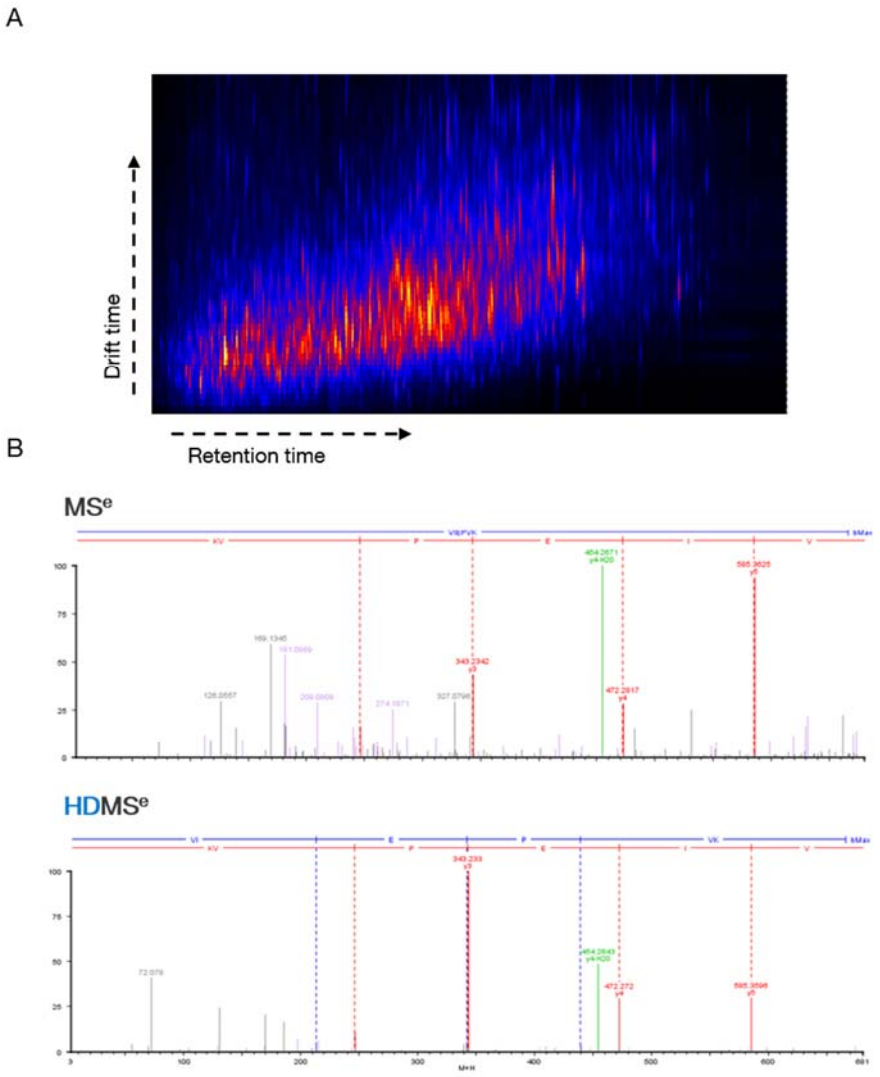


Figure 4 Increased peak capacity by IMS (A) Shows the separation effect achieved by ion mobility. Vertically aligned signals represent features that could be resolved by drift time but not by retention time. (B) Spectrum of the peptide (R)VIEPVK(R) (683.43 Da) of the *E. coli*. tryptophanase measured with MS^e and with HDMS^e. Showing the improved spectrum quality by incorporation of ion mobility.

On the spectrum level, this lead to a "clearance" of the resulting reconstructed and identified MS2 spectra for peptide/protein identification, exemplified here with the spectrum of the peptide (R)VIEPVK(R) (683.43 Da) of the *E. coli*. Tryptophanase (Fig. 4B).

Since the instrument is lacking a system to control the ion flux (available in the newest version G2Si) onto the detector, it is likely to saturate the detector which leads to disturbed isotopic patterns and to the appearance of "ghost peaks". It was observed that this detector saturation is occurring at a signal intensity of approx. 1e6 for MS^e whereas at 1e5 for HDMS^e. Comparing the base peak intensity (BPI) chromatograms (Fig. 5A) of e.g. the *E. coli* samples analysed with MS^e to HDMS^e showed a reoccurring signal attenuation in the BPI of the HDMS^e

run of approx. one-third. The effect of this observation on protein quantitation was further investigated (Fig. 5B).

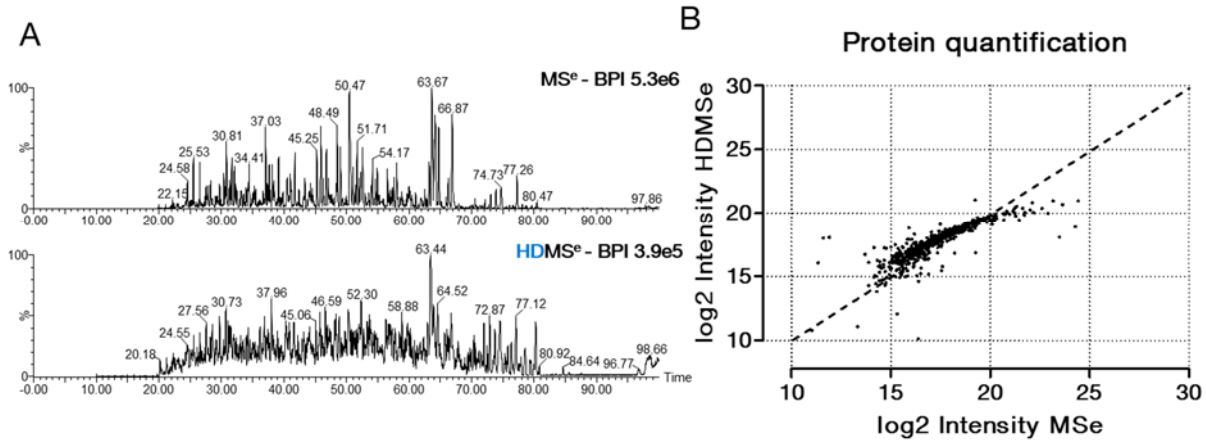


Figure 5 Protein quantification (A) Shows the base peak intensity (BPI) chromatograms of a 100 ng sample of E.coli analysed with MS^e and with HDMS^e. Indicating the signal attenuation effect of the ion mobility. (B) Correlating the obtained label-free protein quantification results from both methods. Coherent with the in (A) observed effect of signal attenuation, resulting in an underestimation of high abundant proteins and a tendency of overestimating low abundant proteins.

The graph reveals that both methods do not lead to similar quantitative results. In more detail it can be seen that there is a tendency to underestimate the amount of high abundant proteins in the HDMS^e results. On the other end the amount of low abundant proteins seems to be overestimated compared in HDMS^e. In addition, it is notable that depending on the setting for the detector voltage, which needs to be checked on a weekly basis, the processing parameter in the software can vary quite drastically. Demanding adjustment after every change on the detector voltage.

Due to the performance increase by incorporation of ion mobility in the DIA workflow, in the following sections the focus was set on the HDMS^e acquisition scheme.

EVALUATION OF DIA DATA PROCESSING USING PROTEINLYNX GLOBAL SERVER

The claimed advantage of DIA acquisition methods compared to DDA is the ability to record all precursors (above LOD) of a sample and all fragments over the whole acquisition time. In DDA the principle assumption is that all fragment ions in a recorded MS/MS spectrum originate from the previously selected precursor. This “specificity” is lost during a DIA acquisition. In consequence the main purpose of software processing DIA data is to re-establish the source (precursor) of a fragment prior to database search. In PLGS the Apex3D

algorithm is responsible for the family reunion of the precursor and its fragment ions, resulting in EMRTs. In a nutshell, the fragments from the "elevated energy" trace are stepwise associate with the intact precursor masses from the "low energy" trace based on their analogous retention and drift time profiles. The algorithm has fixed windows for these alignment procedure. For the retention time alignment it considers ions that have an apex that is within 1/10 of the chromatographic peak width, whereas the apex of the drift time profile is narrowed to two bins. If one fragment fits these parameters for more than one precursor, it is associated to all of them. In this part, the performance of the proprietary Waters software PLGS and its underlying algorithms for the processing of a set of different HDMS^e data sets was evaluated.

PHOSPHOPEPTIDE IDENTIFICATION AND -SITE LOCALISATION

In 2013 Marx et al.[18] synthesized a large synthetic peptide and phospho peptide library containing 96 individual libraries, together yielding >100,000 synthetic modified or unmodified peptides. Due to the fact that the sequence of each synthetic peptide plus it's site of modifications in the library is known, this sample enabled us to address the question how well PLGS is able to identify phosphorylated peptides and its ability to localize the phosphorylation site. For the purposes of this analysis it was sufficient to use only one out of the 96 libraries (Nr. 4). The underlying "seed" peptide had the following sequence: L-L-L-X-$_1$-X_0-X_1-E-T-K. The positions X-$_1$ and X_1 can here be permutated with all 20 naturally occurring amino acids. In contrast X_0 was permutated to contain only the amino acids S, T or Y either modified or unmodified by phosphorylation. Theoretical this would lead to 2,400 peptides in this library. An initial analysis in the publication using an LTQ-Orbitrap Velos (HCD) revealed a successful synthesis of approx. 1,000 non-redundant sequences with 50% being phosphorylated and 50% being unmodified.

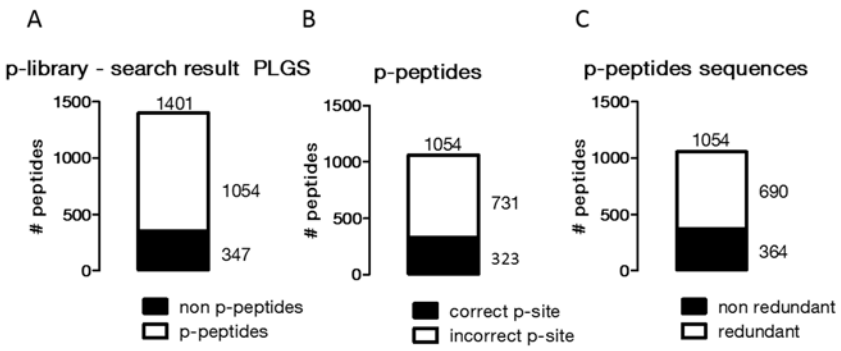


Figure 6 Phospho peptide identification and site localisation (A) The search result of PLGS identifies in total 1401 peptides from the phospho peptide library A4. 347 are non-phosphorylated and 1054 are containing a phosphorylation. (B) Of the 1054 phospho peptides from (A) 731 are assigned with an incorrect sight of phosphorylation and 323 are assigned with the correct sight of phosphorylation. (C) Filtering the 1054 phosphorylated peptides from (A)/(B) for sequence redundancy revealed 364 non redundant sequences and 690 redundant sequences.

For this study an amount of 500 pmol of the library was analysed over a 90 min gradient (Fig. 6) and processed using recommended settings from Waters (details next section). Overall (Fig. 6A) PLGS was able to identify 1401 peptides of which it assigned ca. 350 (25%) as non-phosphorylated and 1054 (75%) as phosphorylated peptides. As previously mentioned the advantage of a synthetic library is the content certainty. Consequently it can be seen that of the 1054 phospho-peptides (Fig. 6B) for only around 300 the site of phosphorylation was correctly localised (30%). Creating a non-redundant list of peptide sequences from the 1054 peptides annotated by PLGS as phosphorylated (Fig. 6C) shows that about 35% are non-redundant. This indicated that PLGS was in almost all cases listing the entire possible phospho peptide isomers for a distinct sequence.

INFLUENCE OF DATABASE SEARCH PARAMETER

Prior to the database search, the user is able to specify criteria in PLGS for peptide and protein identification (in bracket is the standard setting recommended by Waters), e.g. fragment matches per peptide (1), fragment matches per protein (3) and peptides for protein identification (1). A manual inspection of the peptide/protein identifications for a HeLa digest revealed several implausible identifications (Fig. 7) e.g. the peptide with the sequence KPTYDVSEDQDPLSSDFKR from the CDC42 kinase 1 (gene: TNK2).

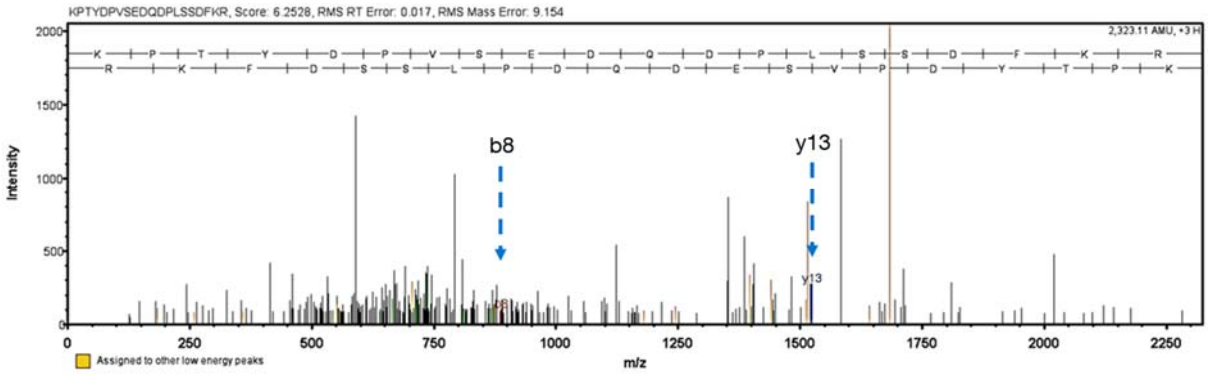


Figure 7 Questionable peptide ID by PLGS Shows the MS/MS spectrum assigned by PLGS to the peptide KPTYDVSEDQDPLSSDFKR from the CDC42 kinase 1. The identification is based on the b8 ion and the y13 highlighted by the blue arrow. The yellow peaks have been assigned to other peptide identifications and the black peaks remained unassigned.

In this case the identification of the whole peptide is based on a b8 and y13 ion, which poses only weak intensity compared to the overall spectrum. Further a number of more dominant peaks in this MS2 spectrum cannot be correlated to these or any other sequence in the database (which would be indicated by a yellow highlighting of the peak).

In order to investigate if an identification may be plausible by this few ions, the standard deviation of the drift time over 4 *E.coli* samples was used as an estimate for certainty. Since the mobility of a peptide-ion is only dependent on inherent properties of the ion (*m/z*, charge, collision cross section), the drift time should be stable for the same peptide in different analysis. In Fig. 8 the correlation between the average number of ions, that were used for the identification of a certain peptide (average over 4 runs), and the standard deviation of the drift time was plotted.

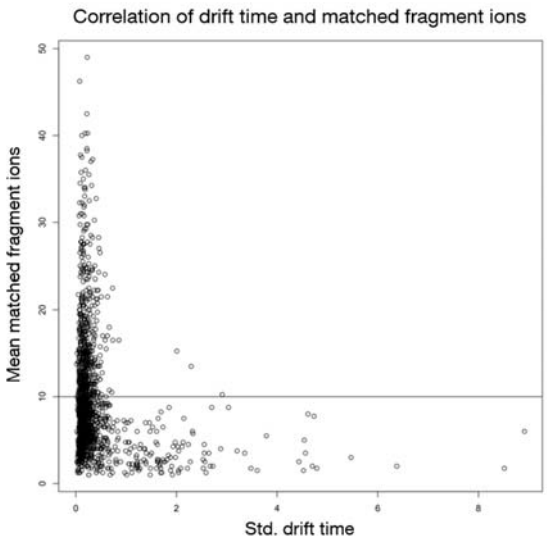


Figure 8 Correlation of drift time and number of matched fragments The correlation is based on 4 runs of 100 ng E.coli. The peptides have been matched based on the sequences identified by PLGS. It can be seen that the distribution of the standard deviation of the drift time is quite narrow until the mean of matched fragment ions gets under 10 ions.

It can be seen that the variations in the standard deviation of the drift time increase with fewer matched fragments. Especially when less than 10 ions are used for the identification. Translated, this means that by using less than 10 ions for an identification, the probability in identifying the same peptide decreases.

With the purpose to establish a reference for the number of fragments that are needed to confidently identify a peptide, two data sets from a LTQ-Orbitrap Elite[24] (HCD and CID) of 100 ng HeLa on a 60 min gradient were used[23]. The data have been analysed with the free software package MaxQuant[25](Fig. 9). The Protein FDR in this program was set to 1% and no further filtering on peptide level was applied.

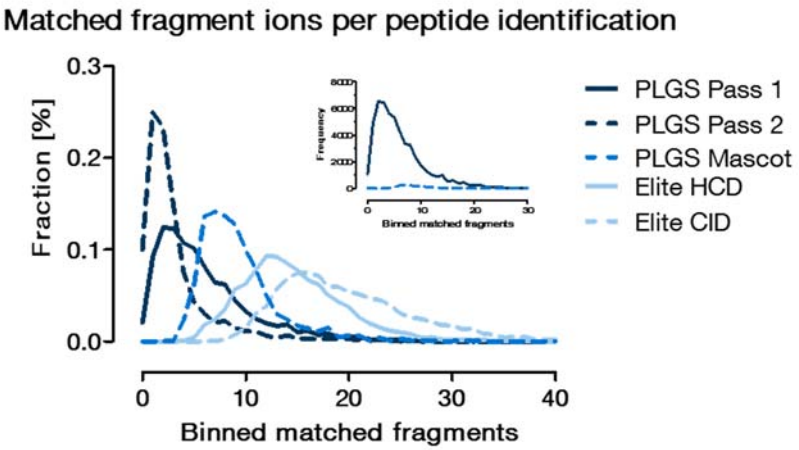


Figure 9 Matched fragment ions per peptide identification The figure shows the frequency distributions of the binned number of matched fragment ions for peptide identifications for a 100 ng HeLa sample from the G2S analysed by PLGS (black lines), the peak list from PLGS searched with the Mascot search engine (dark blue dashed line) and measured on a LTQ-Orbitrap Elite with HCD (light blue line) or CID (light blue dashed line) processed with MaxQuant. The data are shown as percent of the single data set. The interest shows the same data for PLGS Pass 1 and PLGS Mascot but in absolute numbers.

Interestingly there is a difference in the distributions between both acquisition methods used on the LTQ-Orbitrap Elite. Although both distributions look almost identical, the maximum for CID is shifted towards a higher number of fragments compared to HCD. In summary, the Elite is identifying 75% of its peptides with at least 17 or 23 fragments in HCD and CID respectively, whilst not identifying anything with less than 4 or 7 ions. Correlating this to the data gained from the PLGS Pass 1 and Pass 2 searches (100 ng HeLa, 60 min gradient) exposed a significant difference. Both graphs are drastically shifted towards low numbers with the maxima being far under 10 fragment ions per identified MSMS. Pass 1 e.g. bases 75% of its identifications on 8 or less fragments. Prominent is also a tendency of the Pass 2 search results to reveal peptide identifications with even less ions compared to Pass 1 (75% percentile 4 fragments or less). The peak list from PLGS was exported and searched with the Mascot search engine. The resulted distribution lay between that of the Elite MaxQuant and the G2S PLGS data, with 75% of the data based on 11 or less ions with a minimum of 3 ions per peptide identification. The insertion shows the total frequency of the peptide identifications between Pass 1 and the Mascot results both founding on the same peak list. It can be seen that only a marginal amount of the spectra could be identified from the Mascot search engine compared to PLGS Pass 1.

Table 1 Database search parameter and protein/peptide ID

Fragments			
per peptide	per protein	Proteins	Peptides
1	3	2337	27597
3	3	1886	22215
3	7	1783	24910
5	5	1465	16368
5	7	1223	16467
5	10	1097	16183
10	10	461	5605
Mascot	Search	674	2194

How adjusting the database search parameters influences the performance for protein and peptide identifications can be seen in Table 1. The effect is significant, considering that increasing the number of required ions for a peptide identification from 1 to 10 lowers the number of identified proteins as well as peptides by a factor of 5 for the 100 ng HeLa sample. During all processing iterations the minimum peptide number for a confident protein identification was kept at 1. The Mascot search engine identified 674 proteins and 2194 peptides, only being able to identify 1.9% of the 98,016 searched spectra (1% Protein FDR in Scaffold).

PEPTIDE SCORING IN PLGS

PLGS is considered as a protein centric search program. Thus the FDR is only calculated on the protein level. Nevertheless, during the analysis the identified peptides receive a score that represents the strength of the peptide-spectrum match. In this section we investigate the applicability of the PLGS scoring system in order to better understand the identifications made.

The scoring in PLGS is i.a. based on a set of physico-chemical models and information, e.g. accurate product ion masses, total product ion intensity, number of matched product ions conform to model, and neutral losses conforming to amino acid composition.

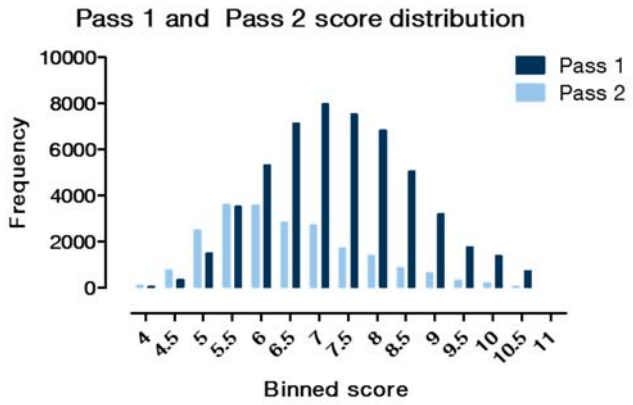


Figure 10 Score distribution of Pass 1 and Pass 2 The search results from the Pass 1 and the Pass 2 search of PLGS have been binned according to the received score.

In a first iteration, the score distributions between the Pass 1 and the Pass 2 results of the ion accounting algorithm (100 ng HeLa) was investigated (Fig. 10). This showed that the results of Pass 1 and Pass 2 formulate two different distributions, with Pass 1 having a trend of higher scoring results centring around a score of 7 whereas the Pass 2 score distribution having a apex around 5.5.

In addition, the potential correlation between the number of matched fragments for an identification and the corresponding score for the peptide-spectrum-match was investigated (Fig. 11A). Only Pass 1 peptides were used for this analysis. As a reference, the same was plotted for the Andromeda[26] search engine of the MaxQuant software (Fig. 11B).

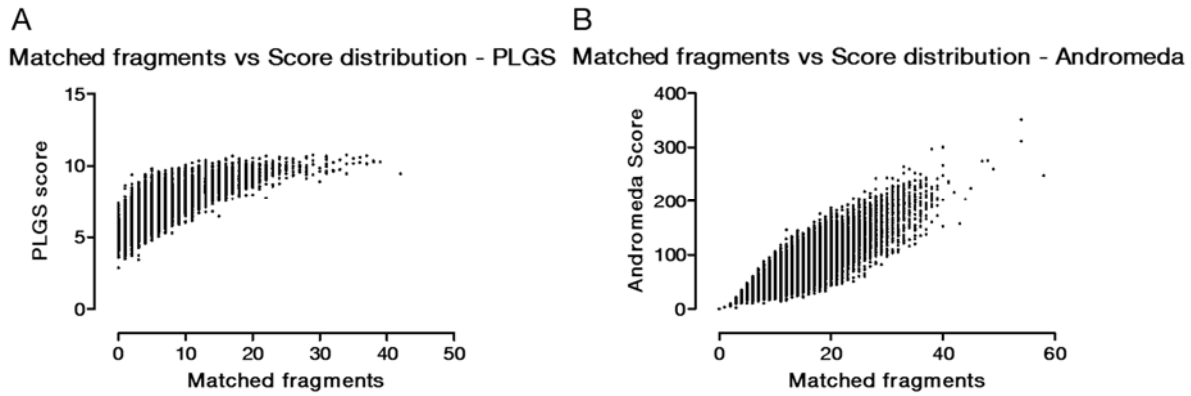


Figure 11 Number of matched fragments vs. Score (A) Shows the correlation between the number of matched fragment ions and the obtained score from the PLGS search engine (Pass 1 only) for a 100 ng HeLa sample search with the standard settings. (B) Same as (A) but for 100 ng HeLa measured on a LTQ-Orbitrap Elite in HCD mode.

The PLGS scoring for pass 1 peptides shows no clear correlation between the number of matched fragment ions and the score given to an identified peptide from an EMRT. Moreover, a peptide identified with e.g. 20 fragments can be given the same score as one identified with 1 fragment.

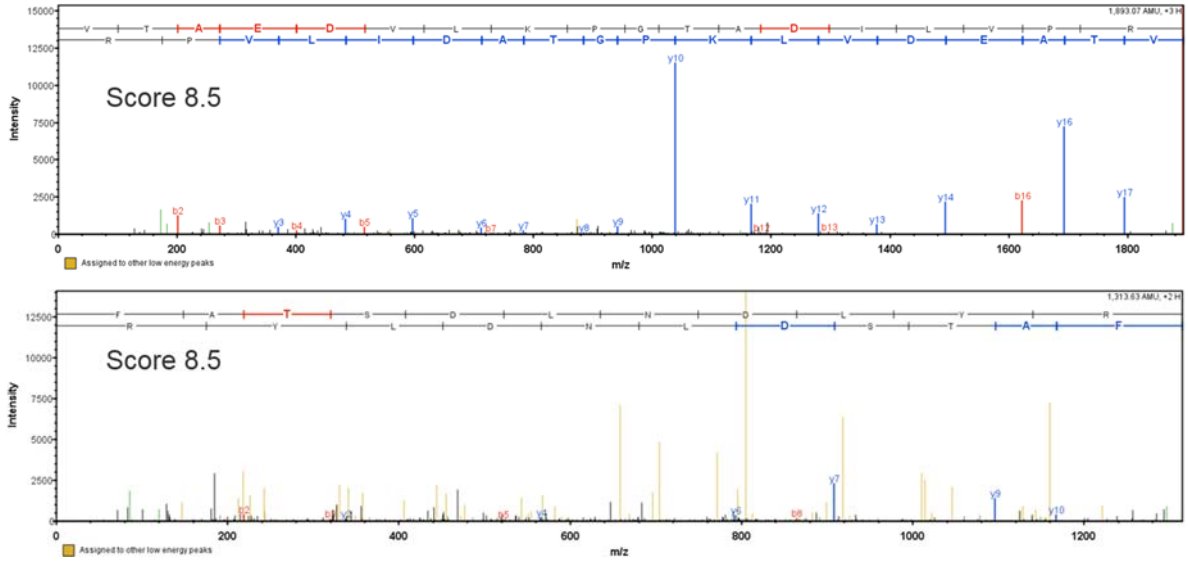


Figure 12 PLGS scoring Two spectra from different peptides are depicted that show obvious differences in quality but have been assigned with the same score by PLGS. Blue and red peaks indicate annotated b and y ions. Yellow peaks show fragments that have been assigned to another peptide.

This observation can be further supported by the two example spectra in Fig. 12, which show a visible difference in quality but were assigned with the same score of 8.5. The picture changes when reviewing this correlation for the Andromeda score. Here a linear correlation between the number of matched fragments and annotated score is visible (Fig. 11B).

CLASSIFICATION OF THE DIA PERFORMANCE IN COMPARISON TO LTQ-ORBITRAP INSTRUMENTS

In order to rank the performance of HDMS^e on the Synapt G2S the results were mainly compared to results from the characterization of a state of the art LTQ-Orbitrap Elite[23].

The pure sensitivity of an instrument can be determined by the measurement of a single protein digest dilution series e.g. bovine serum albumin (BSA, Table 2). Both, MS^e and HDMS^e, show the same limit of detection (LOD) of 100 amol BSA digest on column. Compared to the predecessor of the G2S, the G2 (1 fmol, HDMS^e) the LOD improved by one order of magnitude. Linking the results of the G2S to two generations of LTQ-Orbitrap

instruments (XL and Elite), it can be seen that the LOD of the Orbitrap XL is at an amount of 50 amol BSA on column as the LOD for the Elite could be determined with <1 amol.

Table 2 Sensitivity of the G2S in comparison to the G2, the LTQ-Orbitrap XL and LTQ-Orbitrap Elite

	G2-S		G2	Orbitrap XL	Orbitrap Elite
	HDMS^e	MS^e	HDMS^e	DDA	DDA
LOD	100 amol	100 amol	1 fmol	50 amol	<1 amol

Considering the evaluation results of the data processing with ProteinLynx Global Server, the processing in the rest of this section was performed requiring 10 fragment ions for peptide and protein identifications. For the purpose of elaborating the performance in conjunction with a more complex sample, a dilution series with a complex HeLa digest was performed. Ranging from 100 pg to 1000 ng digest on column (Fig. 13 only showing data 1 – 1.000 ng).

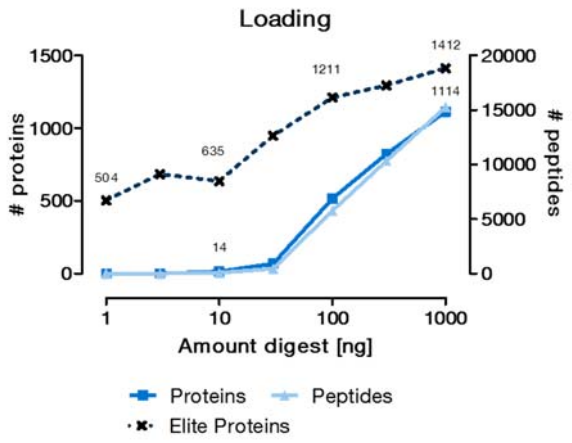


Figure 13 HeLa dilution series Different dilutions of a complex HeLa digest have been measured on the G2S (blue and back lines) and the Elite (dashed black line). The dilution series was ranging from 0.1 ng up to 1.000 ng. Data are only displayed from 1 ng to 1000 ng due to the lack of IDs on the G2S.

It can be seen that the LOD for the G2S for the complex sample is at 10 ng (14 proteins) HeLa digest on column whereas the Elite is already able to detect 500 proteins at 100 pg (data not shown)[23]. A linear correlation between sample loading and an increase in identified proteins and peptides is visible. Ranging from 14 proteins at 10 ng up to 1114 proteins at 1.000 ng sample amount. An additional notation is that a loading test analysed (100 – 1000 ng) using MS^e did not reveal any protein identification for 300 and 1000 ng. The performance of the Elite is at any data point superior to the G2S in HDMS^e mode.

For DDA experiments it is known that extending the measurement time can lead to an improvement in protein identification. Also higher acquisition speeds enable shorter gradient

times with a constant sampling depth[3]. Thus the next question that was addressed, was the impact of HDMS[e] on the sample throughput and depth of proteome coverage (Fig. 14). Here we applied different analysis times for 100 ng of HeLa digest ranging from 15 min up to 210 min.

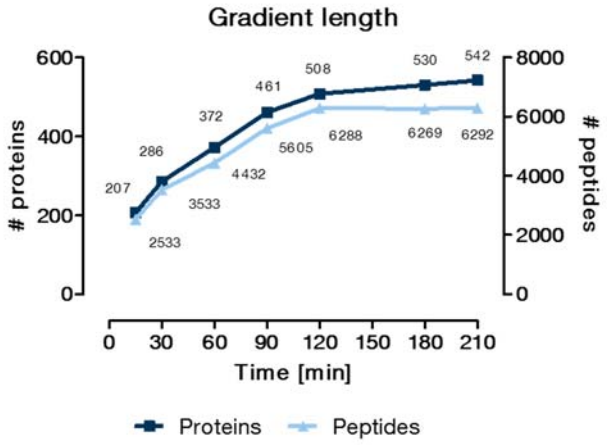


Figure 14 Variation of gradient length The gradient length was varied during the analysis of a 100 ng HeLa sample from 15 min up to 210 min.

Increasing the analysis time from 15, 30, 60 and up to 90 min linearly raised the number of identifications of the system. The slope of the increase gets shallower from 90 to 210 min whereas the gain in IDs between 120 min (508 proteins) and 210 min (542 proteins) remains incremental. The optimal gradient time for this system seems to between 60 and 90 min.

One of the mayor applications in our lab in the field of chemical proteomics is the use of the Kinobead technology for the enrichment of protein kinases out of cell lysate. For a standard DDA based analysis of one Kinobead pulldown, a 90 min gradient is applied on an LTQ Orbitrap Elite. Leading to the identification of 205 protein kinases and an average sequence coverage of 24.3%. For comparative reasons a Kinobead pulldown was analysed using HDMS[e] and a 90 min gradient. In this experiment 133 protein kinases could be identified, which possessed an average sequence coverage of 27.5%.

APPLICABILITY OF TWIMS FOR THE SEPARATION OF POSITIONAL PHOSPHO PEPTIDE ISOMERS

The separation of positional phosphopeptides by using liquid chromatography alone is sometimes not complete[19, 27]. Thus in this passage the use of TWIMS for the additional separation of positional phosphopeptides isomers in the gas phase is investigated. Therefore positional phosphopeptide isomers of a synthetic library with precisely known sites of phosphorylation were used[19, 28]. Since the charge, mass and the shape determines the ion mobility of an ion, differences of positional isomers may result in a different ion structure in the gas phase and may lead to gas phase separation.

In a first step the stability of the drift time of ions throughout the ion mobility separation was assessed (Fig. 15A).

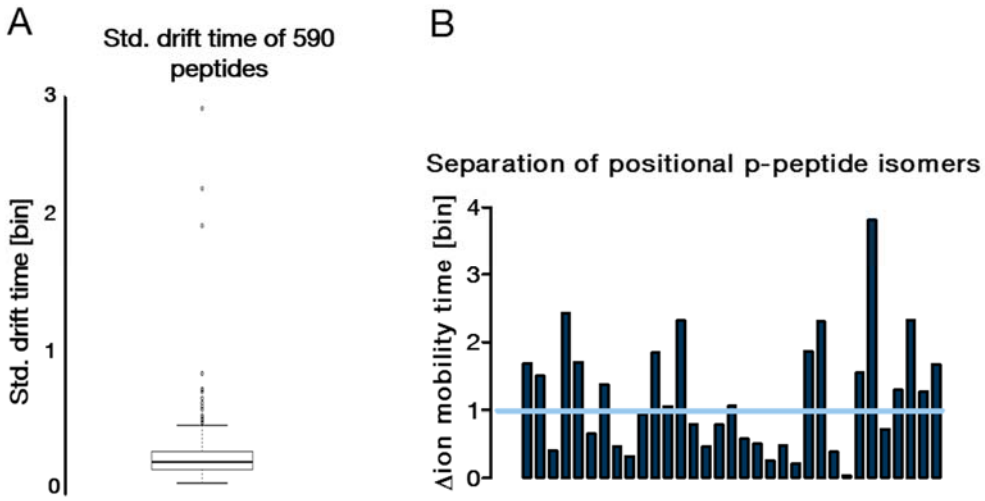


Figure 15 Drift time standard deviation and drift time difference of positional phospho peptide isomers (A) The consistency of the drift time was calculated over 4 E.coli runs. In order to maintain certainty in the peptide identification only peptides were considered that have been identified by more than 10 fragments. The data show a high consistency in the drift time with 99.9% of the data having a standard deviation of 1 bin. (B) The bars represent the drift time difference of the 33 measured pairs of phospho peptide isomers. The bars exceeding the blue line have drift time difference bigger than the standard deviation.

For this purpose, the standard deviation of the drift time of 590 peptides from *E.coli* was calculated. As criteria for trustworthy identifications, only peptides that have been identified in all 4 runs with on average more than 10 product ions per identification were included. In conclusion >99.9% of the peptides show a standard deviation of their drift time of 1 bin. In consequence, two peptides may be distinguished if the drift time differs by more than 1 bin. A bin is the pusher period during an IMS separation and the time can vary depending on the used settings. Here 1 bin = 69 µs.

In total the drift time differences for 33 positional isomers could be determined during this experiment (Fig. 15B). Each peptide has been analysed in single experiments in order to avoid the occurrence of mixed spectra.

In over half the cases (17 pairs) the difference in drift time was >1 bin and would allow a discrimination of these peptides only based on the time spend in the ion mobility device. However, taking also the differences in retention time into account (Fig. 16) the additional separation impact gets reduced.

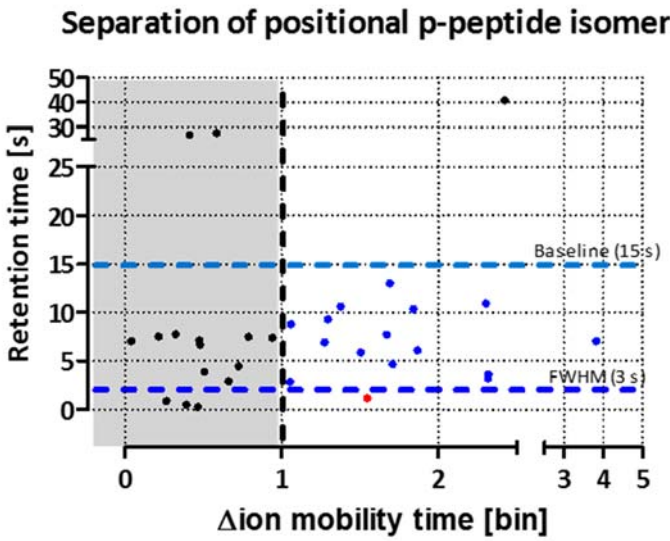


Figure 16 Separation of positional phospho peptides isomers by ion mobility and retention time Plotting the peptide pairs by their difference in drift time and retention time shows that all peptides in the grey area have not drift time difference bigger than 1 bin (black dashed line). The possible criteria on the level of liquid chromatography (baseline or FWHM) are highlighted the blue dashed lines.

If considering two peaks separated at the FWHM (3 s), only one pair out of 33 would benefit from an additional gas phase separation. Regarding a baseline separation of 15 s 15 pairs of peptides would profit from an additional separation using a travelling wave ion mobility separation.

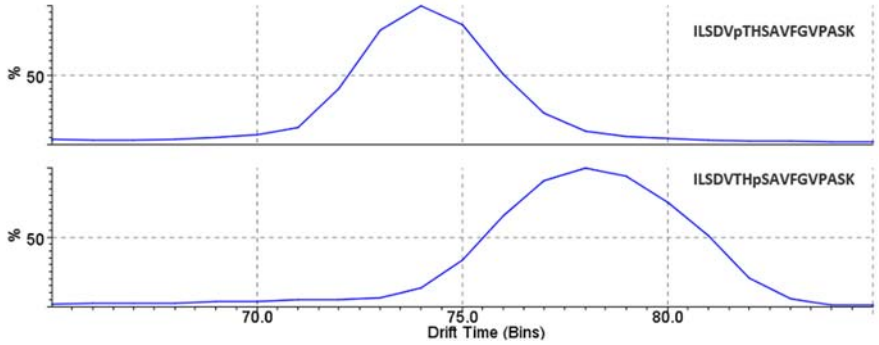


Figure 17 Gas phase chromatograms of two phospho peptide isomers The two peptides ILSDVpTHSAVFGVPASK and ILSDVTHpSAVFGVPASK are showing a difference in drift time of 4 bins. It can also be seen that a baseline separation in this case would be earliest achieved at approx. 20 bins.

Nonetheless, as the resolution of this ion mobility device is 45 a drift time difference of one bin is not enough for a baseline separation in the gas phase. As exemplified by the gas phase chromatograms of the isomeric peptides ILSDVpTHSAVFGVPASK and ILSDVTHpSAVFGVPASK, which are separated by 4 bins (276 µs, Fig. 17).

DISCUSSION

INTEGRATION OF TRAVELLING WAVE ION MOBILITY FOR DIA IN PROTEOMICS STUDIES

The incorporation of ion mobility into mass spectrometers using DIA provides an additional dimension of separation and consequently increases the peak capacity of the analysis. This lead to a decrease in the number of chimeric and interfered spectra compared to a plain MS^e workflow [29-31]. In a DIA workflow additional possibilities for a more selective reconstruction of the filiation of the fragment ions is of relevance (Fig. 4)[32] due to the absence of precursor selection prior to fragmentation. This in consequence explains the approx. 50% higher number of IDs when analysing the same *E. coli* sample with $HDMS^e$ in comparison to MS^e (Fig. 3). As a side note: For the here presented data this difference in performance even holds if applying the later established more stringent criteria for database search (data not shown). Several studies have confirmed the performance enhancing effect of TWIMS in a DIA workflow, already using the predecessor instrument Synapt G2[33, 34]. Interestingly, this enhancement in system capacity seems to have an even larger impact when analysing even more complex samples. Distler et al. reported a 100% gain in IDs by analysing 200 ng of a complex HeLa digest with or without additional gas phase separation[35]. Further they implemented a workflow that applies ion-mobility dependent collision energy (CE) profiles, named $UDMS^e$. By the usage of $UDMS^e$ they have been able to raise the number of identified proteins additionally by 47% percent. This significant improvement is attributed to their observation of under- or overfragmentation of peptides during $HDMS^e$ that by default uses a CE ramp over the full IMS cycle.

The observed signal attenuation (Fig. 5) in $HDMS^e$ and the accompanying disturbed quantification results for high respectively low abundant proteins is a well described issue of TWIMS[33, 34, 36]. Prior to the TWIMS, the ions undergo an accumulating process in the trapping region of the TWIMS device. This results in a saturation of the detector at already lower signal intensity than in MS^e and consequently lead to an underestimation of the amount of high abundant proteins. For the difference in quantitation for the low abundant proteins Shliaha et al.[33] reported the observation of a loss in sensitivity by employing TWIMS. They refer this to the pressure differences on the instrument's ion path, which consist of two high vacuum

regions (Quadrupole and TOF) separated by a region of higher pressure (TWIMS) for optimal ion separation. Consequently, only the most abundant peptides of a low abundant protein would get identified and would lead to an overestimation of the protein abundancy. One proposed solution to approach the discrepancies in quantitation between MS^e and $HDMS^e$, was presented by Bond et.al [37]. The program synapter is combining the information from a "quantification run" (MS^e) and of an "identification run" ($HDMS^e$) to further achieve the sampling depth of the $HDMS^e$ workflow but combine it with the superior quantification from MS^e. Nevertheless, this approach is not applicable in high throughput projects where: I) the sample amount is often limited e.g. patient samples and II) a doubling of measurement time is not justifiable. A further drawback for the $HDMS^e$ workflow in respect to longer running high throughput projects is the constantly changing detector voltage. Meanwhile, the newest update of the Synapt MS instruments G2Si capable of a software operated total ion current (TIC) control, which is designed to avoid detector saturation. In order to gain maximum performance out of the TOF detector, it has to be checked for maximum signal-to-noise ratio in a regular manner. This alone would not be an issue, but in correspondence with the changing detector voltage the initial processing parameter for the raw DIA data is changing and affects the obtained results for protein and peptide ID significantly.

EVALUATION OF DIA DATA PROCESSING USING PROTEINLYNX GLOBAL SERVER

ProteinLynx Global Server (PLGS) is proprietary processing software for $HDMS^e$ data prior to database search. The obtained list of EMRTs from PLGS are undergoing a tentative database search based on three steps (passes)[11]. For the database search the user is able to specify some criteria for protein and peptide ID. This include the minimum number of fragments that are required for a successful peptide ID, the minimum number of fragments for a successful protein ID and the minimum number peptides required for a protein ID. The default criteria, recommended by Waters, for this database search have (1-3-1) revealed promising numbers but at a closer look are questionable e.g. peptide IDs (Fig. 7). Consequently a deeper look in the consistency of the obtain search results was taken. By taking the standard deviation of the drift time as a criterion for consistent peptide ID two things can be seen: I) the general distribution is quite narrow and the deviation is around 1 bin (Fig. 8) II) at a certain point (less than 10 fragment ions for peptide ID), the standard deviations for several peptides start to scatter significantly. Although, in the PLGS output being presented as the same sequence.

In conclusion, the results in this area with <10 fragments are becoming inconsistent. In order to further learn how many ions are needed for reliable peptide identification, the results from a LTQ-Orbitrap Elite (DDA), processed with MaxQuant[25], were analysed (Fig. 9). Here four things raise the attention: I) the frequency distributions between MaxQuant/Elite and PLGS/G2S differ significantly II) there is a change in the different acquisition schemes used on the Elite III) the distributions for the PLGS/G2S results are narrower compared to the result from MaxQuant/Elite IV) pass 1 and Pass 2 from PLGS show differences with Pass 2 even basing most of its IDs on even less fragments than Pass 1. PLGS claims to identify 75% of its peptides with maximum 8 (Pass 1) or 4 (Pass 2) fragments. This is a huge gap compared to the results from the DDA data, which for HCD use up to 17 ions (23 CID) for 75% of the IDs, but without identifying a single peptide with less than 4 (7 CID) ions. The difference between CID and HCD can be explained unambiguously from the higher number of present b-ions in ion trap CID compared to HCD. The result from PLGS may be interpreted in two ways. Either the PLGS search algorithm is more sophisticated compared to e.g. MaxQuant and can base valid IDs on only minimal information, or it creates unreliable peptide and protein identifications. If the first case would be true, the results should not decrease dramatically by applying more stringent criteria for the database search, since the peptides are present and, as learned from the DDA results, deliver an adequate amount of fragment ions that can be matched. Yet ramping the number of necessary fragment ions for a peptide ID leads to a collapse of the numbers for peptide and protein IDs. Demanding 5 fragment ions (approx. starting point of the HCD distribution) leads to a reduction by 40% on the peptide level. Setting the number to 10 fragments (consistency in drift time standard deviation) lets the ID numbers drop by 80% on the peptide and protein level, thus supporting the point that PLGS is optimistically assigning a protein or peptide ID. For more robust protein IDs one might argue that raising the number of required peptides from 1 to e.g. 2 could help. Recalling that PLGS identified 75% (Pass 1) of its peptides with less than 8 fragments this suggestion can be neglected. Issues were also reflected in the analysis of the phospho peptide library[18], which revealed that the PLGS algorithm is not suitable for assigning the sequence of the contained phospho peptides correctly (Fig. 6). The received correctly assigned phospho peptides were obtained by chance, since the algorithm included the same peptide sequence multiple times in the results lists with phosphorylation sites assigned to almost every S, T or Y. Interestingly using more stringent criteria here (10-10-1) leads to a complete loss in the identification of any phosphorylated peptide at all, a result which cannot reflect the sample since the library contains 50% phosphorylated peptides[18].

The idea of scoring peptide spectra matches in proteomics is to give a probability based estimation for the quality of a peptide identification and correspondingly might represent the quality of the underlying spectra. The score distributions between results from the Pass 1 and the Pass 2 search show that Pass 2 tends to identify more peptides that are assigned with a lower score, consequence of basing its IDs on spectra of poorer quality consistent with the information in Fig. 10 (fragments). The correlation of the score and the number of matched fragments is quite inconsistent for PLGS, whereas it is showing a linear behaviour for the Andromeda[26] search engine from MaxQuant[25]. This is clear considering that the Andromeda score represents the -10log of the probability to match a certain number of matched fragment ions of a spectrum to the theoretical expected ions for a peptide by chance[26]. Thus clearly considering the amount of matched fragment ions. On the other hand PLGS does not consider the score for the peptide to spectrum match for the correctness of a protein ID. Moreover it considers 13[11] physico-chemical properties, partly based on empiric based models, of the proteins/peptides from liquid chromatography and gas phase fractionation for validation in combination with a protein FDR. Overall it can be questionable, if just applying a FDR on the protein level is valid enough to make IDs confident if peptides are what are identified in the MS[38](and Wilhelm, in preparation).

Interestingly searching a peak list from PLGS with the Mascot search engine also revealed only the identification of a marginal number of peptides and proteins. This led to the conclusions that indeed the database search performed by PLGS is experiencing more reliability when applying stringent criteria and that the peak capacity of the system used for the generation of the EMRTs is not high enough and still results in a huge number of chimeric spectra. Certainly the PLGS data processing part of the algorithm would benefit from more stringent criteria for the generation of EMRT list to reduce the number of chimeric spectra. Unfortunately these settings cannot be accessed by the user. In light of this results the data reconstruction might be challenging if applying only retention time in combination with high resolution mass spectrometers like it is done in AIF[9] or MS^e.

SWATH-MS is another DIA method developed by the group of Ruedi Aebersold[10]. During a SWATH acquisition the data are generated by consecutively acquiring fragment ion spectra during the whole LC chromatography by repeatedly stepping through 32 discrete precursor isolation windows of 25-Da width across the entire MS1 mass range. The series of isolation windows acquired for a given precursor mass range and across the LC is referred to as a “swath”. In contrast to HDMS^e the correlation of precursor and fragment ion is performed in a targeted approach based on a spectral library that originates from a set of DDA data[39].

Offering high confidence on the level of identifications and enabling accurate quantification. This approach has already been applied in several studies in yeast[40] or human samples[41, 42]. The use of spectral libraries also for DDA data in order to overcome the problems intrinsic of the target-decoy database search approach was recently revisited[43]. Applying spectral libraries for the identification of peptides based on the comparison of the reference spectrum and the measured spectrum was introduced in 2007 and is a valid approach for DDA[44, 45] and DIA[46] data. HDMS^e would certainly benefit from the implementation of a spectral library based peptide and protein identification to avoid the aforementioned shortcomings during the database search. A major piece of software in the field of (targeted) proteomics is Skyline[47]. Skyline is capable of supporting different DIA formats from almost all major vendors i.a. HDMS^e data. With the release of v1.3 in the end of 2012 it became possible to import MS^e data in the system but only in the end of 2014 (v2.6) being able to make use of the ion mobility function. Constantly new software tools are emerging that are able to deal with DIA data including *ab initio* data reconstruction and published vendor independent[48] but have not been evaluated during this study owned to the time point of the software release. Nevertheless this shows the emerging interest in the field of proteomics for DIA approaches.

CLASSIFICATION OF THE DIA PERFORMANCE AND COMPARISON TO LTQ-ORBITRAP INSTRUMENTS

In the beginning of mass spectrometry based proteomics Q-TOF instruments were the standard platform. However, these instruments got replaced by linear ion trap instruments coupled to the Orbitrap mass analyser[49, 50], which are until now the gold standard in the field of proteomics due to their fast acquisition speed and high mass accuracy/resolution[3, 51]. In the first experimental set-up for the classification of the DIA performance on the Synapt G2S, we determined the LOD on a single protein digest using a dilution series of BSA (Table 2). First of all there is no improvement or loss in absolute sensitivity visible between HDMS^e and MS^e and both methods reveal a LOD of 100 amol. This is 10x more sensitive than the predecessor, the G2 (LOD 1 fmol). This has been accomplished by technical improvements in the G2S due to the incorporation of a Stepwave ion guide in the front end and a bigger orifice diameter for improved ion entrance into the instrument. In conclusion, however, the G2S in HDMS^e performs on the same sensitivity level as the LTQ-Orbitrap XL from the year 2005[52]. At the time point of the experiments the most recent LTQ-Orbitrap was the Elite[23], which is >100x more sensitive than the Q-TOF G2S. This difference in sensitivity is also

reflected in the HeLa dilution from 1.000 ng down to 0.1 ng (Fig. 13). Here the minimum loading for the G2S was also 100x higher than for the LTQ-Orbitrap. Over the whole dilution series the ID numbers for the Elite always excided the numbers of the G2S. This discrepancy in performance was also visible in the measured Kinobead pulldown sample (Elite 205 kinases vs. G2S 133 kinases). A supporting observation for the lack of sensitivity of the G2S in combination with the HDMSe is the fact that the difference in protein IDs is decreasing with higher sample loadings. In addition this number of approx. 1100 protein IDs (60 min, 1.000 ng) is never reached for 100 ng HeLa at even extended measurement times (210 min, 550 IDs, Fig. 14), showing that the issue is not the sampling time but the sensitivity of the system.

On the first sight the here presented numbers for a HeLa digest analysed with HDMS^e are in contrast to a recently published paper from Distler et al.[35]. The publication reports the identification of ca. 2,600 protein groups from 200 ng HeLa digest using a 90 min gradient and HDMS^e. This disparity is due to the fact that we applied stronger criteria (10-10-1) for the database search in PLGS compared to Distler et al. (3-3-2) otherwise our data reach the same magnitude.

Summarising, the experimental data show a performance difference between the DIA approach on the G2S and the DDA on the Elite. A straightforward explanation for this difference can be the simple fact that an LTQ is able to accumulate ions until a specific value that is sufficient to produce a proper MS/MS spectrum. Theoretically the Q-TOF instrument can be programmed to a prolonged MS/MS acquisition time to gain sensitivity, but only when operating it in DDA mode. In case of HDMS^e the cycle time for one MS (low energy) and one MS/MS (high energy) scan is fixed and cannot adapt to peaks with low intensity.

APPLICABILITY OF TWIMS FOR THE SEPARATION OF POSITIONAL PHOSPHO PEPTIDE ISOMERS

A primary challenge for proteomics is the comprehensive characterizing and the function of post-translational modifications (PTMs) in global studies[53-55]. Due to its immense impact in the regulation and signalling of almost all cellular functions the field of phospho proteomics[56] has emerged during the last years and gained huge impact. One challenge in the field of phospho proteomics is the assignment of the correct phosphorylation site on a peptide. Several attempts to improve the site localization have already been made using bioinformatics approaches[18, 19, 57, 58] or different fragmentation techniques like ETD or HCD[59,

[60]. Chromatography separation of phospho- and non-phospho-peptides by liquid chromatography has shown to be possible[61]. Nevertheless the chromatographic separation of positional phospho peptide isomers sometimes becomes less likely[19, 27]. Using a gas phase separation approach Shvartsburg et al.[62, 63] have already been shown that conceptual the use of field asymmetric waveform IMS for the separation of positional phospho peptide isomers is possible. In this passage, the use of travelling wave ion mobility separation for the additional segregation of positional phosphopeptide isomers was investigated. Therefore positional phosphopeptide isomers of a synthetic library with precisely known sites of phosphorylation sites were used[19, 28]. Since the charge, mass and the shape determines ion mobility, positional isomers can be resolved in the gas phase. This resolution may arise from differences in the 3D structure of a peptide deriving from different phosphorylation sites. The stability of the drift time in the TWIMS device has proven to be remarkably stable (1 bin, 69 µs) across 4 runs. In consequence, two peptides can be distinguished if the drift time differs by more than 1 bin.

Over 50% of the measured pairs showed a difference in drift time bigger than the standard deviation. In the field of chromatography, two peaks are considered as to be separated if they can be separated at FWHM. If this rule is applied to these data (FWHM 3 s), only one pair out of 33 would benefit from TWIMS (Fig. 16). Demanding the LC peaks to be separated at the baseline, 15 additional pairs of peptides would profit from using a travelling wave ion mobility separation. Gas phase separation of isomeric peptides is primarily thought to avoid chimeric MS2 spectra from co-eluting peptides[19], hence its predominantly impact would be on all 16 pairs of peptides that cannot be baseline resolved by the LC.

Nonetheless, as the resolution of this ion mobility device is 45[64] a drift time difference of one bin is not enough for a baseline separation in gas phase. Consequently the TWIMS device in the G2S would greatly benefit from improved resolution. On the ASMS 2014 Waters presented a prototype of a looping TWIMS device capable of resolving with up to 140[65]. In principle it is possible to produce IMS devices with up to 500 resolution[66].

A possible avenue for future development of this technology would be to combine peptide spectral libraries with drift time libraries to improve the detection of peptides in DIA approaches.

CONCLUSION

In this chapter the applicability of a DIA acquisition method on a Waters Synapt G2S Q-TOF for shot gun and chemical proteomics was investigated. The results demonstrated that the system offers several opportunities for improvement: The inherent settings of the software for the realignment of precursor and the fragments needs more stringent filter criteria to further reduce the number of chimeric spectra. In addition, increasing the peak capacity of the system by e.g. a higher resolution ion mobility device could lead to an improved spectrum reconstruction. This in consequence would also have a positive influence on the identification of PTMs (e.g. phosphorylation), which is so far not possible. Further, the amount of information the PLGS software bases its identifications on may be considered as to low in comparison to state of the art DDA systems. The results also showed that the hardware is performing on the level of the LTQ-Orbitrap XL from the year 2005 in terms of sensitivity. The difference in performance to the in 2012 newest LTQ-Orbitrap Elite becomes less obvious at higher sample amounts. This may be additionally caused by the conceptually higher acquisition speed of a DIA system compared to DDA. In the middle of 2013[3] a newer instrument, the Q-Orbitrap-LTQ Fusion was launched, which is operating at >2x higher acquisition speed (25 Hz) and exceeding the number of IDs of the Elite by 25%. Thus for DDA instruments operating at higher acquisition speeds and instruments in general reaching higher sensitivity, a higher proteome coverage in shorter time is possible[3-6]. So the limitation of DDA by prolonged instrument cycle times[7] becomes insignificant. One would expect that with improved system performance, the stochastic nature and irreproducibility[7] of the precursor selection would disappear. In reality this would only be true if all possible features in a sample would be accessible by the instrument. Currently only “the tip of the iceberg” is worked on and the problem of undersampling remains[8]. Ways to circumvent this issues are implemented by software tools in order to make the results between technical or biological replicates more consistent[25], available even for DIA data[35, 67].

Conceptual MS^e and HDMS^e are still very interesting approaches but lacking state of the art hardware and software support, which makes this workflow not applicable for high performance, high throughput proteomics.

ACKNOWLEDGMENTS

I want to acknowledge Dr. Christopher Hughes and Dr. Johannes Vissers from Waters Corporation for extensive MS and software training. I am also thankful to Andrea Hubauer, Michaela Kroetz-Fahning and Dr. Fiona Pachl for preparing the HeLa digest as well as Mathias Wilhelm for helping with the data analysis and Dr. Charlotte Daly for fruitful discussions.

ABBREVIATIONS

AIF	All ion fragmentation
BSA	Bovine serum albumin
BPI	Base peak intensity
CID	Collision induced dissociation
DB	Database
DDA	Data dependent acquisition
DIA	Data independent acquisition
DT	Drift time
E.coli	Escherichia coli
EMRT	Exact mass and retention time
FAIMS	Field asymmetric waveform ion mobility separation
FDR	False discovery rate
HCD	High energy C-trap dissociation
HDMS^e	Data independent acquisition using ion mobility separation (Waters, Corporation)
HeLa	Human cervical cancer cell line – Henrietta Lacks
ID	Identification
IMS	Ion mobility separation
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography coupled with tandem mass spectrometry
LOD	Limit of detection
LTQ	Linear trap quadrupole
MS	Mass spectrometer
MS^e	Data independent acquisition (Waters, Corporation)
PTM	Post translational modification
PLGS	ProteinLynx Global Server
Q-TOF	Quadrupole Time-of-flight mass spectrometer
RT	Retention time
SWATH	Data independent acquisition method (ABISciex)
TIC	Total ion current
TWIMS	Travelling wave ion mobility separation

REFERENCES

1. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
2. Mallick, P. & Kuster, B. Proteomics: a pragmatic perspective. *Nat Biotechnol* **28**, 695-709 (2010).
3. Hebert, A.S. et al. The One Hour Yeast Proteome. *Mol Cell Proteomics* (2013).
4. Hahne, H. et al. DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Methods* **10**, 989-991 (2013).
5. Pirmoradian, M. et al. Rapid and Deep Human Proteome Analysis by Single-dimension Shotgun Proteomics. *Mol Cell Proteomics* **12**, 3330-3338 (2013).
6. Helm, D. et al. Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol Cell Proteomics* **13**, 3709-3715 (2014).
7. Geromanos, S.J. et al. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683-1695 (2009).
8. 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. *Journal of Proteome Research* **10**, 1785-1793 (2011).
9. Geiger, T., Cox, J. & Mann, M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol Cell Proteomics* **9**, 2252-2261 (2010).
10. Gillet, L.C. 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* **11**, O111 016717 (2012).
11. Li, G.Z. et al. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696-1719 (2009).
12. Myung, S. et al. Development of high-sensitivity ion trap ion mobility spectrometry time-of-flight techniques: a high-throughput nano-LC-IMS-TOF separation of peptides arising from a Drosophila protein extract. *Anal Chem* **75**, 5137-5145 (2003).
13. Valentine, S.J. et al. Toward plasma proteome profiling with ion mobility-mass spectrometry. *J Proteome Res* **5**, 2977-2984 (2006).
14. Hoaglund-Hyzer, C.S., Li, J. & Clemmer, D.E. Mobility labeling for parallel CID of ion mixtures. *Anal Chem* **72**, 2737-2740 (2000).
15. Taraszka, J.A. et al. Mapping the proteome of Drosophila melanogaster: analysis of embryos and adult heads by LC-IMS-MS methods. *J Proteome Res* **4**, 1223-1237 (2005).
16. Wisniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).
17. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2**, 1896-1906 (2007).
18. Marx, H. et al. A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat Biotechnol* **31**, 557-564 (2013).
19. Savitski, M.M. et al. Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics* **10**, M110 003830 (2011).
20. Bantscheff, M. et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* **25**, 1035-1044 (2007).
21. Medard, G. et al. Optimized Chemical Proteomics Assay for Kinase Inhibitor Profiling. *J Proteome Res* (2015).

22. Pringle, S.D. et al. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry* **261**, 1-12 (2007).
23. Pachl, F., Ruprecht, B., Lemeer, S. & Kuster, B. Characterization of a high field Orbitrap mass spectrometer for proteome analysis. *Proteomics* **13**, 2552-2562 (2013).
24. Michalski, A. 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* **11**, O111 013698 (2012).
25. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
26. Cox, J. et al. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research* **10**, 1794-1805 (2011).
27. Hahne, H. et al. in 57th ASMS conference on Mass Spectrometry and allied Topics (Philadelphia, USA; 2009).
28. Lemeer, S. et al. Phosphorylation site localization in peptides by MALDI MS/MS and the Mascot Delta Score. *Anal Bioanal Chem* **402**, 249-260 (2012).
29. Lee, S. et al. Extracted Fragment Ion Mobility Distributions: A New Method for Complex Mixture Analysis. *Int J Mass Spectrom* **309**, 154-160 (2012).
30. Valentine, S.J. et al. Using ion mobility data to improve peptide identification: intrinsic amino acid size parameters. *J Proteome Res* **10**, 2318-2329 (2011).
31. Baker, E.S. et al. An LC-IMS-MS platform providing increased dynamic range for high-throughput proteomic studies. *J Proteome Res* **9**, 997-1006 (2010).
32. Geromanos, S.J., Hughes, C., Ciavarini, S., Vissers, J.P. & Langridge, J.I. Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples. *Anal Bioanal Chem* **404**, 1127-1139 (2012).
33. Shliaha, P.V., Bond, N.J., Gatto, L. & Lilley, K.S. Effects of traveling wave ion mobility separation on data independent acquisition in proteomics studies. *J Proteome Res* **12**, 2323-2339 (2013).
34. Daly, C.E., Ng, L.L., Hakimi, A., Willingale, R. & Jones, D.J. Qualitative and quantitative characterization of plasma proteins when incorporating traveling wave ion mobility into a liquid chromatography-mass spectrometry workflow for biomarker discovery: use of product ion quantitation as an alternative data analysis tool for label free quantitation. *Anal Chem* **86**, 1972-1979 (2014).
35. Distler, U. et al. Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat Methods* **11**, 167-170 (2014).
36. Helm, S., Dobritzsch, D., Rodiger, A., Agne, B. & Baginsky, S. Protein identification and quantification by data-independent acquisition and multi-parallel collision-induced dissociation mass spectrometry (MS(E)) in the chloroplast stroma proteome. *J Proteomics* **98**, 79-89 (2014).
37. Bond, N.J., Shliaha, P.V., Lilley, K.S. & Gatto, L. Improving qualitative and quantitative performance for MS(E)-based label-free proteomics. *J Proteome Res* **12**, 2340-2353 (2013).
38. Marx, H., Lemeer, S., Klaeger, S., Rattei, T. & Kuster, B. MScDB: a mass spectrometry-centric protein sequence database for proteomics. *J Proteome Res* **12**, 2386-2398 (2013).
39. Schubert, O.T. et al. Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat Protoc* **10**, 426-441 (2015).
40. Selevsek, N. et al. Reproducible and Consistent Quantification of the Saccharomyces cerevisiae Proteome by SWATH-mass spectrometry. *Mol Cell Proteomics* **14**, 739-749 (2015).
41. Collins, B.C. et al. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat Methods* **10**, 1246-1253 (2013).
42. Liu, Y. et al. Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *Proteomics* **13**, 1247-1256 (2013).

43. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587 (2014).
44. Lam, H. Building and searching tandem mass spectral libraries for peptide identification. *Mol Cell Proteomics* **10**, R111 008565 (2011).
45. Lam, H. et al. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **7**, 655-667 (2007).
46. Weisbrod, C.R., Eng, J.K., Hoopmann, M.R., Baker, T. & Bruce, J.E. Accurate peptide fragment mass analysis: multiplexed peptide identification and quantification. *J Proteome Res* **11**, 1621-1632 (2012).
47. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).
48. Tsou, C.C. et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* **12**, 258-264 (2015).
49. Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **72**, 1156-1162 (2000).
50. Makarov, A. et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* **78**, 2113-2120 (2006).
51. Scheltema, R.A. 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* **13**, 3698-3708 (2014).
52. Olsen, J.V. et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **4**, 2010-2021 (2005).
53. Mann, M. & Jensen, O.N. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255-261 (2003).
54. Molina, H., Horn, D.M., Tang, N., Mathivanan, S. & Pandey, A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* **104**, 2199-2204 (2007).
55. Gupta, N. et al. Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome research* **17**, 1362-1377 (2007).
56. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355-365 (2001).
57. Beausoleil, S.A., Villen, J., Gerber, S.A., Rush, J. & Gygi, S.P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **24**, 1285-1292 (2006).
58. Bailey, C.M. et al. SLoMo: automated site localization of modifications from ETD/ECD mass spectra. *J Proteome Res* **8**, 1965-1971 (2009).
59. Frese, C.K. et al. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res* **10**, 2377-2388 (2011).
60. Olsen, J.V. et al. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **4**, 709-712 (2007).
61. Steen, H., Jebanathirajah, J.A., Rush, J., Morrice, N. & Kirschner, M.W. Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. *Mol Cell Proteomics* **5**, 172-181 (2006).
62. Shvartsburg, A.A., Creese, A.J., Smith, R.D. & Cooper, H.J. Separation of peptide isomers with variant modified sites by high-resolution differential ion mobility spectrometry. *Anal Chem* **82**, 8327-8334 (2010).
63. Shvartsburg, A.A., Singer, D., Smith, R.D. & Hoffmann, R. Ion mobility separation of isomeric phosphopeptides from a protein with variant modification of adjacent residues. *Anal Chem* **83**, 5078-5085 (2011).
64. Giles, K., Williams, J.P. & Campuzano, I. Enhancements in travelling wave ion mobility resolution. *Rapid Commun Mass Spectrom* **25**, 1559-1566 (2011).

65. Giles, K. et al. in 62st ASMS conference on Mass Spectrometry and allied Topics (Baltimore; 2014).
66. Shvartsburg, A.A. et al. High-definition differential ion mobility spectrometry with resolving power up to 500. *J Am Soc Mass Spectrom* **24**, 109-114 (2013).
67. Kuharev, J., Navarro, P., Distler, U., Jahn, O. & Tenzer, S. In-depth evaluation of software tools for data-independent acquisition based label-free quantification. *Proteomics* (2014).

CHAPTER III

PERFECT TIMING

ION MOBILITY MS FOR ENHANCED PERFORMANCE

ABSTRACT

One of the limiting factors in determining the sensitivity of tandem mass spectrometry using hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) instruments is the duty cycle of the orthogonal ion injection system. As a consequence, only a fraction of the generated fragment ion beam is collected by the TOF analyzer. The here described method utilises post-fragmentation ion mobility spectrometry of peptide fragment ions in conjunction with mobility time synchronized orthogonal ion injection leading to a substantially improved duty cycle and a concomitant improvement in sensitivity of up to 10-fold for targeted and bottom-up proteomic experiments. High resolution targeted proteomics using HD-MRM on the Q-TOF demonstrated similar quantification results compared to LFQ based quantification on a LTQ-Orbitrap. Further HD-DDA enabled the identification of 7,500 human proteins within one day and 8,600 phosphorylation sites within 5h of LC-MS/MS time in bottom-up proteomic experiments. The method also proved powerful for multiplexed quantification experiments using tandem mass tags exemplified by the chemoproteomic interaction analysis of histone deacetylases with Trichostatin A.

INTRODUCTION

Mass spectrometry-based proteomics enables the large scale interrogation of proteins to study time and cell-type dependent protein expression, posttranslational modifications, protein degradation, protein interactions, as well as protein activities[1]. The *de-facto* standard analytical platform for the identification and quantification of peptides and proteins is the coupling of reversed phase liquid chromatography to nano-electrospray ionization tandem mass spectrometry (LC-MS/MS)[2]. The field is still driven by advances in chromatographic and mass spectrometric technology, which enables the analysis of ever more complex proteomes at ever-decreasing time and sample quantity requirements[3-6]. In particular, hybrid tandem mass spectrometers such as ion trap-orbitrap, quadrupole-orbitrap and quadrupole-time-of-flight (Q-TOF) instruments have dominated the field for many years. These configurations offer high data acquisition speed and high mass accuracy for peptide precursor and fragment ions both of which are advantageous when analyzing complex proteomes[7].

Although it is technically possible to acquire tandem mass spectra at a rate of up to 100 Hz[8] on Q-TOF instruments, only relatively few of these contain a sufficient number of fragment ions to yield a productive peptide identification. One reason for this is the low duty cycle of orthogonal acceleration (oa) TOF instruments which is a consequence of sampling a continuous ion beam into an orthogonally mounted pulsed mass analyzer. The duty cycle of such instruments is given by:

$$Duty\ cycle(mass) = \frac{Wb}{Sep}\sqrt{\frac{mass}{max.mass}}$$

In which *Wb* is the width of the ion beam pulsed orthogonally into the TOF, *Sep* is the distance from the pusher to the TOF detector, *mass* is the mass of an ion of interest and *max.mass* is the maximum mass of the detection system. The geometrical figure Wb/ Sep is typically 0.25 for commercial oaTOF mass spectrometers. The sampling of the fragment ion packet into the oaTOF is normally selected such that consecutive spectra do not overlap, i.e. that the heaviest ions have time to reach the detection system, leading to the second (mass dependent) term in the equation[9]. As a consequence, the practical duty cycle of Q-TOF instruments is often no higher than about 10% leading to substantial losses in sensitivity. Past approaches addressing this limitation have not lead to entirely satisfactory results for the analysis of complex mixtures by LC-MS/MS[10-12]. For example,

Chernushevich and colleagues were able to improve the duty cycle to up to 100% by trapping ions in the collision cell and using short ejection bursts synchronized with the oaTOF extraction[13]. However, this method had the limitation that the maximum improvement could only be achieved over a small mass range. Similarly, Giles and co-workers demonstrated that including ion mobility separation of ions can lead to 100% duty cycle, again, over a small mass range[12].

However, traveling wave ion mobility spectrometry (TWIMS)[14, 15], on peptide fragment ions prior to their measurement in the oaTOF analyser greatly improves the duty cycle over a wide mass range and, therefore, sensitivity of tandem mass spectrometry of peptides. This consequently allows for faster data acquisition rates and higher sensitivity in targeted proteomics and leading to improved peptide identification and quantification rates in bottom up proteomics.

EXPERIMENTAL PROCEDURES

SAMPLE PREPARATION

HELA DIGEST

Mycoplasma-free HeLa S3 cervix carcinoma cells (DMSZ, Braunschweig, Germany) were cultured in Dulbecco's Modified Eagle Medium (DMEM) with high glucose (PAA, Pasching, Austria) supplemented with 10% (v/v) fetal bovine serum (FBS, PAA, Pasching, Austria) at 37°C in humidified air and 10% CO_2. Cells were washed with phosphate buffered saline (PBS) and harvested by lysis using 50 mM Tris/HCl pH 7.5, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 0.8% NP-40, 1 mM dithiothreitol and 25 mM NaF with freshly added protease inhibitors and phosphatase inhibitors (5x phosphatase inhibitor cocktail 1, Sigma-Aldrich, Munich, Germany, 5x phosphatase inhibitor cocktail 2, Sigma-Aldrich, Munich, Germany, 1 mM sodium ortho-vanadate and 20 nM Calyculin A, LC Laboratories, Woburn, MA, USA). Protein extracts were clarified by ultracentrifugation for 1 h at 145,000 xg at 4°C and protein concentration was determined by the Bradford method. Lysates were denatured in 8 M urea, 0.1 M Tris/HCl, subsequently diluted to 2 M urea followed by protein digestion with trypsin (Promega Corp., Madison, WI, USA) according to the FASP protocol[16]. After overnight digestion, peptides were eluted from the filters with 50 mM TEAB and purified on C_{18} StageTips as described[17]. Human epidermoid A431 cells used for IMAC phosphopeptide enrichment experiments were grown in Iscove's Modified Dulbecco's Medium (IMDM) medium supplemented with 10% (v/v) FBS and 1% antibiotic/antimycotic solution. Prior to lysis, cells were treated with 1 mM sodium orthovanadate for 10 min. After harvesting, cells were washed twice with ice cold PBS and lysed in 8 M Urea, 50 mM TEAB, 1 x EDTA free protease inhibitor mixture (complete mini, Roche), 1 mM sodium orthovanadate and Phosphatase inhibitor cocktail (Sigma). The lysate was centrifuged at 14.500 rpm for 1 hour at 4°C. Protein concentration was determined using the Bradford method. Proteins in the supernatant were reduced with 10 mM DTT at 56°C for 1 h and alkylated with 55 mM chloro-acetamide for 45 min at room temperature in the dark. The protein mixture was diluted 5-fold with 50 mM TEAB to reduce the urea concentration to 1.6 M. A first digestion was performed using sequencing grade trypsin (Promega; 1:100 enzyme:substrate ratio) and incubation at 37°C for 4 h. Subsequently, a second aliquot of 1:100 trypsin was added. After digestion

at 37°C overnight, samples were acidified with formic acid (FA). SepPack columns (C18 cartridges Sep-Pak Vac 1cc [50 mg], Waters Corp.) were used for peptide desalting and concentration, and eluates were dried in a vacuum centrifuge and stored at -80°C. IMAC enrichment of phosphopeptides from human cell lines was performed as described[18]. For the tissue proteomes, 50 µg of lysate were reduced, alkylated (see above) and separated via an LDS-PAGE gel and cut into 12 equally sized bands. In-gel trypsin digestion was performed according to standard procedures[19]. The chemoproteomic profiling of Trichostatin A was performed according to Bantscheff et al. (2011)[20] and kindly provided by Isabell Becher from Cellzome.

KINOBEAD PULLDOWN – LAPATINIB

The used Kinobead sample was generated as described in Bantscheff et al (2007)[21] and Medard et al. (2015)[22]. In short, lysates of a high EGFR expressing BT474 cells were diluted with equal volumes of 1x compound pulldown (CP) buffer (50 mM Tris/HCl pH 7.5, 5% glycerol, 1.5 mM MgCl 2, 150 mM NaCl, 25 mM NaF, 1 mM dithiothreitol and freshly added protease inhibitors and phosphatase inhibitors (5x phosphatase inhibitor cocktail1, (Sigma-Aldrich, Munich, Germany), 5x phosphatase inhibitor cocktail 2, (Sigma-Aldrich, Munich, Germany), 1 mM sodium ortho-vanadate and 20 nM Calyculin A, (LC Laboratories, Woburn, MA, USA)). Lysates were diluted to a final protein concentration of 5 mg/ml using 1x CP buffer supplemented with 0.4% NP-40. Kinobeads (100 µL settled beads) were incubated with lysates (total of 5 mg of protein) for 1 h at 4°C. Subsequently, beads were washed with 1x CP buffer and collected by centrifugation. Bound proteins were eluted with 2x NuPAGE LDS Sample Buffer (Invitrogen, Darmstadt, Germany) and eluates were reduced and alkylated by 50 mM dithiothreitol and 55 mM iodoacetamide. Samples were then run into a 4–12% NuPAGE gel (Invitrogen, Darmstadt, Germany) for about 0.5 cm to concentrate the sample prior to in-gel tryptic digestion. In-gel trypsin digestion was performed according to standard procedures. The samples were kindly provided by Benjamin Ruprecht.

PREPARATION OF THE BSA DIGEST

A commercially available protein digestion standard of bovine serum albumin (BSA) (Michrom Bioresources, CA, USA) was diluted to the desired concentration using LC grade water with 0.1% formic acid.

LC-MS/MS MEASUREMENTS

Liquid chromatography tandem mass spectrometry was performed by coupling a nanoAcquity UPLC (Waters Corp., Manchester, UK) to a Q-TOF SYNAPT G2Si instrument (Waters Corp., Manchester, UK). Unless otherwise described, one microgram of the HeLa digest was delivered to a trap column (180µm x 20mm nanoAcquity UPLC 2G-V/MTrap 5µm Symmetry C18, Waters Corp, Manchester, UK) at a flow rate of 5 µL/min in 99.9% solvent A (0.1% FA, in HPLC grade water). After 3 min of loading and washing, peptides were transferred to an analytical column (75µm x 250mm nanoAcquity UPLC 1.8µm HSST3, Waters Corp, Manchester, UK) and separated at a flow rate of 300 nL/min using a gradient from 1% to 40% solvent B (0.1% FA in acetonitrile) of varying length (15-360 min). The gel cuts for the tissue proteomes were analyzed using a 90 min gradient with a total turnaround time of 2 hrs. The eluent was sprayed via PicoTip Emitters (Waters Corp, Manchester, UK) at a spray voltage of 2.6-3.0 kV and a sampling cone voltage of 25 V and a source offset of 30 V. The source temperature was set to 100°C. The cone gas flow was 20 l/h with a nano flow gas pressure of 0.4 bar and the purge gas was turned off. When the SYNAPT G2Si instrument was operated in HD-MRM mode, a list of targeted peptides (for BSA or EGFR) was assessed based on data derived from a discovery experiment (HD-DDA). The LC time window for transition scanning was set to ±2 min of the apex of the elution profile and the scan time was set to 1 s. The CID collision was performed in the trap region of the Triwave device and the CE was adjusted to the precursor *m/z*-value. When the instrument was operating in data-dependent mode, it was automatically switching between MS and MS2. Full scan MS and MS2 spectra (*m/z* 50 – 5000) were acquired in resolution mode (20,000 resolution FWHM at *m/z* 400). MS spectra were acquired using a fixed acquisition time of 150 ms. Tandem mass spectra of up to 12 precursors were generated in the trapping region of the ion mobility cell by using a collisional energy ramp from 6/9 V (low mass, start/end) to up to 147/183 (high mass, start/end). For the TMT samples, the collision energy ramp was adjusted by +2 V resulting in more efficient fragmentation. The precursor ion isolation width of the quadrupole was set to 2.0 Th (low mass resolution 10, high mass resolution 15). The MS2 scan time was set to 80 ms and the “TIC stop” parameter was set to 500,000 intensity/s allowing a maximum accumulation time of 250 ms (i.e. up to three tandem MS spectra of the same precursor). Dynamic exclusion of fragmented precursor ions was enabled and the exclusion time window was set to 30 s. For HD-DDA/-MRM methods using the IMS configuration, a variable IMS wave velocity was used. Wave velocity was ramped from

2500 m/s to 400 m/s (start to end) and the ramp was applied over the full IMS cycle. A manual release time of 500 µs was set for the mobility trapping and a trap height of 15 V with an extract height of 0 V. The IMS wave delay was set to 1000 µs for the mobility separation after trap release. The pusher/ion mobility synchronization for the HD-DDA and the HD-MRM method was performed using MassLynx VRP1214 and DriftScope v2.4. LockSpray of Glufibrinopeptide-B (*m/z* 785.8427) was acquired every 60 s for lock mass correction was applied post acquisition. For DDA experiments, the basic settings were the same as for HD-DDA but using a reduced mass range of 50-2000 m/z and without fragment separation via ion mobility.

PEPTIDE AND PROTEIN IDENTIFICATION AND QUANTIFICATION

Raw MS data were either processed by Mascot Distiller (version 2.4.3.1) for peptide and protein identification and isobaric quantification or by Progenesis QI (Nonlinear Dynamics, United Kingdom) for label free quantification. MS/MS spectra were searched against the Uniprot human database (release date 21.7.13 containing 88,354 sequences and including the PFAM domain identifiers for each entry) using the Mascot search engine (version 2.4.1) with the following search parameters: full tryptic specificity, up to two missed cleavage sites; carbamidomethylation of cysteine residues was set as a fixed modification and N-terminal protein acetylation and methionine oxidation as variable modifications. For phosphopeptide enriched samples also phosphorylation for serine, threonine and tyrosine was selected as variable modification. In order to obtain the quantification result for the chemoproteomic profiling TMT 6-plex was selected as quantification method in Mascot. Mass spectra were re-calibrated within Mascot Distiller and searched with a mass tolerance of 25 ppm. Fragment ion mass tolerance was set to 0.1 Da. Mascot search results were further processed using the Percolator algorithm[23, 24] and subsequently loaded into Scaffold (v4.1.1). The data was filtered using a Mascot Percolator score of 13, corresponding to a posterior error probability (PEP) of 0.05. The resulting protein false discovery rate (FDR) was 0.2% for the data shown in Figure 7 (different gradient times), 0.4% for the tissue proteomes shown in Figure 10 and 3.8 % for the chemoproteomic experiment shown in Figure 12/13 (please note that fewer than 300 proteins were contained in the latter experiment which renders the protein FDR estimate less meaningful. Mascot search results of the phosphopeptide analysis were

further analyzed by the percolator algorithm embedded in the program RockerBox [25]. Subsequently, the data was filtered at a peptide spectrum match (PSM) FDR of 0.01 and an additional Mascot ion score cut off of 14 was applied[26]. Phosphopeptide sequences were filtered for redundancy in order to obtain the number of identified phosphopeptides. The number of unique phosphopeptide sequences was determined by counting only non-redundant peptide sequences, regardless of how many and where phosphorylation events occurred within the peptide. False localization rates (FLR) for confidently assigned phospho-sites (<0.05 FLR) were estimated using the Mascot Delta Score as described previously[26, 27].TMT quantification was performed by Mascot using default settings (minimum of two peptides, ion score above homology threshold, automatic outlier removal and intensity weighted ratio calculation). LFQ data were generated using Progenesis QI for proteomics (v.1.0.5156.29278; Waters, Corporation) Visualization of quantitative TMT data was performed using GraphPad Prism v5.03. For the analysis of the HD-MRM data the targeted proteomics software Skyline (v.2.6.06851)[28] was used.

DATA AVAILABILITY

The mass spectrometric raw files were converted into mz5 [29] format using ProteoWizard [30] (version 3.0.5245 64-bit). The mz5 files are available along with peaklist files (mgf format) and search result files (Scaffold files and Scaffold exported protein and peptide lists) via proteomeXchange [31] (http://www.proteomexchange.org; accession number PXD000863).

RESULTS AND DISCUSSION

DUTY CYCLE ENHANCEMENT

In this study, traveling wave ion mobility spectrometry (TWIMS)[14, 15], available on a commercial mass spectrometer was coupled with the established data dependent acquisition scheme (DDA) typically used in bottom-up proteomics (termed high definition or HD-DDA) and with a multi-reaction monitoring approach (termed HD-MRM). As will become apparent below, this enhances the duty cycle of the instrument and, therefore, sensitivity (Fig. 1) The principle and performance characteristics of TWIMS have been described before[12] . Briefly, in the instance of tryptic peptides, TWIMS separation is dependent upon the ions charge (z) and ions are nested, for a given charge state, by mass and drift time. This charge state separation and nesting can be used to discriminate against single charge background and to focus upon and select exclusively multiply charged peptides (or other components) for tandem MS. Subsequently, peptide precursor ions are sequentially selected by the quadrupole mass analyzer and fragmented by collision induced dissociation (CID) in the first stacked ring ion guide of the Triwave device and prior to reaching the ion mobility cell. Product ions are trapped within this first travelling wave region of the Triwave device and gated into the high-pressure ion mobility cell where they are separated according to their gas phase mobility within the cell which is predominantly determined by mass, charge, size and shape. As a result, fragment ions of the same mobility exit the cell as a series of compact packets. Hence, by synchronizing the pusher pulse that accelerates the fragment ions into the oaTOF mass analyzer with the arrival of product ions from the TWIMS cell into the pusher region, fragment ions are sequentially injected into the TOF analyser with greatly enhanced duty cycle (~100%) across the mass scale[12].

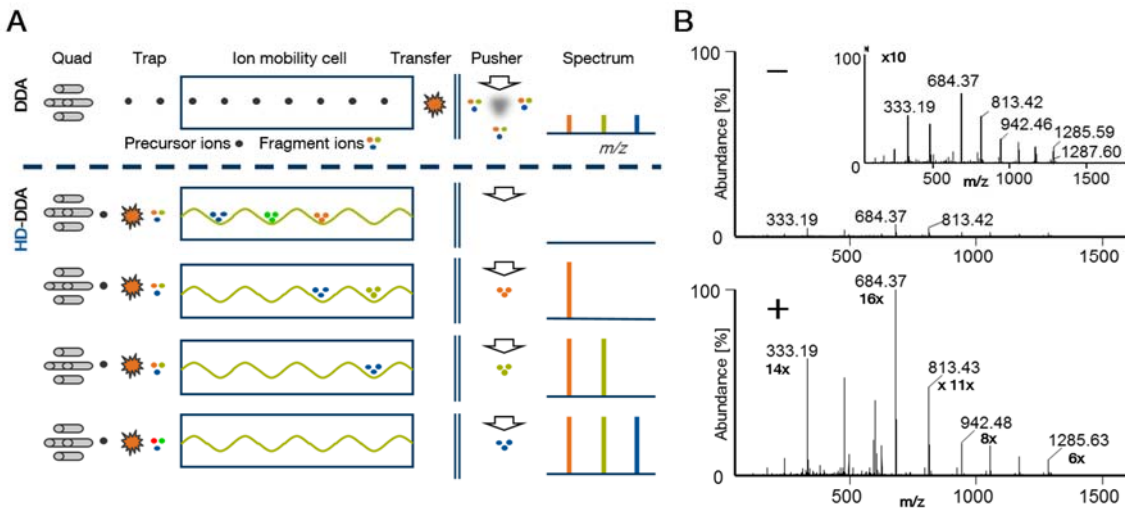


Figure 1 Synchronization of fragment ion mobility and pusher frequency improves the system performance of a Q-TOF instrument. (A) Schematic illustration of operation of conventional DDA (above dashed line) and HD-DDA (below dashed line) on a Q-TOF instrument. (B) Tandem mass spectrum of GluFib without (-) and with (+) fragment ion mobility synchronization enabled. The insert shows the 10x magnified tandem mass spectrum of the conventional DDA experiment.

This ion separation and packeting leads to an increase in sensitivity. Each ion mobility separation occurs over a time scale of ~10 ms and consists of 200 discrete bins. Each mobility time bin contains a part of the full spectrum and, therefore, the complete tandem mass spectrum of a peptide is acquired by integrating the signal across the entire mobility separation. Since the ion mobility time frame is in the order of milliseconds, it nests well between the timeframe of liquid chromatography (seconds) and the oaTOF mass spectrometer which operates in the microsecond timeframe. For a single peptide, enabling the ion mobility enhancement lead to an average of ~10-fold higher fragment ion intensity compared to the respective “standard” experiment without TWIMS based concentration of fragment ions (Fig. 1B).

HD-MRM

The ability to quantify multiple proteins with high precision and reproducibility across a range of biological samples is an important task in modern systems biology and biomedical research. Targeted proteomic methods such as SRM on tandem quadrupole instruments (Fig. 2A) have developed into the *de-facto* standard for this purpose[32, 33]. However, the very high molecular complexity of biological samples is challenging for the selectivity of the classical SRM approach using low resolving power mass analysers if it comes to the point of discrimination of targeted signal an interference[34, 35]. In this experiment, the HD-MRM acquisition method was used on a high resolution, high mass accuracy quadrupole time of flight mass spectrometer that can potentially overcome the selectivity issues presented by tandem quadrupole instruments[36].

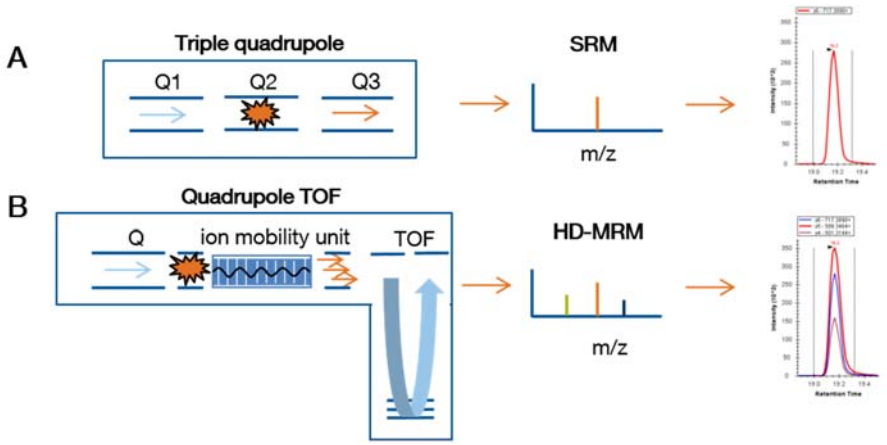


Figure 2 Principles of SRM and HD-MRM. (A) In an SRM experiment, a triple quadrupole instrument is used as a low resolution mass filter for one precursor and one fragment at a time. (B) In contrast, a quadrupole TOF instrument allows for the analysis of one precursor and all its fragments at the same time and at high resolution improving speed and selectivity.

The conceptual advantage of a TOF over a triple quadrupole mass analyser for MRM assays is i) high mass accuracy and resolution which should improve selectivity and ii) the fact that all fragment ions generated from the targeted peptide are recorded in one spectrum. Hence the determination and the extraction of the transitions can be performed post-acquisition without the need to optimize any of these beforehand.

In an initial experiment, the LOD and LOQ of the assay were established using a BSA dilution series and targeting (10 fmol - 1 amol on column, Fig. 3) BSA specific fragment ions over a 30 min LC gradient. BSA specific fragments could be identified until 1 amol (LOD). In terms of quantification, the assay also showed a linear behaviour down to 1 amol (LOQ, R^2=0.995) showing that at least in principle, a Q-TOF based SRM/MRM experiment can be highly sensitive.

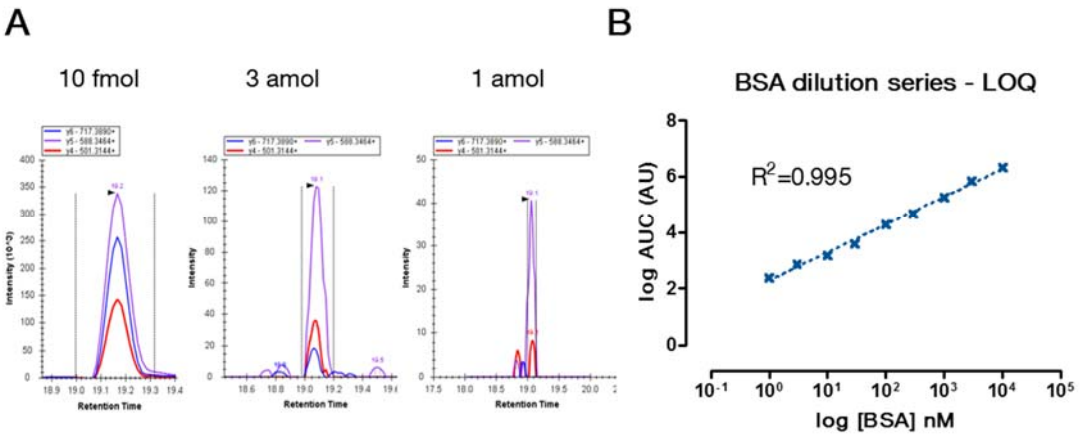


Figure 3 LOD and LOQ (A) Fragments of a BSA specific peptide (m/z = 570,7366, CCTESLVNR) could be detected from 10 fmol down to 1 amol BSA digest on column at a resolution in the MS2 of 25,000. (B) Integration of the areas under the curve of the fragment ions gave a linear behavior from 10 fmol down to 1 amol, coverring 4 orders of magnitude.

The HD-MRM method was next applied for the quantification of the Lapatinib:EGFR interaction (Fig. 4A). Samples from kinobead pulldowns with increasing doses of Lapatinib were subjected to HD-MRM measurements and analysed using a 60 min LC gradient (50% of the standard DDA approach). A set of 15 unique peptides of EGFR was monitored by selecting precursor ions with the quadrupole across the chromatographic time frame, following CID fragmentation with all corresponding fragment ions (HD-MRM) recorded by the oaTOF analyser. Extraction and quantification of all transitions determined a KD value of approx. 50 nM for the Lapatinib:EGFR interaction in the background of >1,000 other proteins in the sample confirming the high selectivity of this HD-MRM approach. The determined KD value is consistent with previous data obtained by classical DDA in conjunction with MS1 quantification as well as with the literature on the cellular activity of Lapatinib[37]. In a post-acquisition evaluation step (Fig. 4B), the slight fluctuations of the individual transitions of a single peptide could be revealed. Further, it could be shown that using only one transition from one peptide is even enough to get a decent quantification result (SRM type experiment).

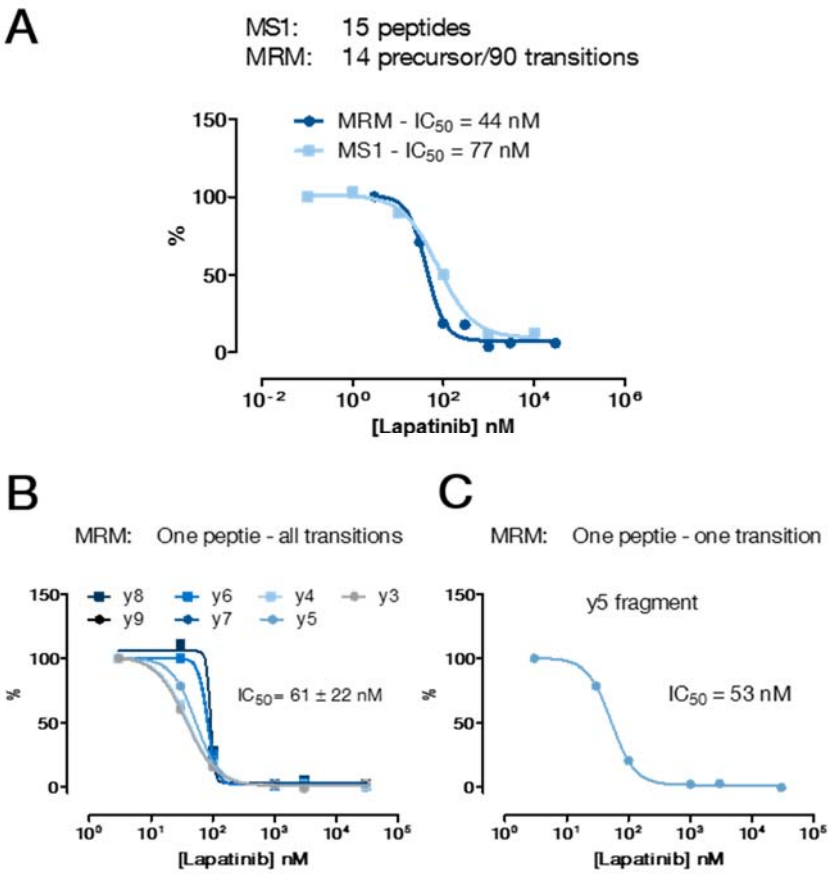


Figure 4 HD- MRM and MS1 based quantification of the Lapatinib : EGFR interaction (A) Standard label-free quantification using MS1 intensity and HD-MRM result in very similar inhibition profiles providing the proof-of-concept of the HD-MRM measurement in this assay. (B) Post-acquisition analysis/validation of the true MRM strategy revealed slightly fluctuations of individual transitions for one peptide. Owing to its high selectivity, a single transition (C) of a single EGFR peptide can suffice in order to generate a reasonable quantification result.

This approach was not, however, applicable for targeting multiple kinases (approx. 250[22]) present in a standard kinobead screen. Further, despite the previously shown results, it could not be applied in a high throughput screen, since this iteration of the acquisition software allowed only the selection of max. 40 peptides in a targeted experiment. Nevertheless, in summary this data suggest that high resolution and high mass accuracy MRM experiments on Q-TOF instruments are suitable for targeted quantification experiments in a complex biological background. A recent study using targeted proteomics on a Q-Exactive supported the here demonstrated idea of using high resolution and high mass accuracy instruments for targeted proteomics to overcome the aforementioned issues of using low resolution instruments[38].

HD-DDA vs. DDA

Next the ion mobility enhancement was used in combination with a standard DDA acquisition scheme and the merits of the approach using a dilution series of shotgun digested HeLa cell lysates (Fig. 5A) were investigated. HD-DDA analysis identified >2,000 proteins (>7,000 peptides) from 1 µg of digest on column using a 60' LC gradient.

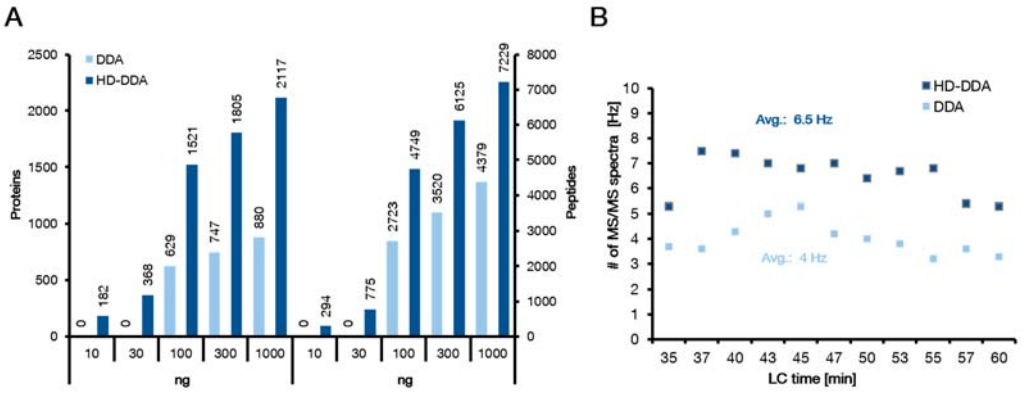


Figure 5 Performance comparison regarding sensitivity of the HD-DDA and DDA methods. (A) Dilution series of a HeLa digest from 1,000 ng to 10 ng on column (each analysed on a 60-min LC-gradient) via DDA (light blue) and HD-DDA (dark blue), illustrating the differences of both methods for peptide and protein identification. (B) Comparison of the rate of tandem mass spectrum acquisition (in Hz) across the LC-gradient for DDA (light blue) and HD-DDA (dark blue).

The method also more than doubled the number of identified proteins at any point across the dilution series and improved the detection limit 10-fold over classical DDA. The increased quality of HD-DDA tandem mass spectra allowed for a 60% higher average acquisition rate of tandem mass spectra compared to DDA (6.5 Hz vs 4 Hz, Fig. 5B).

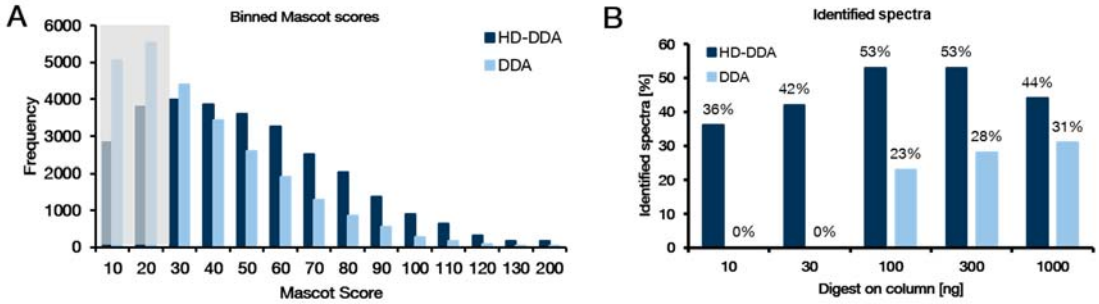


Figure 6 Comparison of DDA and HD-DDA at the level of peptide identification score (here Mascot ion score) and the productivity of tandem MS measured by the fraction of MS/MS spectra leading to a successful identification. (A) Binned Mascot score distribution obtained from the measurement of 1,000 ng of HeLa digest on column on a 60-min LC gradient for DDA (light blue bars) and HD-DDA (dark blue bars). Only "rank 1" peptide spectrum matches (PSMs) were used in this plot. The grey zone indicates PSMs below the Mascot identity threshold. (B) Comparison of the fraction of successfully identified MS2 spectra from DDA (light blue bars) and HD-DDA (dark blue bars) measurements across a dilution series of a HeLa digest from 10 – 1,000 ng digest on column and analyzed by a 60-min LC gradient.

For the same reason, Mascot identification scores increased by ~40% (corresponding to 2.5-fold better identification probability because the Mascot ion score is a log10 score, Fig. 6A), thereby giving more confidence in the assigned peptides compared to DDA data. Another consequence of the improved MS2 data quality was an increase in the productivity of peptide identification from ~25% to ~50% of all acquired CID spectra (Fig. 6B).

SAMPLE THROUGHPUT AND PROTEOME COVERAGE

Sample throughput and depth of proteome coverage is an important criteria in mass spectrometry based bottom-up proteomics. For this purpose different LC gradient times were used for the analysis of 1 µg of HeLa digest. The number of identifications from triplicate analysis of HeLa digests ranged from ~900 proteins (2,600 peptides) using a 15 min LC gradient to ~4,400 proteins (20,000 peptides) using a 6 hour gradient (Fig. 7).

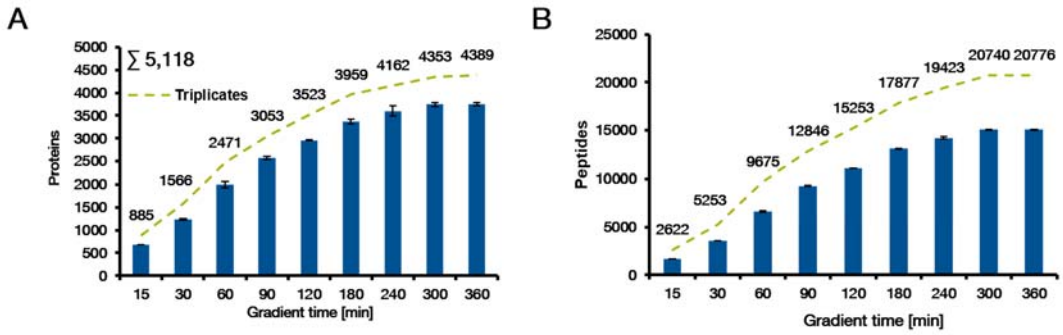


Figure 7 Analysis of triplicate measurements of 1,000 ng HeLa digest using different LC gradient times. (A) Number of identified proteins as a function of LC gradient time. (B) Same as (A) but for identified peptides. The numbers above the bars represent the sum of the triplicate identifications.

The high number of protein identifications at very short separation times makes HD-DDA an interesting option for the analysis of simple proteomes (e.g. protein complexes obtained by affinity purification) or for the targeted analysis of particular proteins within a given sample). The apparent reproducibility of the number of protein and peptide identifications was surprisingly high (illustrated by the very tight error bars in Fig. 7).

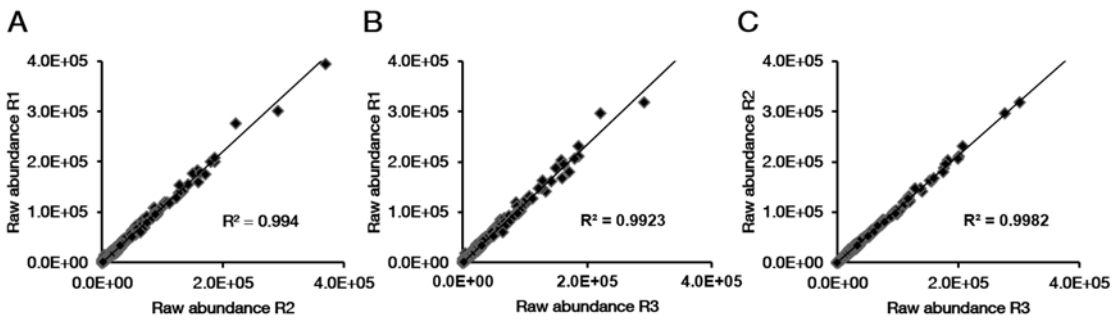


Figure 8 Reproducibility of peptide intensities (expressed as Progenesis raw abundances) between replicate analysis (denoted R1, R2, R3). The data shown represents pair-wise comparisons of three 60 min LC-MS/MS experiments using 1 mg HeLa digest on column each. (A) R1 vs. R2 (B) R1 vs R3 (C) R2 vs R3.

While reproducibility in terms of signal intensity was also very high (Fig. 8), closer inspection revealed that the actual protein and peptide recall rates were much lower (Fig. 9 and Fig.10). For example, for 60 min gradients, only 62% of all proteins and 39% of all peptides were identified in all replicates. As one would expect, recall rates were worse for 15 min gradients (55% of all proteins and 33% of all peptides were identified in all replicates) but better for 360 min gradients where identification recalls increased to 71% for proteins and 47% for peptides (Fig. 10).

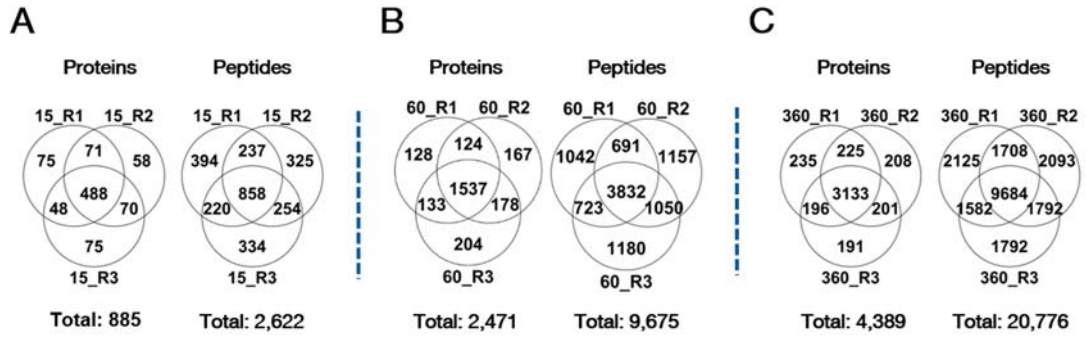


Figure 9 Reproducibility analysis of triplicate measurements of 1.000 ng HeLa digest using different LC gradient times. Analysis of the reproducibility/recall of protein and peptide identification for (A) 15 min, (B) 60 min and (C) 360 min LC gradients.

As noted before[39] and confirmed here again, this data indicates that the number of peptides present in shotgun digests of cell lines still overwhelms the capacity of the mass spectrometer highlighting the need for yet further improvements in the speed of data acquisition (see also below).

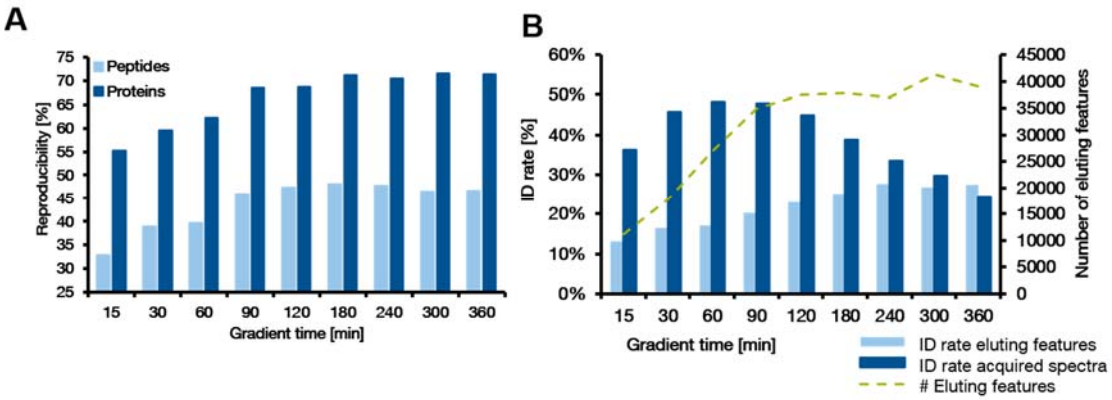


Figure 10 Reproducibility analysis of triplicate measurements of 1.000 ng HeLa digest using different LC gradient times (A) Same as in Fig. 9 but for all gradient times. It is apparent that the recall rate increases with longer gradients but does not substantially improve beyond 90' gradient time. (B) Qualitative reproducibility of the gradient replicates based on the number of identified proteins and peptides.

It could be observed that there was little improvement in recall rates beyond gradients of 90 min and that the number of detected (peptide) features in MS1 spectra also stabilized at a gradient time of 90 min (Fig. 10). In addition, the identification rate of the acquired tandem mass spectra also maximized for 90 min gradients indicating that this is the most productive setting for this particular LC-MS/MS setup which is why this method was used for subsequent experiments.

HD-DDA IN DIFFERENT PROTEOMIC STUDIES

To illustrate the usefulness of the developed HD-DDA method, it was applied to a number of topics frequently investigated in current proteome research. This first concerned the proteomic profiling of human tissues. This type of analysis is more challenging than that of cell lines grown in culture due to the presence of proteins from e. g. body fluids or connective tissue which both increases the number of proteins present as well as the dynamic range of protein concentrations in a sample. In addition, the considerable proteomic variance between individuals generally requires the analysis of multiple replicates in order to arrive at statistically robust results. Given that this puts a high demand on sample throughput, it was investigated how deep one might cover a human tissue proteome when using the HD-DDA approach and allowing a maximum of one day of measurement time. Two-hours turnaround HD-DDA analysis (90 min gradients, see above) of 50 µg total protein separated into 12 fractions by SDS-PAGE from each of 8 human tissues led to the identification of between ~3,700 (pancreas) and ~7,500 (ovary) proteins and a total proteome coverage of ~10,500 proteins (protein FDR of 0.4%; Fig.

11) indicating that large-scale proteomics of human tissues is feasible using HD-DDA in a reasonable amount of time.

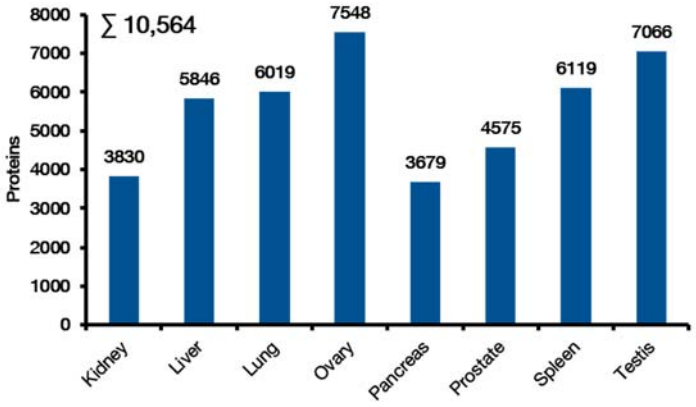


Figure 11 Proteome profiles of eight human organs using a total of 24h of LC-MS/MS time per organ. Two-hours turnaround HD-DDA analysis of 50 µg total protein separated into 12 fractions by SDS-PAGE from each of 8 human tissues led to the identification of between ~3,700 (pancreas) and ~7,500 (ovary) proteins and a total proteome coverage of ~10,500 proteins

As a second example, phosphoproteomics was chosen because the analysis of post-translational modifications (PTMs) is gaining more and more attention notably in signal transduction and drug discovery research[40]. Depth and throughput are also key parameters here because the plasticity of signaling pathways requires deep sampling as well as perturbation or time course experiments[41]. HD-DDA analysis of phosphopeptides enriched by IMAC[42] from A431 skin carcinoma cells led to the identification of ~8,600 phosphorylation sites requiring only 5 hours of instrument time. For ~6,600 of these, the phosphorylation site could be determined with better than 5% false localization rate [26](Fig. 12). Among the identified phospho-proteins are kinases (e.g. EGFR, ABL1, AKT1) as well as transcription factors (e.g. FOXO3, MYC, IRF6) showing that HD-DDA analysis of phosphoproteomes provides considerable depth which should allow it to be applied more generally and easily to a whole range of research topics in the future.

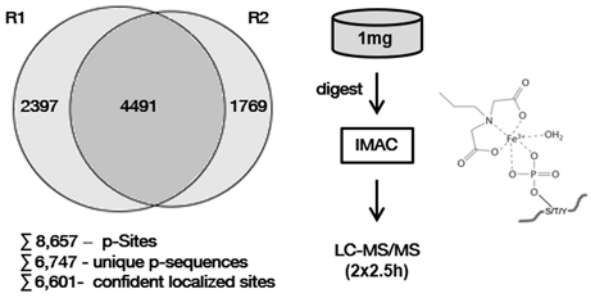


Figure 12 Evaluation of the suitability of HD-DDA for PTM analysis. Analysis of IMAC enriched phosphopeptides in two technical replicates using 5 h total analysis time.

A recent report has shown that TWIMS can improve the quantitative performance of isobaric mass tagging experiments [43, 44]. Notably, the ion mobility separation of peptide precursor ions reduced the well described issue of co-isolation of peptides and the concomitant decrease in accuracy and precision of TMT quantification [45]. In the setting in which TWIMS is applied to peptide fragment ions to enhance duty cycle and sensitivity in MS2 spectra, isobaric tagging is also an attractive application.

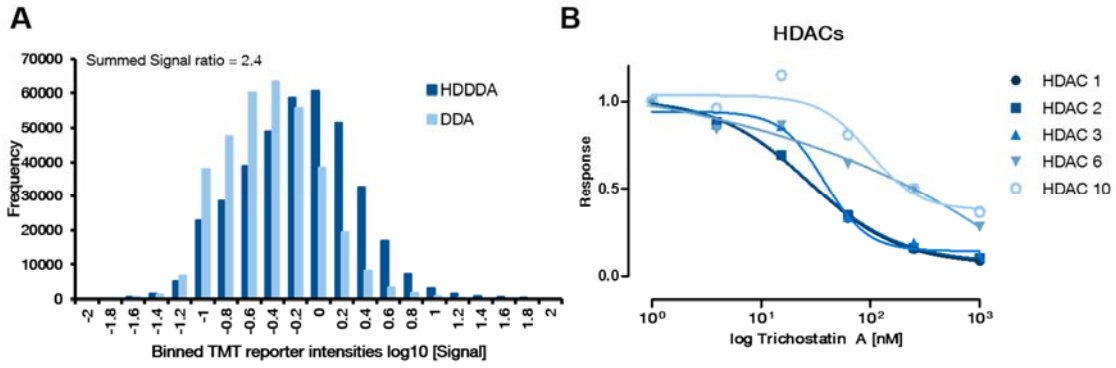


Figure 13 Using MS2 based quantification in conjunction with HD-DDA for chemoproteomic profiling of the HDAC inhibitor Trichostatin A (A) Extracted (log10) TMT intensity from a sample analyzed by standard DDA (light blue bars) and HD-DDA (dark blue bars). (B) Inhibition profiles of all identified HDACs in response to Trichostatin A.

As a proof-of-principle experiment, HD-DDA was used for a chemoproteomic experiment measuring the interaction of the small molecule drug Trichostatin A with its cellular target proteins in K562 myelogenous leukemia cells. Briefly, the nonselective histone deacetylase (HDAC) inhibitor SAHA (suberoylanilide hydroxamic acid, Vorinostat) was immobilized to sepharose beads and pulldowns were performed from lysates of K562 myelogenous leukemia cells. Following tryptic digestion and labeling by tandem mass tags (TMT), the labeled peptides showed a considerably higher reporter ion intensity in

HD-DDA over DDA measurements (Fig. 13A) indicating the expected higher performance of HD-DDA in such applications too. Next SAHA beads were used to profile the interaction of the antibiotic and HDAC inhibitor Trichostatin A with proteins from K562 cells in a quantitative fashion. As reported before [20], SAHA beads capture HDACs class I (HDAC1, 2, 3 and 8) and class II (HDAC6 and 10). Competition of binding using increasing concentrations of Trichostatin A in lysates revealed the potent but differential interaction of Trichostatin A with HDAC1,2 and 3 (low nM) over HDAC6 and 10 (Fig. 13B). The same analysis also revealed co-purification and co-competition of members of the CoRest complex including HDAC1 and 2, LSD1, GSE1, RCOR 1 and 3 (Fig. 14A). These results are in very close agreement with data obtained on the very same samples using a Q-Exactive instrument[20] showing that HD-DDA delivers quantitative data of high quality (Fig. 14B).

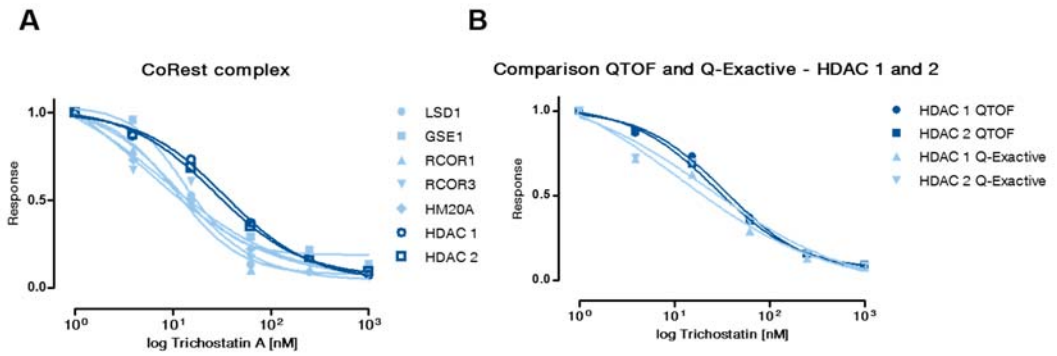


Figure 14 Chemoproteomic characterization of the HDAC inhibitor Trichostatin A using a SAHA matrix and a TMT quantification read out. (A) TMT-quantification of the interaction of Trichostatin A with cellular proteins employing a competition binding assay using immobilized SAHA and increasing doses of Trichostatin A (B) Comparison of the results obtained by HD-DDA vs. the results obtained by Q-Excactive analysis.

CONCLUDING REMARKS

The above results show that the integration of ion mobility spectrometry into DDA acquisition schemes removes the long standing limitation of low duty cycles of Q-TOF instruments and leads to considerably improved sensitivity and productivity of proteomic analysis on such instruments. While ion trap instruments may still offer higher absolute sensitivity owing to their ability to accumulate precursor ions to a desired level prior to fragmentation (a feature that is not available on Q-TOF instruments), TOF instruments have the potential for very fast data acquisition and are high in-spectral dynamic range, both of which are highly desirable for the analysis of the very complex samples. The actual speed of HD-DDA in this first implementation of the method is approximately 6-8 Hz, mainly owing to the time required for on-the-fly data processing within the instrument control software. However, it can be projected that improvements to the software alone will increase speed to 15-20 Hz in the foreseeable future and the method as such has the potential to scale to 30-40 Hz further increasing the prospects of HD-DDA for the rapid and comprehensive analysis of complex proteomes or for the use in high throughput chemical proteomics studies.

ACKNOWLEDGMENTS

Johannes PC Vissers, Christopher J Hughes, Arkadiusz Grzyb, Keith Richardson, Jason Wildgoose for helping with the method development. Hannes Hahne, Benjamin Ruprecht, Isabelle Becher, Barbara Ferreira da Silva and Andrea Hubauer for assistance with sample preparation. Further Dr. Charlotte Daly for fruitful discussions.

ABBREVIATIONS

CID	Collision induced dissociation
DDA	Data dependent acquisition
EGFR	Epidermal growth factor receptor
ESI	Electrospray ionization
HDAC	Histone deacetylase
HD-DDA	High definition data dependent acquisition (ion mobility enhanced DDA)
HD-MRM	High definition multi reaction monitoring (ion mobility enhanced MRM)
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
oaTOF	Orthogonal acceleration time-of-flight
Q-TOF	Quadrupole Time-of-flight
SAHA	Suberoylanilide hydroxamic acid
TMT	Tandem mass tags
TWIMS	Traveling wave ion mobility separation

REFERENCES

1. Pinto, S.M. et al. Functional annotation of proteome encoded by human chromosome 22. *Journal of Proteome Research*, 140326051637001 (2014).
2. Mallick, P. & Kuster, B. Proteomics: a pragmatic perspective. *Nat Biotechnol* **28**, 695-709 (2010).
3. Hebert, A.S. et al. The One Hour Yeast Proteome. *Mol Cell Proteomics* (2013).
4. Hahne, H. et al. DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Methods* **10**, 989-991 (2013).
5. Pirmoradian, M. et al. Rapid and Deep Human Proteome Analysis by Single-dimension Shotgun Proteomics. *Mol Cell Proteomics* **12**, 3330-3338 (2013).
6. Bantscheff, M. & Kuster, B. Quantitative mass spectrometry in proteomics. *Anal Bioanal Chem* **404**, 937-938 (2012).
7. Zubarev, R.A. & Makarov, A. Orbitrap mass spectrometry. *Anal Chem* **85**, 5288-5296 (2013).
8. 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* **83**, 5442-5446 (2011).
9. Guilhaus, M., Selby, D. & Mlynski, V. Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrom Rev* **19**, 65-107 (2000).
10. Hashimoto, Y., Hasegawa, H., Satake, H., Baba, T. & Waki, I. Duty cycle enhancement of an orthogonal acceleration TOF mass spectrometer using an axially-resonant excitation linear ion trap. *J Am Soc Mass Spectrom* **17**, 1669-1674 (2006).
11. Brenton, A.G. et al. Improvement of the duty cycle of an orthogonal acceleration time-of-flight mass spectrometer using ion gates. *Rapid Commun Mass Spectrom* **21**, 3093-3102 (2007).
12. Giles, K. et al. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* **18**, 2401-2414 (2004).
13. Chernushevich, E.I. Duty cycle improvement for a quadrupole-time-of-flight mass spectrometer and its use for precursor ion scans. *European Journal of Mass Spectrometry* **6**, 417-479 (2000).
14. Shvartsburg, A.A. & Smith, R.D. Fundamentals of traveling wave ion mobility spectrometry. *Anal Chem* **80**, 9689-9699 (2008).
15. Pringle, S.D. et al. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry* **261**, 1-12 (2007).
16. Wisniewski, J.R., Zougman, A. & Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res* **8**, 5674-5678 (2009).
17. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-670 (2003).
18. Ficarro, S.B. et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* **20**, 301-305 (2002).
19. Wilm, M. et al. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466-469 (1996).
20. Bantscheff, M. et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotechnol* **29**, 255-265 (2011).
21. Bantscheff, M. et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* **25**, 1035-1044 (2007).

22. Medard, G. et al. Optimized Chemical Proteomics Assay for Kinase Inhibitor Profiling. *J Proteome Res* (2015).
23. Brosch, M., Yu, L., Hubbard, T. & Choudhary, J. Accurate and sensitive peptide identification with Mascot Percolator. *J Proteome Res* **8**, 3176-3181 (2009).
24. Kall, L., Canterbury, J.D., Weston, J., Noble, W.S. & MacCoss, M.J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **4**, 923-925 (2007).
25. van den Toorn, H.W. et al. RockerBox: analysis and filtering of massive proteomics search results. *J Proteome Res* **10**, 1420-1424 (2011).
26. Marx, H. et al. A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat Biotechnol* **31**, 557-564 (2013).
27. Savitski, M.M. et al. Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics* **10**, M110 003830 (2011).
28. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).
29. Wilhelm, M., Kirchner, M., Steen, J.A. & Steen, H. mz5: space- and time-efficient storage of mass spectrometry data sets. *Mol Cell Proteomics* **11**, O111 011379 (2012).
30. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534-2536 (2008).
31. Vizcaino, J.A. et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* **41**, D1063-1069 (2013).
32. Domon, B. & Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology* **28**, 710-721 (2010).
33. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212-217 (2006).
34. Abbatiello, S.E., Mani, D.R., Keshishian, H. & Carr, S.A. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clinical chemistry* **56**, 291-305 (2010).
35. Sherman, J., McKay, M.J., Ashman, K. & Molloy, M.P. How specific is my SRM?: The issue of precursor and product ion redundancy. *Proteomics* **9**, 1120-1123 (2009).
36. Helm, D. et al. in HUPO 12th Annual World Congress (Yokohama, Japan; 2013).
37. Anastassiadis, T., Deacon, S.W., Devarajan, K., Ma, H. & Peterson, J.R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1039-1045 (2011).
38. Gallien, S. et al. Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol Cell Proteomics* **11**, 1709-1723 (2012).
39. 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. *Journal of Proteome Research* **10**, 1785-1793 (2011).
40. Olsen, J.V. & Mann, M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics* **12**, 3444-3452 (2013).
41. Phanstiel, D.H. et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Methods* **8**, 821-827 (2011).
42. Ruprecht, B. et al. Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. *Mol Cell Proteomics* **14**, 205-215 (2015).
43. Thompson, A. et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).
44. Shliaha, P.V. et al. Additional Precursor Purification in Isobaric Mass Tagging Experiments by Traveling Wave Ion Mobility Separation (TWIMS). *J Proteome Res* **13**, 3360-3369 (2014).

45. Savitski, M.M. et al. Measuring and managing ratio compression for accurate iTRAQ/TMT quantification. *J Proteome Res* **12**, 3586-3598 (2013).

CHAPTER IV

SELECTIVITY PROFILING OF PONATINIB

ABSTRACT

Protein kinases are key components in signal transduction pathways and are in consequence attractive drug targets for diseases like cancer and inflammation. About 30 small molecule drugs targeting kinases are approved for clinical use and more than 200 further compounds are under investigation in clinical trials. Comprehensive knowledge about the target space of a kinase inhibitor is crucial for the proper interpretation of its biological effects. The Kinobeads technology, a chemical proteomics approach, has proven to be a complementary technique for a systematic assessment of kinase inhibitor selectivity. Ponatinib (AP24534) is a potent tyrosine kinase inhibitor which was approved by the FDA in the year 2012 for the treatment of chronic myeloid leukemia. Ponatinib was chosen for a selectivity profiling since it was withdrawn from the market in 2013 due to the occurrence of severe side effects. Here, the described selectivity profile of ponatinib was assessed using Kinobeads in a four cancer cell line mixture as well as human placenta, enabling the exposure of ponatinib to 308 kinases. This led to the identification of a previously unknown set of kinases that show inhibition by ponatinib. Among these are MAPK14 (p38α) and ZAK (MLTK) as two new potent targets, both of which are known to play a critical role in the regulation of the JNK and p38 MAPK pathways involved in cancer and inflammation. These results could be supported with data from molecular docking experiments as well as by an activity assay and helped to explain previous inflammation inhibitory effects of ponatinib.

INTRODUCTION

Protein kinases play a key role in signal transduction and consequently are attractive drug targets in diseases such as cancer and inflammation[1-5]. Currently, about 30 small molecule drugs have been approved for clinical use and more than 200 further compounds are under investigation in clinical trials[6]. Among these approved compounds is ponatinib (AP24534[7]}), a potent tyrosine kinase inhibitor which was approved by the FDA at the end of 2012[8] for the treatment of chronic myeloid leukemia (CML). The advantage of ponatinib compared to other CML drugs like imatinib, nilotinib and dasatinib is the potent inhibition of several mutated forms of BCR-ABL[7]. The initial PAGE (Ponatinib Ph+ ALL and CML Evaluation)[9] study with patients intolerant or resistant to other tyrosine kinase inhibitors showed significant improvements and lead to ponatinib's approval. A follow up study, started in the end of 2013 to evaluate the use of ponatinib as a first-line treatment for CML, was terminated due to the occurrence of cardio-vascular side effects[10-13]. Considering that ponatinib, like most kinase inhibitors (KIs), targets the structurally conserved ATP binding pocket within the kinase domain, these small molecules have a high probability of targeting multiple protein kinases or ATP-hydrolysing/binding enzymes. Indeed, many molecules have been found to inhibit a wider range of kinases than that for which they had initially been designed[14-17]. Therefore, the better one can elucidate the real target spectrum of a KI, the better one understands its mechanism-of-action (MoA) and, in consequence, learn its full therapeutic potential or manage toxic side effects[5]. *In vitro* methods based on large panels of recombinant kinase domains are traditionally used to assess the selectivity of kinase inhibitors (e.g. ponatinib)[7, 18-20]. Albeit powerful and commonly employed, these assays are blind for many additional factors that influence kinase activity in cells (e.g. regulatory domains, interacting proteins, the proteins' conformational state and/or its post translational modification (PTM) status)[17, 20, 21]. In addition, panels of enzyme assays are by default limited to measure only the included proteins. Consequently they may miss targets that are important to understand or predict the efficacy of a compound in cells.

Quantitative mass spectrometry based chemical proteomics has proven to be an complementary technique for assessing the selectivity of kinase inhibitors[14, 16, 20, 22-24]. The use of immobilized unselective small molecule inhibitors as affinity purification reagents enables a straight forward approach towards the identification of the interaction space of

a drug[22, 23, 25-28]. Kinobeads technology can be configured as a competition binding assay in combination with a quantitative mass spectrometry read out. This set-up allows for the determination of the selectivity of a small molecule kinase inhibitor against hundreds of proteins in a single experiment[22, 29, 30].

In this study, the target spectrum of ponatinib has been evaluated in cell lines and human tissue known from proteome profiling data to express 451 respectively 309 and protein kinases. With the Kinobead technology it was possible to enrich 308 of these in total. This enabled the identification of new potent targets of ponatinib possessing a high potency (K_D < 100 nM) in the range of ponatinibs' plasma concentration[31]. Single results could be supported by molecular docking and additionally activity assay data. Among the newly identified target kinases are ZAK (MLTK) and MAPK14 (p38α). For both exist evidence for central rolls in inflammatory and cancer signaling[32, 33]. ZAK was even found to be inhibited with similar potency as ponatinib's main target BCR-Abl.

EXPERIMENTAL PROCEDURES

SAMPLE PREPARATION

FOUR CELL LINE MIX

K562 and MV4-11 cells were grown in roller culture in RPMI1640 medium supplemented with 10% FBS and antibiotic antimycotic solution (Sigma A5955). Cells were harvested upon density (approx. 5 x 10^6), centrifuged and washed with cold PBS. SKNBE2 cells were grown in stationary culture (15 cm dishes) in DMEM/HAMS medium supplemented with 10% FBS and antibiotic antimycotic solution (Sigma A5955). Cells were harvested upon confluence by mechanical detachment followed by centrifugation and washing with cold PBS. Colo205 cells were produced by fed-batch fermentation in 1.8 L scale in a Braun Biostat B2 (37 °C, pH = 7.1 CO_2 regulated, 30 rpm, O_2 maintained at starting value) in RPMI1640 medium supplemented with 10% FBS and antibiotic antimycotic solution (Sigma A5955). Cells were harvested upon density (approx. 5 x 10^6), centrifuged and washed with cold PBS. The cells were lysed in lysis buffer containing 50 mM Tris/HCl, pH 7.5, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 1 mM Na_3VO_4, 0.8% NP40, 0.375 mM NaF, 1 mM DTT including protease inhibitors (SigmaFast protease inhibitor tablet S8820) and phosphatase inhibitors (Phosphatase Inhibitor Cocktail 3, Sigma-Aldrich, Munich, Germany). The lysate was ultracentrifuged for 1 h at 4 °C and 145,000g and afterwards stored at -80°C until usage. The four cell line mix was kindly provided by Huichao Qiao.

PLACENTA LYSATE

Placental tissue was obtained from the local hospital in Freising following informed consent by the donor. Approximately 20 mg of fresh frozen placental tissue was washed three times with 1 ml ice-cold phosphate buffered saline (PBS) supplemented with protease inhibitor (SigmaFAST, Sigma-Aldrich, Germany) and 250 µl lysis buffer was added. The lysis buffer comprised 8 M urea, 50 mM TrisHCl (pH 8.5), 10 mM using tris(2-carboxyethyl)phosphine (TCEP) and 40 mM chloro-acetamide (CAA), supplemented with protease inhibitors. TCEP and CAA enable one-step reduction and alkylation (PMID: 24487582). Placental tissue was lysed by bead milling using the Precellys 24 homogenizer (Bertin Technologies, France) at 5,000 rpm for 2x10s with 5s pause. After the tissue homogenization, the protein concentration was determined via Bradford assay (Pierce, Germany). The lysate (300 µg total protein) was diluted with 50 mM TrisHCl to <2 M and

subsequently digested with trypsin (2x 1:50 protease:protein at 37°C; 4h pre-digestion followed by digestion overnight). Trypsin digestion was stopped by the addition of 100% FA to adjust the pH <2 and precipitates were removed by centrifugation (20,000 x g, 2min). Tryptic digests were desalted using solid-phase extraction with Sep-Pak C18 cartridges (Waters, Germany). Briefly, Sep-Pak columns were conditioned with buffer B (0.1% FA in 60% acetonitrile/40% water v/v) and equilibrated with buffer A (0.1% FA). The acidified sample was loaded onto the column and washed with buffer A and then eluted with 3x 100 µl of buffer B. The eluted samples were dried *in vacuo* and stored at -80 °C prior to usage.

KINOBEAD PULLDOWN AND ACTIVITY ASSAY

The procedure has already been described elsewhere[22]. In Briefly, cell or placenta lysates were diluted with equal volumes of 1× compound pull down (CP) buffer (50 mM Tris/HCl, pH 7.5, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 20 mM NaF, 1 mM sodium ortho-vanadate, 1 mM DTT, protease inhibitor (SigmaFast protease inhibitor tablet S8820) and phosphatase inhibitors (Phosphatase Inhibitor Cocktail 3, Sigma-Aldrich, Munich, Germany). If required, lysates were further diluted to a final protein concentration of 5 mg/mL using 1× CP buffer supplemented with 0.4% NP-40 (CP-0.4). The cell mixes were prepared as 1:1:1:1 mixtures regarding the total amount of proteins and as determined by Bradford assay.

For selectivity profiling experiments in 96 well plates, the diluted cell mix lysates/placenta lysate (5 mg total proteins / well) were incubated for 45 min at 4 °C in a head-over shaker with 0 nM (DMSO control), 3 nM, 10 nM, 30 nM, 100 nM, 300 nM, 1 µM, 3 µM, or 30 µM of the kinase inhibitor dissolved in DMSO. KBγ[22] or KBɖ, F2[34], Compound 18[35], VI compound[36] and Compound 2 (linkable CZC8004[16], (35 µL settled beads resuspended with 50% glycerol, were washed with water, CP buffer and equilibrated with CP-0.4 buffer) were incubated with the lysates at 4 °C for 30 min. The DMSO control lysate was recovered and incubated similarly with KBγ or KBɖ as a pull down of pull down experiment in order to calculate the depletion factor. The beads were then washed (3 mL CP-0.4 buffer followed by 2 mL CP-0.2) and the bound proteins subsequently eluted by incubation for 30 min at 50 °C with 60 µL 2× NuPAGE LDS sample buffer (Invitrogen, Darmstadt, Germany) containing 50 mM DTT and centrifugation. The reduced eluates (30 µL) were alkylated with chloroacetamide (3 µL, 550 mM) and the proteins were desalted and concentrated by a short electrophoresis (about 0.5 cm) on a 4–12% NuPAGE gel (Invitrogen). In-gel trypsin digestion was performed according to standard procedures[37].

The coupling of the inhibitors to the sepharose beads was kindly performed by Stefanie Heinzlmeier.

The activity assay was performed by Reaction Biology Corporation (Malvern, Pennsylvania) and its procedure is described elsewhere[20].

HOMOLOGY DOCKING

The search for a template structure was carried out by using the position-specific iterated BLAST (PSI-BLAST) algorithm [38]. For this search the standard parameter were used with one limitation to the RCSB PDB databank [39]. The blast search resulted in a number of similar structures for the first 270 amino acids of ZAK. By this, it was possible to identify the most similar template to be the mitogen-activated protein kinase kinase kinase 9 (MAP3K9) [40]. They share an identity of 42% and a similarity of 58% with an expectation value of 3e-55. In an additional step using the Uniprot database it was confirmed that the binding site (ATC-binding loop) of ZAK is covered by the template structure [41]. The binding sites of ZAK and MAP3K9 are highly similar [42]. Based on this knowledge the homology model was built on the template structure of MAP3K9. To build the sequence alignment, the alignment type “clustal omega” was used [43]. The homology modelling process was performed by a Python script of MODELLER [ref:modeller]. Building the three-dimensional structure of ZAK the option “automodel” was used. For this, all atoms except water molecules were considered for building the model. The resulting homology model was cleaned to correct problems due to structural disorder, protein residue connectivity and bond orders, missing side-chains or backbone atoms (Discovery studio, Accelrys Inc., http://www.accelrys.com/dstudio). For this purpose Discovery studio 4.1 function “clean protein” was used (Discovery studio, Accelrys Inc., http://www.accelrys.com/dstudio). Further steps for energy minimization were also performed with Discovery Studio. Conjugate gradients with standard parameters and 1000 steps for energy minimization for the complete model were applied. After this step, the P-loop was additionally refined. The loop refinement searches for low energy conformations for the P-loop by optimising the side chains for the residues of this loop (Discovery studio, Accelrys Inc., http://www.accelrys.com/dstudio). Homology docking results were kindly provided by Björn Gohlke from the Charité Berlin.

MOLECULAR DOCKING

Docking studies were carried out using GOLD 5.2 with the fitness score (ChemScore) scoring function [44]. The binding site was defined from the bound ligand of the crystal structure of MAP3K9. The distance around this ligand (radius 10.0 Å) was defined as the active site, which also covered the P-loop. The docking process was performed by using the standard parameters. Molecular docking results were kindly provided by Björn Gohlke from the Charité Berlin.

LC-MS/MS MEASUREMENTS

The optimized conditions for the liquid chromatography tandem mass spectrometry measurements featured an Eksigent nanoLC-Ultra 1D+ (Eksigent, Dublin, CA) coupled to an Orbitrap Elite instrument (Thermo Scientific, Bremen, Germany). The peptides were delivered to a trap column (100 µm×2 cm, packed in-house with Reprosil-Pur C18-AQ 5 µm resin, Dr. Maisch, Ammerbuch, Germany) at a flow rate of 5 µL/min in 100% solvent A (0.1% formic acid, FA, in HPLC grade water). After 10 min of loading and washing, peptides were transferred to an analytical column (75 µm×40 cm, packed in-house with Reprosil-Gold C18, 3 µm resin, Dr. Maisch, Ammerbuch, Germany) and separated at a flow rate of 300 nL/min using a 90 min gradient ranging from 4% to 32% solvent C in B (solvent B: 0.1% FA and 5% DMSO in HPLC grade water, solvent C: 0.1% FA and 5% DMSO in acetonitrile). The eluent was sprayed via stainless steel emitters (Thermo) at a spray voltage of 2.2 kV and a heated capillary temperature of 275°C. The Orbitrap Elite instrument was operated in data-dependent mode, automatically switching between MS and MS2. Full scan MS spectra (m/z 360 – 1300) were acquired in the Orbitrap at 30,000 (m/z 400) resolution using an automatic gain control (AGC) target value of 1e6 charges. Tandem mass spectra of up to 15 precursors were generated in the multipole collision cell by using higher energy collisional dissociation (HCD) (AGC target value 2×1e4, normalised collision energy of 30%) and analysed in the Orbitrap at a resolution of 15,000. Precursor ion isolation width was set to 2.0 Th, the maximum injection time for MS/MS was 100 ms and dynamic exclusion was set to 20 s. Measurements were kindly performed by Susan Kläger.

PROTEIN IDENTIFICATION AND QUANTIFICATION

For the drug profiling, MaxQuant (MQ) software (version 1.4.0.5) and Uniprot database (v22.07.143) were used for intensity based label free quantification. For the database

search, carbamidomethyl cysteine was set as a fixed modification and N-terminal protein acetylation and methionine oxidation as variable modifications. Trypsin was specified as the proteolytic enzyme and up to two missed cleavages were allowed. The mass tolerance of the precursor ion was set to 10 ppm and for fragmentations to 0.05 Da. For identification, 0.01 peptide and protein FDRs were used. Feature matching between raw files was enabled, using a match time window of 1 min. Averaged label free quantification (LFQ) intensity values were used to calculate protein ratios with the DMSO sample as reference. MQ data were filtered for reverse identifications (false positives) and contaminants. IC_{50} curves were drawn using an in-house software tool adapting an open source statistical software tool[45]. Selected dose-inhibition curves where analysed in GraphPad Prism (v. 5.04).

RESULTS AND DISCUSSION

SELECTIVITY PROFILING USING KINOBEADS

In order to get a broad coverage of ponatinib's binding spectrum, it was subjected to a set of different kinase selectivity profilings using the Kinobead technology[16, 22] (Fig 1A). In brief, competition binding assays using Kinobeads (Fig. 1B) were performed in a cell mix lysate expressing 451 protein kinases and placenta tissue lysate expressing 309 protein kinases. To this end, lysates were pre-incubated with increasing concentrations of ponatinib and subsequently subjected to Kinobead pulldowns and MS-based protein identification and quantification.

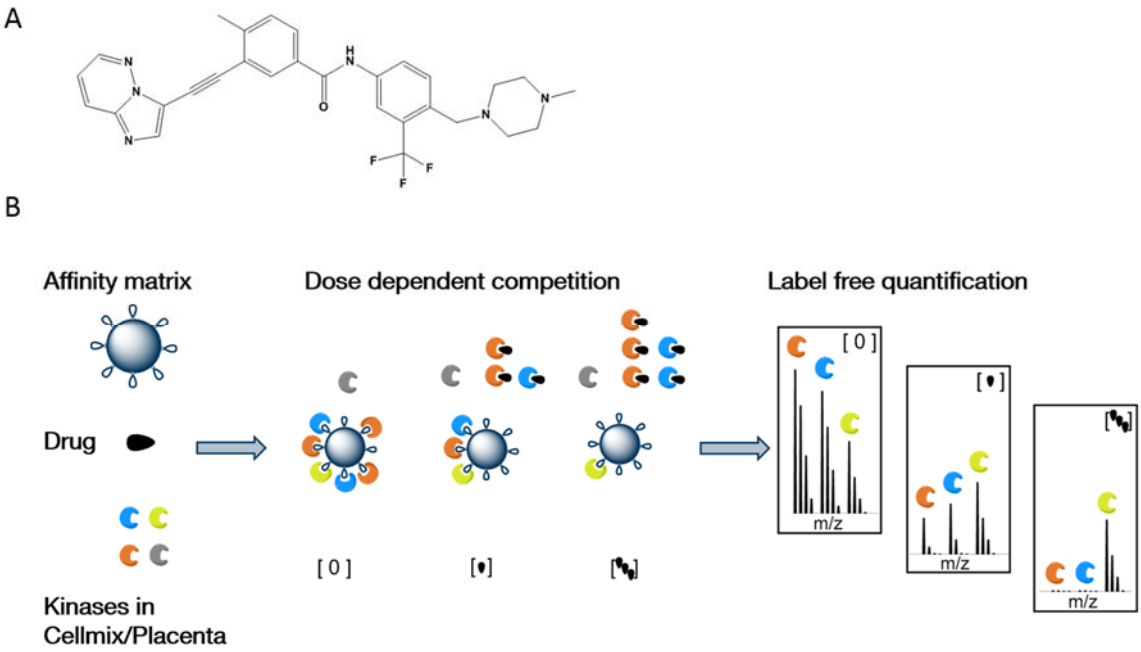


Figure 1 Overview (A) Chemical structure of ponatinib (B) Illustration of the Kinobead workflow as a drug competition experiment. Protein targets that bind the drug in the lysate show a dose-dependent reduction in binding to the unspecific kinase inhibitors immobilized on the Kinobeads, whereas unaffected kinases do not show this effect. (adapted from Medard et al, 2015[22])

Protein targets that bind the drug in the lysate show a dose-dependent reduction in binding to the unspecific kinase inhibitors immobilized on the Kinobeads, whereas unaffected kinases do not show this effect. In order to obtain a high level of comprehensiveness in the enrichment of the expressed kinases, different experimental set-ups were used. Two sets of affinity matrices (KBɣ and KBɖ) were used. KBɖ contained

among others, compound 18[35] and the F2 probe[34] which are especially designed for the enrichment of vascular endothelial growth factor receptors (VEGFRs) and fibroblast growth factor receptors (FGFRs) respectively.

KINOME COVERAGE

The capabilities of the different enrichment set-ups are depicted in Figure 2. The size of the dot correlates to the MS intensity of the kinase enriched in this particular experiment.

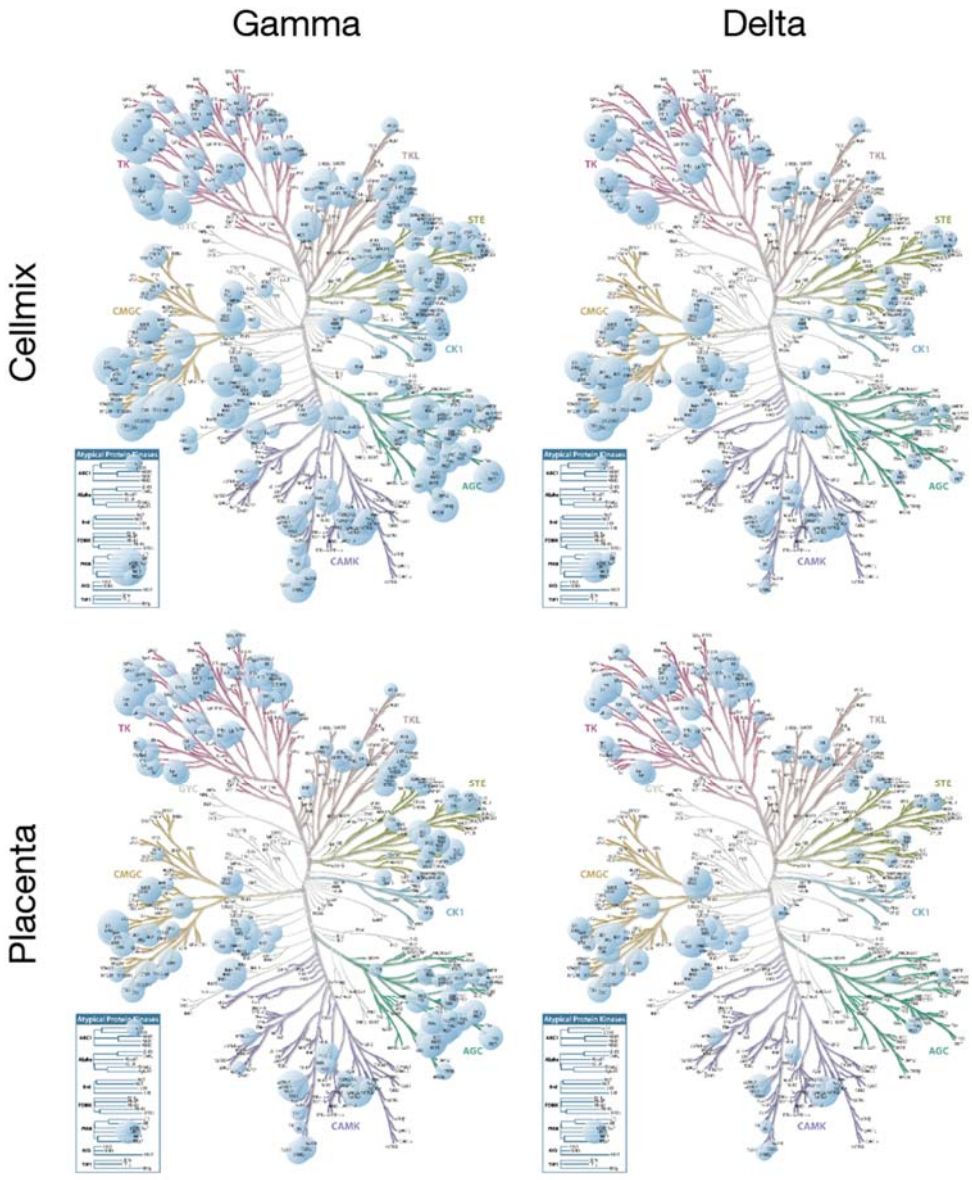


Figure 2 Kinometrees for the four different experiments elucidating the enrichment capabilities of the different set-ups. The size of the dots correspond to MS intensity. Briefly, the combination of cellmix and KBγ performed best with 255 enriched kinases whereas using placenta lysate and KBδ with only 135 kinases.

The combination of cellmix and KBγ performed best in terms of overall kinome tree coverage and as well in the amount of identified kinases (255). Least performant was the pull down using placenta lysate and KBδ with only 135 kinases. The data for the different Kinobead generations (γ and δ) are in agreement with a recently published Kinobead optimization paper[22]. Further, it can be seen that the used mixture of 4 different cancer cell lysates (Colo205, SKNBE2, K562 and MV11) in general allows a higher kinome coverage compared to placenta lysate.

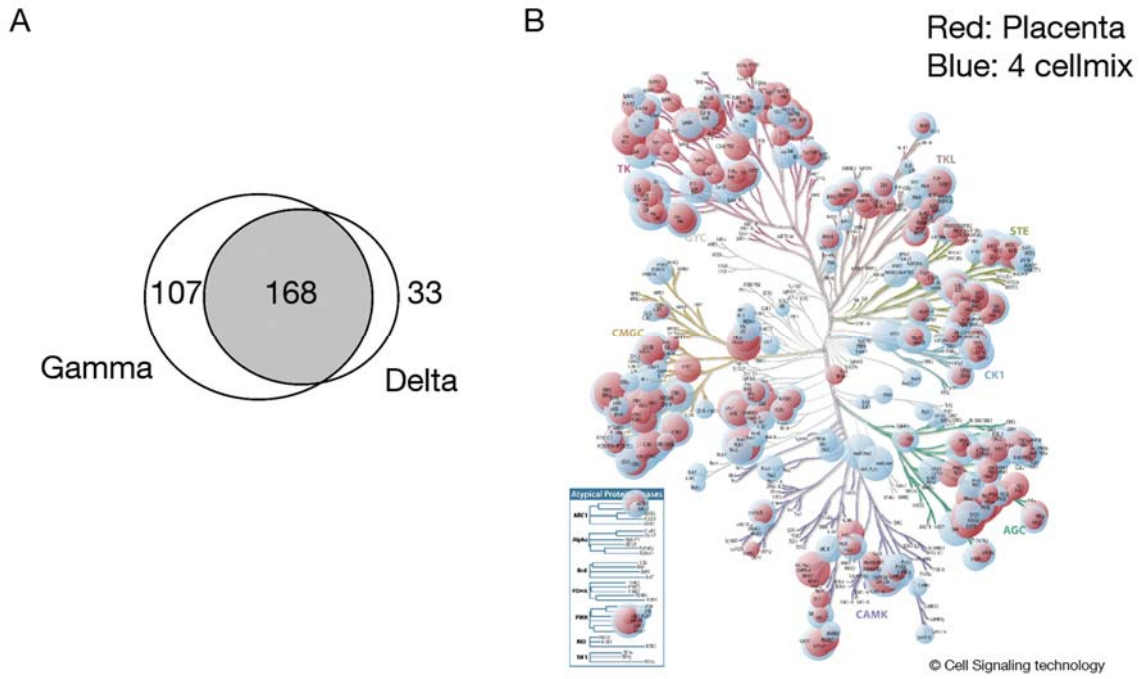


Figure 3 Selectivity screening - summary (A) Venn diagram comparing the kinases enriched by the KBγ and KBδ (B) Kinometree comparing the different biological backgrounds (blue cellmix, red placenta) the experiment was performed in. The size of the dot positively correlates with the MS intensity.

In total it was possible to enrich 308 kinases with these two affinity matrices (Fig. 3A). The use of the two different biological matrices (cell mix and placenta lysate) proved to be a valuable combination for expanding the coverage of the phylogenetic kinome tree (Fig. 3B) showing kinases identified from every branch of the human kinome, covering approx. 60% of the genome encoded kinome. This breadth of kinome coverage provided the ideal basis for identifying previously unknown interaction partners of ponatinib.

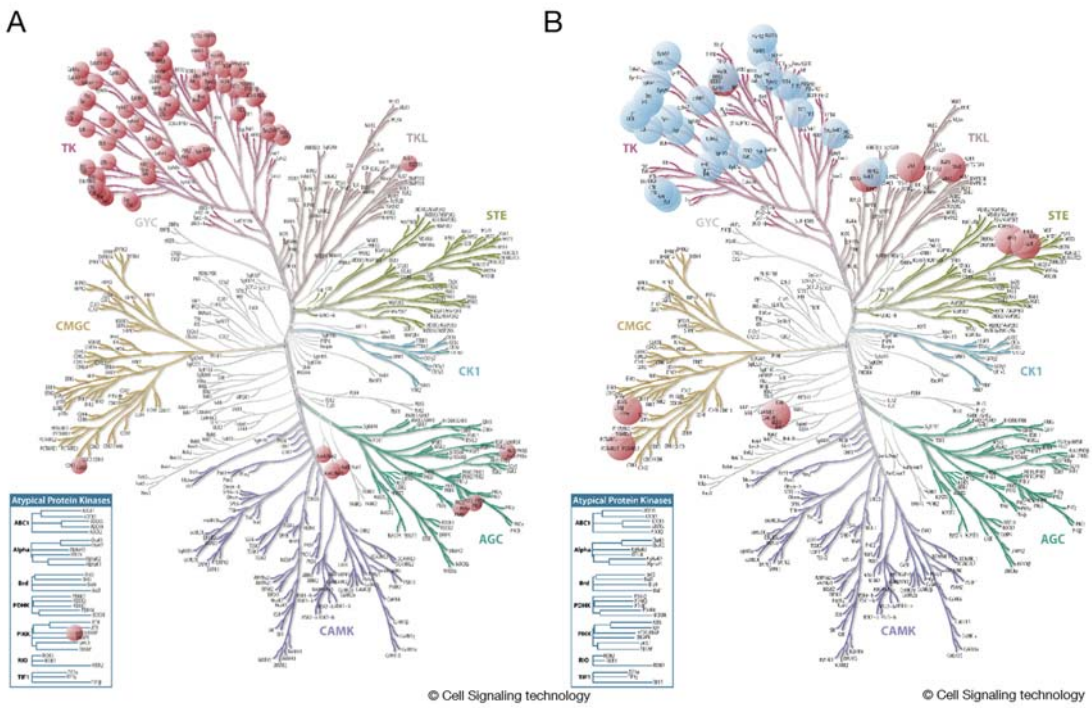


Figure 4 New versus known targets (A) Kinometree highlighting all kinases that have been identified in the initial screening (B) Kinometree highlighting in blue the known targets of ponatinib from the initial screening and in red newly identified targets from this screening with an K_d< 500 nM.

In the initial publication[7] it could already be seen that ponatinib is a kinase inhibitor with a broad target spectrum of >30 targets showing a IC_{50} of < 10 nmol, especially in the tyrosine kinase family (Fig. 4A). Given the structural homology of the ATP binding pocket of kinases, this is not a surprising observation and makes it an interesting compound for further applications. The goal of the present study was to expose ponatinib to a broader set of kinases to elucidate additional interactions across the kinome. Indeed through this experiments it was possible to identify a set of new and highly potent targets deriving from the Tyrosine Kinase Like (TKL), Serine/Threonine-specific kinases (STE) and CMGC families (Figure 4B, Table 1). Among these are kinases involved in the innate and adaptive immune response (IRAK1 and RIPK2) as well as in inflammatory processes e.g. MAPK14 (p38α kinase) and ZAK. Differences between our potencies and the literature values can be explained through the fact that the Kinobead technology uses the fully expressed protein including PTMs (etc.) whereas in the initial screening only truncated kinase domains were expressed.

Table 1 K_d and IC50 [nM] values of the identified targets in this assay and compared to the recombinant kinase assay data from the literature

Kinase	Literature IC50 [nM][7]	Activity Assay EC50 [nM]	Cellmix γ K_d [nM]	Placenta γ K_d [nM]	Cellmix ɖ K_d [nM]	Placenta ɖ K_d [nM]
FLT3*	13		1	no ID	5	no ID
ABL2*	1		1	no Quan	no Quan	no ID
ABL1*	0.4	< 1	2	13	28	no ID
DDR1			3	7	22	21
RET	0.2		3	no ID	11	no ID
EPHA2	2		7	25	no ID	34
MAPK14*		27	7	43	no Quan	no Quan
ZAK*		< 1	8	17	37	66
MAP4K2*			9	no Quan	no ID	no ID
LCK*	0.3		9	no Quan	23	no Quan
LYN*	0.2		10	4	34	16
RIPK2*			15	42	no Quan	no ID
VEGFR2	2		no ID	15	no ID	52
FRK*	1		20	no Quan	no ID	no ID
MAP4K1*		4	26	no ID	no ID	no ID
DDR2	16		29	113	40	9
FYN*	0.4		31	8	207	35
YES1*	1		35	58	153	68
PTK2B*	35		48	no Quan	733	no Quan
IRAK1*		40	59	no ID	no Quan	no ID
EPHA4	1		68	no Quan	no ID	no ID
SRC*	5		90	33	no Quan	98
HCK*	0.1		102	4	143	15
CSK*	13		102	113	no Quan	no ID
BRAF			297	no ID	no ID	no ID
FGFR1	2		no Quan	947	514	571
PDGFRA	1		no ID	no Quan	no ID	no Quan
KIT	13		no ID	no Inhibition	no ID	no Quan

PONATINIB INHIBITS INFLAMMATORY RESPONSE VIA JNK AND P38 MAPK PATHWAYS

The kinase selectivity data showed clearly a reproducible competition effect for the kinases ZAK (Fig. 5A) and MAPK14 (Fig. 5B) by ponatinib. ZAK could be detected in 4/4 selectivity profiles, whereas MAPK14 was only enriched by the KBγ.

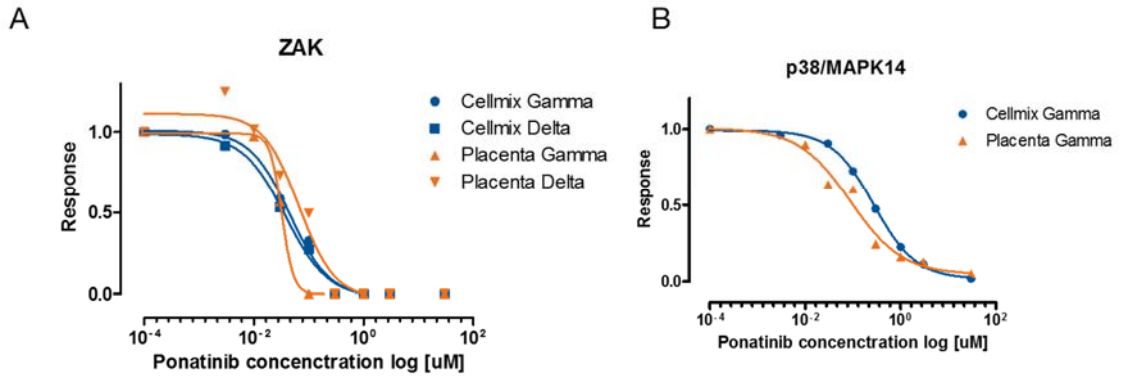


Figure 5 Kinobead inhibition profiles of ZAK and MAPK14 (A) Dose response profiles of ZAK by ponatinib from all 4 biological experiments. (B) Same as (A) but for MAPK14. Note that MAPK14 could only be identified in the KBγ experiments.

For further confirmation of the data molecular docking experiments were performed. Due to the absence of a crystal structure for Mitogen-activated protein kinase kinase kinase MLT (ZAK), a homology model had to be created to analyze the binding pose of ponatinib for ZAK. Based on this model, a docking analysis was performed. The ZAK model was superposed with MAP3K9 to analyze the reliability of the generated model. The superposition of the ZAK model and the template structure is visualized in Fig. 6A and reveals a strong similarity. To validate the quality of the model a Ramachandran plot was created (Fig. 6B). As it is displayed, most of the angles for the residues are in favorable regions.

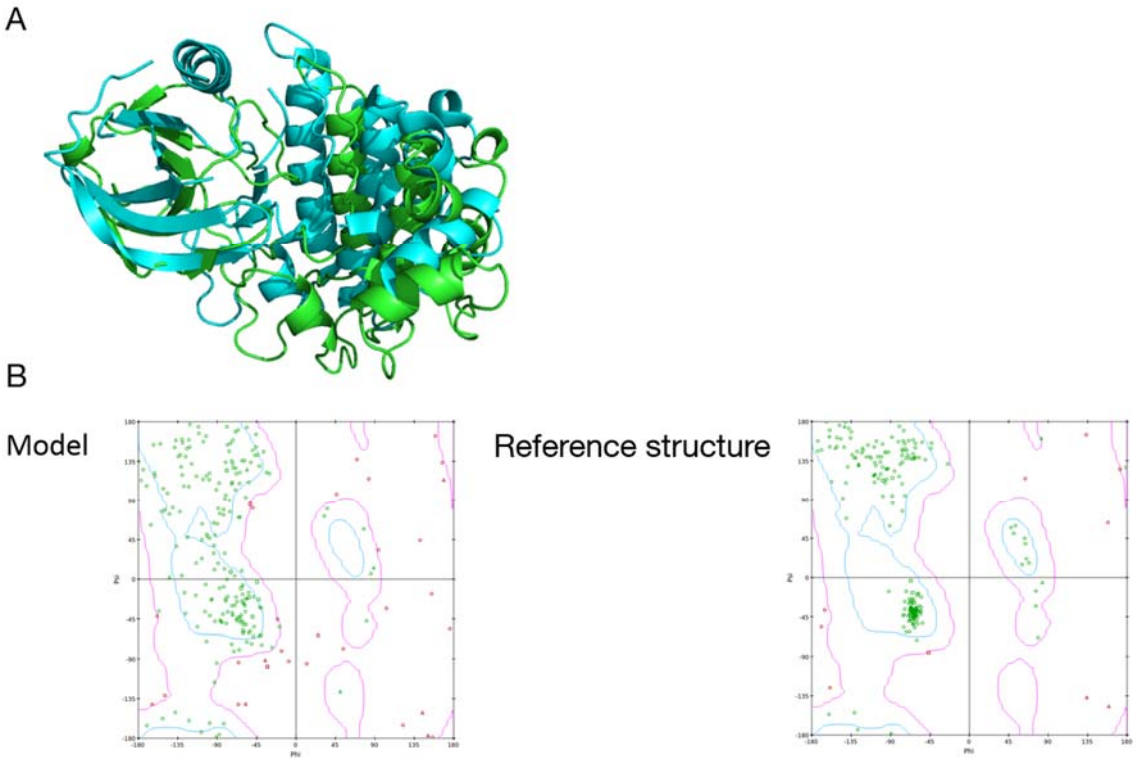


Figure 6 Homology model for ZAK (A) Superposition of homology model (ZAK) with reference structure 3DTC. (B) On the left picture it can be seen that some phi/psi-angles are outside 'allowed' regions in the Ramachandran plot. In comparison the Ramachandran plot for the crystal structure of 3DTC (reference structure) is depicted. Results were kindly provided by Björn Gohlke from the Charité Berlin.

Despite slight differences between the model and the reference structure, the model's angles are still realistic and suitable for purpose. ATP-binding loop (P-loop) of these two proteins were functionally characterized and shown to be similar to each other[41]. This finding also matches the similar superposition of the P-loop from the model at that from the template structure. Fig. 7A and 7B show the docking poses of ponatinib for ZAK and MAPK14.

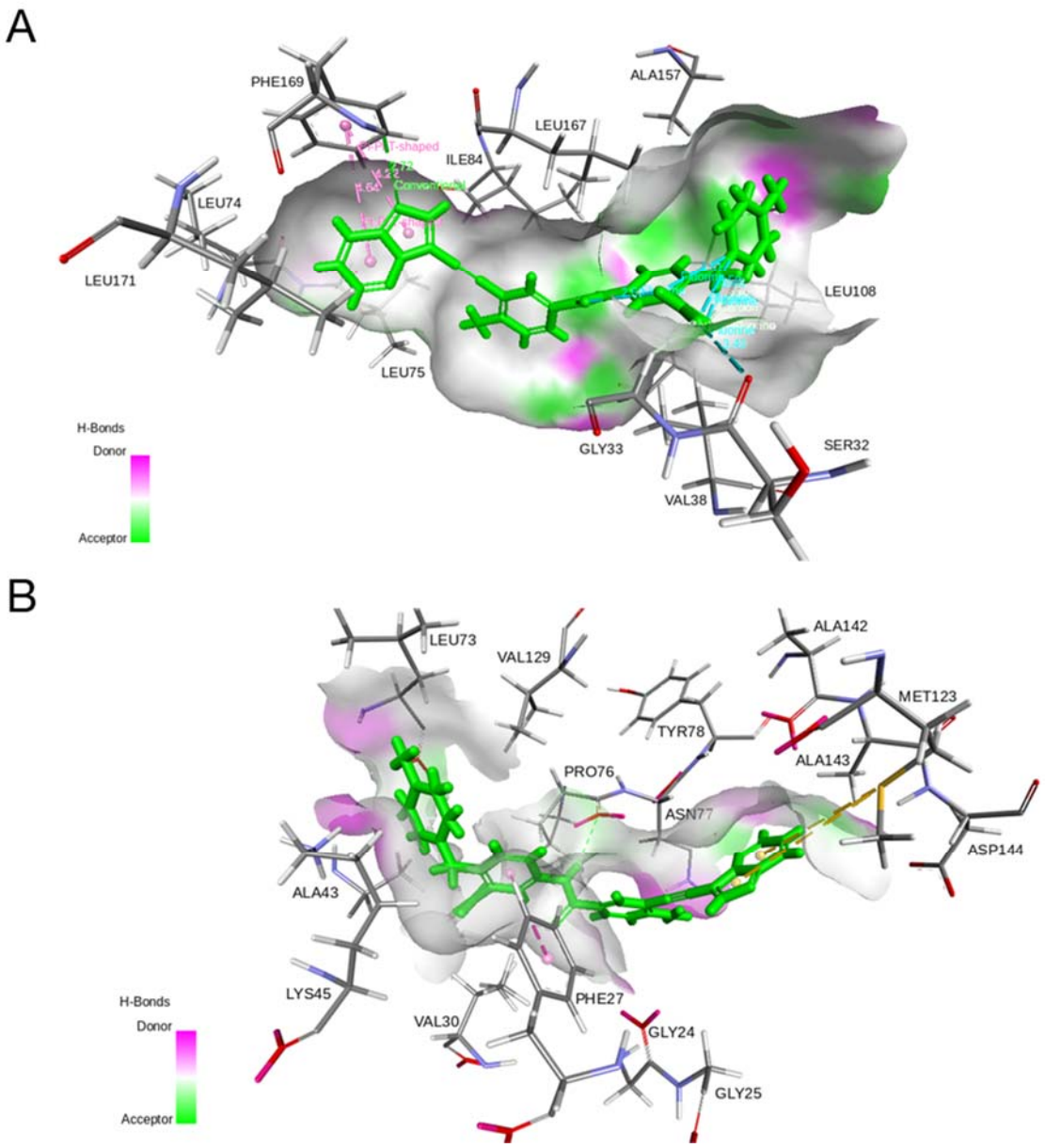


Figure 6 Docking of ponatinib into MAPP14 and ZAK (A) Molecular docking of ponatinib into the active site of MAPK14. (B) same as (A) but for ZAK. In both cases good interactions could be observed that indicate effective binding. Results were kindly provided by Björn Gohlke from the Charité Berlin.

The ChemScores of 41.68 for ponatinib-MAPK14 and 41.75 for ponatinib-ZAK interactions indicated effective binding. The data from the competition assay and the molecular docking was sufficiently encouraging to perform an activity based assay for confirmation. The suspected targets MAPK14 (Figure 8A) and ZAK (Figure 8B) showed clear activity inhibition with rising ponatinib concentration. The IC_{50} values ranged from 27 nM for MAPK14 and down to <1 nM for ZAK. The activity data demonstrated that ponatinib was as potent for ZAK as for its currently accepted target ABL1 (Table 1), underlining the significance of this newly identified target.

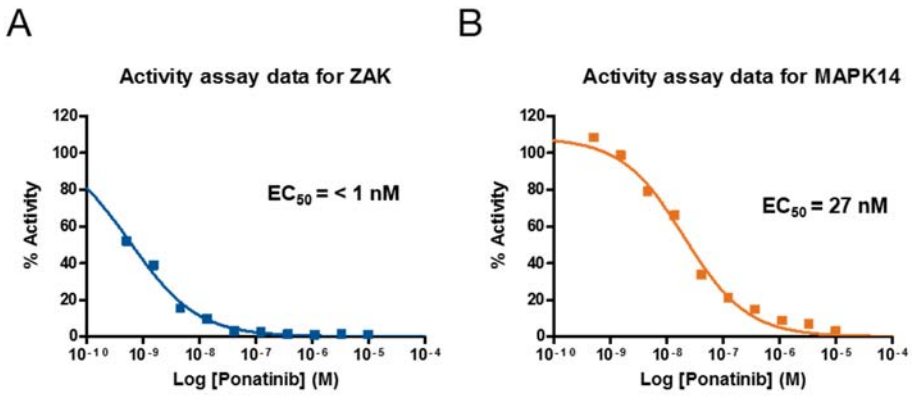


Figure 7 Activity assay data (A) Activity assay data showing the inhibition of the ZAK activity by ponatinib (B) same as (A) but for MAPK14. In both cases a clear inhibition of the kinase activity of both proteins can be seen. The half maximal effective concentration (EC50) for ZAK was even lower than for the primary target BCR-Abl (data not shown). Experiments were performed by Reaction Biology Corporation according to Anastassiadis (2011)[20].

Several studies have already proposed an interesting role of ZAK in the JNK and p38 MAPK pathway in cancer[46] and inflammation[33, 42]. In a publication from 2013, Wong and co-workers[33] state that, among nilotinib and sorafenib, ponatinib was able to suppress the inflammatory response in bone marrow-derived macrophages (BMDM) after treatment with doxorubicin. This resulted in a significant drop in the detection of inflammatory cytokines such as IL-1ß or CXCL1. It could be shown that the JNK mediated pathway as well as the p38 pathway were affected by ponatinib. The authors argued that the main cause for this observation was caused due to the inhibition of ZAK a MAP3K kinase, which they had proven in a previous publication to be involved in the inflammatory response after treatment with doxorubicin[47].

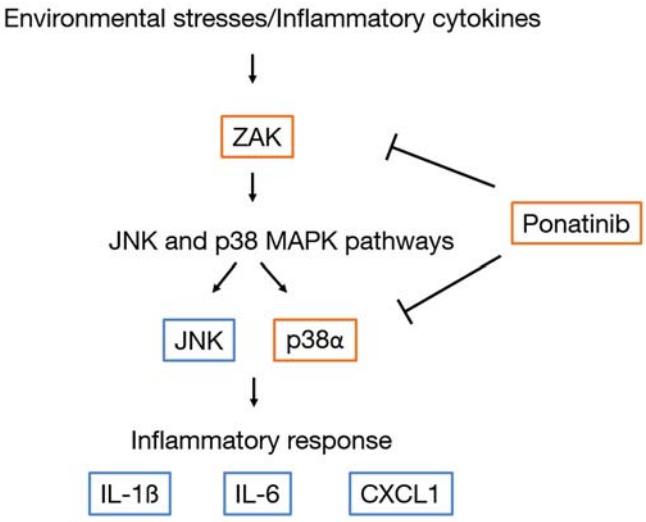


Figure 8 Postulated inhibitory effect on the inflammation processes by ponatinib via targeting MAPK14 and ZAK inline with the biological results observed by Wong et al. (2013)[33].

In the light of our findings it can be stated that the effect observed by Wong et al. originates from the direct inhibition of ZAK as well as from the inhibition of the downstream kinase MAPK14 (p38α) by ponatinib (Fig. 8).

Nevertheless, it could also be argued that the inhibition of the kinases MAP4K1 and MAP4K2 may provide additional effects, since a correlation to the JNK pathway of these kinases has also been reported[48, 49].

The p38 MAPK pathway is involved in many cellular processes affecting proliferation, differentiation and migration in cancer or inflammation[32]. This makes the pathway an interesting target for various therapeutic approaches, however the effect of inhibiting p38 signalling can vary drastically depending on e.g. cell type and applicative context[32]. The first generation of MAPK14 inhibitors showed promising results in the *in vitro* studies but were withdrawn due to liver toxicity[50, 51]. From the over 200 inhibitors in clinical trials, 16 are developed against MAPK14 underlining the importance and the constant interest in targeting these kinases in cancer therapy. From an in-house screen of 150 clinical kinase inhibitors, it can be seen (Fig. 9) that 13 drugs that have not been developed as MAPK14 inhibitors show an affinity to MAPK14 with 9 having a K_D of < 1 µM.

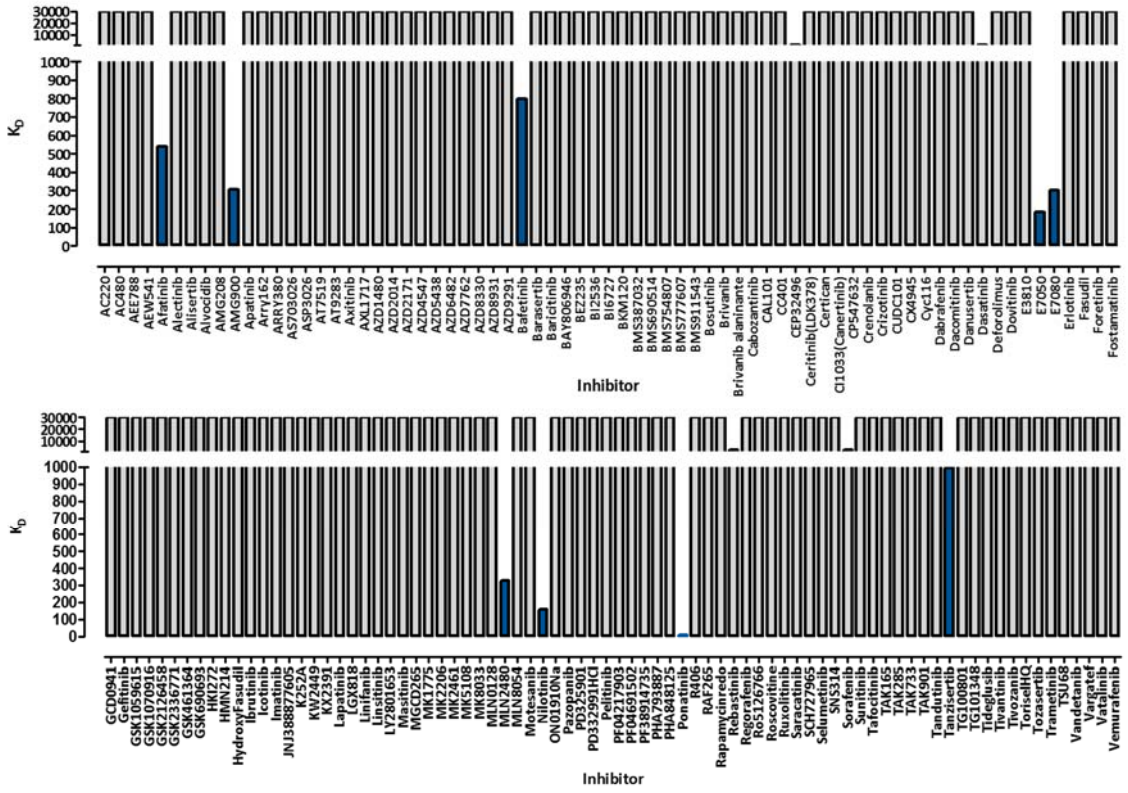


Figure 9 MAPK14 vs. 150 kinase inhibitors 150 clinical kinase inhibitors were screened against MAPK14. Of these, 13 drugs show an affinity to MAPK14, 9 of which have a K_D of < 1 µM, shown by the blue bars.

Nevertheless, targeting only the p38 operated branch in the context of inflammatory response can be compensated by the cells through up regulation of the JNK pathway [52], making drugs that are able to inhibit different or the same pathways on several levels a valuable good in the clinics. In the case of ponatinib (Fig. 10) the aforementioned[33] observed reduction of inflammatory cytokines (e.g. IL-1ß) is caused by inhibition the inhibition of the upstream kinase ZAK of the JNK and p38 pathway as well by the inhibition of the MAPK14 (p38α).

Additionally, recent studies of kinase inhibitors could link further biological effects directly to the inhibition of the mixed lineage kinase ZAK. For the BRAF inhibitor sorafenib, it could be determined that the inhibition of the UV-induced apoptosis in cells is directly mediated by ZAK and independent of the ERK pathway[53]. Interestingly, this study also revealed that the effect is only carried out by the inhibition of ZAK and not supported by MAPK14. Another BRAF inhibitor, vemurafenib, revealed *in vitro* and *in vivo* an apoptosis inhibiting effect that could clearly be linked to its off-target kinase ZAK[54]. Currently there is no specific ZAK inhibitor in clinical development, However, an in-house screening of 150 inhibitors by Kinobeads demonstrates (Fig. 10) that 21 show an inhibitory effect including the previously mentioned nilotinib[19, 55], sorafenib[19, 56, 57] and vemurafenib. From these 21

drugs, 14 inhibitors have a potency of less than 1 µM, making ZAK an interesting and common off-target in kinase inhibitor research that may be considered during the development of inhibitors targeting tyrosine kinases like BRAF and BCR-ABL.

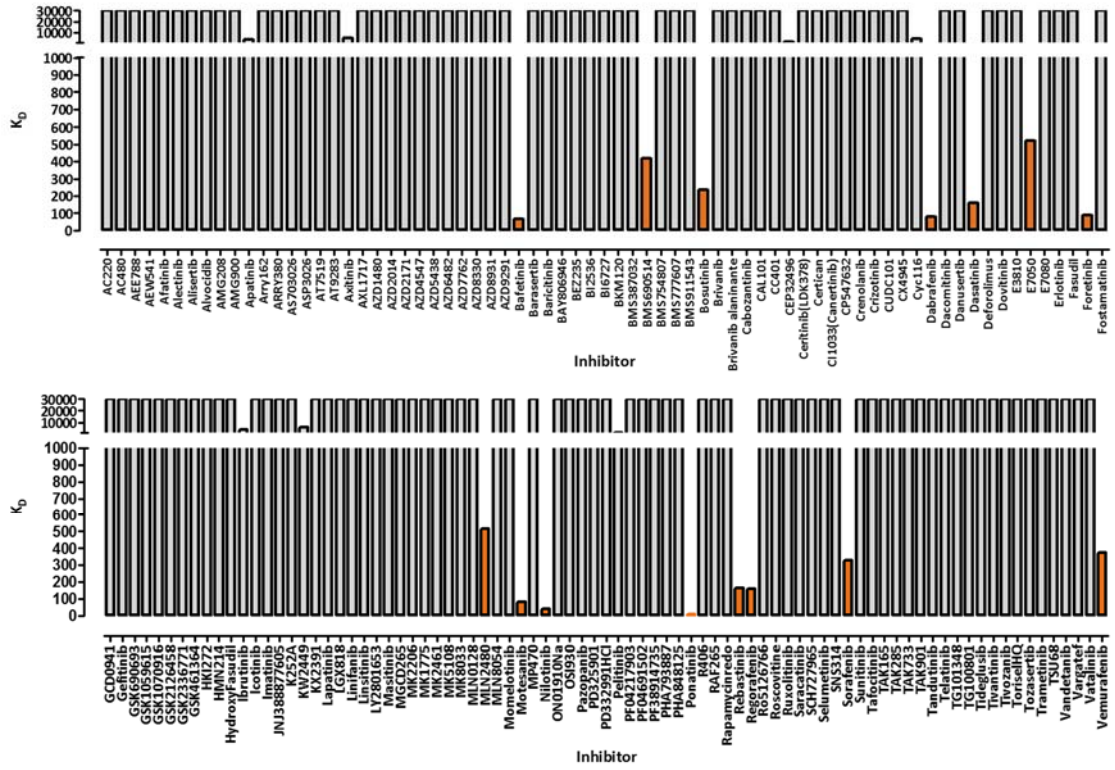


Figure 10 ZAK vs. 150 kinase inhibitors 150 clinical kinase inhibitors were screened against ZAK. Of these, 21 drugs show an affinity to ZAK, 14 of which have a K_D of < 1 µM, shown by the orange bars.

Furthermore, it is noteworthy that a significant proportion (Table 1) of the kinases targeted by ponatinib are reported to be involved in the immune response. Some of them have already been reported (e.g. LYN), but our data suggest that e.g. the activity of IRAK1 and RIPK2 are also affected by ponatinib. IRAK1 is a key regulator for the initiation of the innate immune response, being involved in the Toll-like receptor and IL-1R signalling pathways[58]. This may lead to the establishments of new fields of application for this drug in immune and inflammatory diseases or makes it interesting for combinational treatments with other drugs.

A recent phase III study involving ponatinib was terminated due to the occurrence of severe cardio-vascular effects. So far the reasons for the side-effects seem to be unclear although certain kinases are known to be associated with cardiotoxicity[59] among them several known targets of ponatinib (Table 1). MAPK14, for example, has been reported to be responsible for the death MAPK14 knock-out mice in midgestation caused by a malformed myocardium, however, this effect could be related to general disturbed

placental development and not cardiotoxicity[60]. A number of growth factor receptors regulating tumour cell proliferation/survival as well as tumour angiogenesis including vascular endothelial growth factor receptors (VEGFRs)[61-63], platelet-derived growth factor receptors (PDGFRα and PDGFRß), Kit, FLT3, FMS, and RET[64-66]. This candidates have previously been suspected to cause the cardiotoxicity of sunitinib[67], however, a kinase selectivity study of sunitinib revealed that off-target inhibition of PRKAB1 is the underlying reason for the observed effects[68]. A recent study excluded a ponatinib-based effect on platelet function as the root cause of the acute ischemic cardiovascular events in patients treated with the drug, however, the mechanism of this effect could not be defined[69]. The here presented data may not provide direct insight into this reported toxic side-effects of ponatinib, since maybe the biological link has just not been established yet, but may help understanding the underlying biological processes.

CONCLUDING REMARKS

Here a complementary selectivity analysis of the CML drug ponatinib was presented. This enabled us to identify a set of new high potency kinase targets. Molecular docking and activity based assays confirmed the kinases p38 and ZAK as off-targets of ponatinib, which readily explains the anti-inflammatory effects reported previously[33]. Further, this study underlines the necessity to screen kinase inhibitors against the maximum number of kinases to reveal their full spectrum of targets. Data from 150 kinase inhibitors show that ZAK is a common off-target of kinase inhibitors that are currently in clinical development.

ACKNOWLEDGMENTS

Susan Klaeger Stephanie Heinzlmeir, Huichao Quiao, Andrea Hubauer and Li-Hua Li for helping with the sample preparation and Björn Gohlke for assistance with the molecular docking experiments. Dr. Hannes Hahne and Barbara Ferreira da Silva for discussions.

ABBREVIATIONS

CML	Chronic myeloid leukemia
DMSO	Dimethyl sulfoxide
EC50	Half maximal effective concentration
FDA	Food and Drug Administration
IC50	Half maximal inhibitory concentration
Ki	Kinase inhibitor
Kd	Dissociation constant
LC-MS/MS	Liquid-chromatography coupled to tandem mass spectrometry
LFQ	Label free quantification
MoA	Mode of action
MQ	MaxQuant

REFERENCES

1. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-1934 (2002).
2. Cohen, P. Protein kinases--the major drug targets of the twenty-first century? *Nature reviews. Drug discovery* **1**, 309-315 (2002).
3. Fedorov, O., Muller, S. & Knapp, S. The (un)targeted cancer kinome. *Nat Chem Biol* **6**, 166-169 (2010).
4. Page, T.H., Smolinska, M., Gillespie, J., Urbaniak, A.M. & Foxwell, B.M. Tyrosine kinases and inflammatory signalling. *Current molecular medicine* **9**, 69-85 (2009).
5. Knapp, S. et al. A public-private partnership to unlock the untargeted kinome. *Nat Chem Biol* **9**, 3-6 (2013).
6. Cohen, P. & Alessi, D.R. Kinase drug discovery--what's next in the field? *ACS Chem Biol* **8**, 96-104 (2013).
7. O'Hare, T. et al. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell* **16**, 401-412 (2009).
8. FDA, Vol. 2014 (
9. ARIAD in ARIAD Pharmaceuticals, Inc (2015).
10. Prasad, V. & Mailankody, S. The accelerated approval of oncologic drugs: lessons from ponatinib. *JAMA : the journal of the American Medical Association* **311**, 353-354 (2014).
11. Mayer, K., Gielen, G.H., Willinek, W., Muller, M.C. & Wolf, D. Fatal progressive cerebral ischemia in CML under third-line treatment with ponatinib. *Leukemia* **28**, 976-977 (2014).
12. Ariad suspends ponatinib sales. *Cancer discovery* **4**, 6-7 (2014).
13. Small, S. Marketing and commercial distribution of ponatinib suspended following FDA request. *Clinical advances in hematology & oncology : H&O* **11**, 808-809 (2013).
14. Bantscheff, M. & Drewes, G. Chemoproteomic approaches to drug target identification and drug profiling. *Bioorg Med Chem* **20**, 1973-1978 (2012).
15. Kruse, U. et al. Chemoproteomics-based kinome profiling and target deconvolution of clinical multi-kinase inhibitors in primary chronic lymphocytic leukemia cells. *Leukemia* **25**, 89-100 (2011).
16. Bantscheff, M. et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* **25**, 1035-1044 (2007).
17. Schirle, M., Bantscheff, M. & Kuster, B. Mass spectrometry-based proteomics in preclinical drug discovery. *Chem Biol* **19**, 72-84 (2012).
18. Davis, M.I. et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1046-1051 (2011).
19. Karaman, M.W. et al. A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol* **26**, 127-132 (2008).
20. Anastassiadis, T., Deacon, S.W., Devarajan, K., Ma, H. & Peterson, J.R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1039-1045 (2011).
21. Zinn, N., Hopf, C., Drewes, G. & Bantscheff, M. Mass spectrometry approaches to monitor protein-drug interactions. *Methods* **57**, 430-440 (2012).
22. Medard, G. et al. Optimized Chemical Proteomics Assay for Kinase Inhibitor Profiling. *J Proteome Res* (2015).
23. Bantscheff, M. & Kuster, B. Quantitative mass spectrometry in proteomics. *Anal Bioanal Chem* **404**, 937-938 (2012).

24. Godl, K. et al. An efficient proteomics method to identify the cellular targets of protein kinase inhibitors. *Proc Natl Acad Sci U S A* **100**, 15434-15439 (2003).
25. Graves, P.R. et al. Discovery of novel targets of quinoline drugs in the human purine binding proteome. *Molecular pharmacology* **62**, 1364-1372 (2002).
26. Patricelli, M.P. et al. Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **46**, 350-358 (2007).
27. Liu, Y., Patricelli, M.P. & Cravatt, B.F. Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci U S A* **96**, 14694-14699 (1999).
28. Lemeer, S., Zorgiebel, C., Ruprecht, B., Kohl, K. & Kuster, B. Comparing immobilized kinase inhibitors and covalent ATP probes for proteomic profiling of kinase expression and drug selectivity. *J Proteome Res* **12**, 1723-1731 (2013).
29. Rix, U. et al. Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. *Blood* **110**, 4055-4063 (2007).
30. Bantscheff, M., Scholten, A. & Heck, A.J.R. Revealing promiscuous drug "target interactions by chemical proteomics. *Drug Discovery Today* **14**, 1021-1029 (2009).
31. Narasimhan, N.I., Dorer, D.J., Niland, K., Haluska, F. & Sonnichsen, D. Effects of food on the pharmacokinetics of ponatinib in healthy subjects. *Journal of clinical pharmacy and therapeutics* **38**, 440-444 (2013).
32. Wagner, E.F. & Nebreda, A.R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat Rev Cancer* **9**, 537-549 (2009).
33. Wong, J. et al. Small molecule kinase inhibitors block the ZAK-dependent inflammatory effects of doxorubicin. *Cancer Biol Ther* **14**, 56-63 (2013).
34. Ku, X., Heinzlmeir, S., Liu, X., Medard, G. & Kuster, B. A new chemical probe for quantitative proteomic profiling of fibroblast growth factor receptor and its inhibitors. *J Proteomics* **96**, 44-55 (2014).
35. Ku, X., Heinzlmeir, S., Helm, D., Medard, G. & Kuster, B. New Affinity Probe Targeting VEGF Receptors for Kinase Inhibitor Selectivity Profiling by Chemical Proteomics. *J Proteome Res* (2014).
36. Daub, H. et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol Cell* **31**, 438-448 (2008).
37. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850-858 (1996).
38. Altschul, S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
39. Berman, H.M., Kleywegt, G.J., Nakamura, H. & Markley, J.L. The Protein Data Bank archive as an open data resource. *Journal of computer-aided molecular design* **28**, 1009-1014 (2014).
40. Hudkins, R.L. et al. Mixed-lineage kinase 1 and mixed-lineage kinase 3 subtype-selective dihydronaphthyl[3,4-a]pyrrolo[3,4-c]carbazole-5-ones: optimization, mixed-lineage kinase 1 crystallography, and oral in vivo activity in 1-methyl-4-phenyltetrahydropyridine models. *J Med Chem* **51**, 5680-5689 (2008).
41. Magrane, M. & Consortium, U. UniProt Knowledgebase: a hub of integrated protein data. *Database : the journal of biological databases and curation* **2011**, bar009 (2011).
42. Wang, C., Weerapana, E., Blewett, M.M. & Cravatt, B.F. A chemoproteomic platform to quantitatively map targets of lipid-derived electrophiles. *Nat Methods* **11**, 79-85 (2014).
43. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
44. Jones, G., Willett, P., Glen, R.C., Leach, A.R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267**, 727-748 (1997).

45. Ritz, C. Toward a unified approach to dose-response modeling in ecotoxicology. *Environ Toxicol Chem* **29**, 220-229 (2010).
46. Yang, J.J. et al. ZAK inhibits human lung cancer cell growth via ERK and JNK activation in an AP-1-dependent manner. *Cancer Sci* **101**, 1374-1381 (2010).
47. Sauter, K.A.D., Magun, E.A., Iordanov, M.S. & Magun, B.E. ZAK is required for doxorubicin, a novel ribotoxic stressor, to induce SAPK activation and apoptosis in HaCaT cells. *Cancer Biology & Therapy* **10**, 258-266 (2010).
48. Hu, M.C., Qiu, W.R., Wang, X., Meyer, C.F. & Tan, T.H. Human HPK1, a novel human hematopoietic progenitor kinase that activates the JNK/SAPK kinase cascade. *Genes & development* **10**, 2251-2264 (1996).
49. Yuasa, T., Ohno, S., Kehrl, J.H. & Kyriakis, J.M. Tumor necrosis factor signaling to stress-activated protein kinase (SAPK)/Jun NH2-terminal kinase (JNK) and p38. Germinal center kinase couples TRAF2 to mitogen-activated protein kinase/ERK kinase kinase 1 and SAPK while receptor interacting protein associates with a mitogen-activated protein kinase kinase kinase upstream of MKK6 and p38. *J Biol Chem* **273**, 22681-22692 (1998).
50. Kumar, S., Boehm, J. & Lee, J.C. p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases. *Nature reviews. Drug discovery* **2**, 717-726 (2003).
51. Kim, M.S., Lee, E.J., Kim, H.R. & Moon, A. p38 kinase is a key signaling molecule for H-Ras-induced cell motility and invasive phenotype in human breast epithelial cells. *Cancer Res* **63**, 5454-5461 (2003).
52. Cheung, P.C., Campbell, D.G., Nebreda, A.R. & Cohen, P. Feedback control of the protein kinase TAK1 by SAPK2a/p38alpha. *Embo J* **22**, 5793-5805 (2003).
53. Vin, H. et al. Sorafenib suppresses JNK-dependent apoptosis through inhibition of ZAK. *Mol Cancer Ther* **13**, 221-229 (2014).
54. Vin, H. et al. BRAF inhibitors suppress apoptosis through off-target inhibition of JNK signaling. *eLife* **2**, e00969 (2013).
55. Weisberg, E. et al. AMN107 (nilotinib): a novel and selective inhibitor of BCR-ABL. *Br J Cancer* **94**, 1765-1769 (2006).
56. Iyer, R., Fetterly, G., Lugade, A. & Thanavala, Y. Sorafenib: a clinical and pharmacologic review. *Expert opinion on pharmacotherapy* **11**, 1943-1955 (2010).
57. Wilhelm, S.M. et al. BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res* **64**, 7099-7109 (2004).
58. Uematsu, S. et al. Interleukin-1 receptor-associated kinase-1 plays an essential role for Toll-like receptor (TLR)7- and TLR9-mediated interferon-Î± induction. *The Journal of Experimental Medicine* **201**, 915-923 (2005).
59. Force, T. & Kolaja, K.L. Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes. *Nature reviews. Drug discovery* **10**, 111-126 (2011).
60. Adams, R.H. et al. Essential role of p38alpha MAP kinase in placental but not embryonic cardiovascular development. *Mol Cell* **6**, 109-116 (2000).
61. Demetri, G.D. et al. Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. *Lancet* **368**, 1329-1338 (2006).
62. Motzer, R.J. et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med* **356**, 115-124 (2007).
63. Faivre, S., Demetri, G., Sargent, W. & Raymond, E. Molecular basis for sunitinib efficacy and future clinical development. *Nature reviews. Drug discovery* **6**, 734-745 (2007).
64. Force, T., Krause, D.S. & Van Etten, R.A. Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition. *Nat Rev Cancer* **7**, 332-344 (2007).
65. Verheul, H.M. & Pinedo, H.M. Possible molecular mechanisms involved in the toxicity of angiogenesis inhibition. *Nat Rev Cancer* **7**, 475-485 (2007).

66. Chen, M.H., Kerkela, R. & Force, T. Mechanisms of cardiac dysfunction associated with tyrosine kinase inhibitor cancer therapeutics. *Circulation* **118**, 84-95 (2008).
67. Chu, T.F. et al. Cardiotoxicity associated with tyrosine kinase inhibitor sunitinib. *Lancet* **370**, 2011-2019 (2007).
68. Kerkela, R. et al. Sunitinib-induced cardiotoxicity is mediated by off-target inhibition of AMP-activated protein kinase. *Clin Transl Sci* **2**, 15-25 (2009).
69. Loren, C.P. et al. The BCR-ABL inhibitor ponatinib inhibits platelet immunoreceptor tyrosine-based activation motif (ITAM) signaling, platelet activation and aggregate formation under shear. *Thrombosis research* **135**, 155-160 (2015).

CHAPTER V

GENERAL DISCUSSION

GENERAL DISCUSSION

Before the draft of the human proteome based on mass spectrometry (MS) data was published[1-3], MS had already developed as the standard technology for high-throughput proteomic studies. Over the past decade, a variety of methods have emerged for the analysis of simple and complex (sub) proteomes using MS, allowing proteomics studies to exhaustively detect peptides/proteins in a given biological sample. However, this was only made possible due to the availability of high-quality genomic data, technological advances in the fields of microfluidics and MS. Improved sensitivity of instrumentation facilitated the analysis of even more complex types of samples[4-7]. The technologies used in proteomics studies (particularly LC-ESI-MS/MS) are well suited as a possible experimental strategy for pre-clinical identification and validation of novel drug targets. In this regard, application of chemically probes to enrich protein targets from native biological samples has been introduced under the term chemical proteomics[8-10]. Thus, advances in chemical proteomics as well as in quantitative MS, introducing high-throughput workflows for drug target deconvolution, disease pathway analysis and the understanding of cellular protein dynamics, have recently improved pre-clinical target validation[8,11-16]. Although MS technology is constantly evolving the number of eluting features, i.e. possible peptides, during an experiment[17] still outranges the capacity of even the fastest available instruments. In this context, the primary objective of this thesis was the evaluation and development of new mass spectrometry based technologies for usage in high-throughput chemical proteomics assays.

A substantial challenge in probing biological systems remains the identification of targets being perturbed and causing favorable phenotypic response. Critical to the success of discovery programs is the challenging aspect of deconvoluting targets and pathways elucidating the mechanism of action of chemical compounds. Towards this end, a Synapt G2Si Q-TOF mass spectrometer in combination with a data-independent acquisition (DIA) strategy was chosen for its possible use in a large scale high-throughput chemical proteomics assay. Combining speed and sensitivity, this instrument, would, in principle, be suitable for high-throughput chemical proteomics and drug selectivity screening assays. The DIA method HDMS^e, as described in Chapter II, records all precursors and all corresponding fragments during an analysis, emulating a data-dependent acquisition (DDA) set-up with infinite speed. Although conceptually attractive, in practise, the presented HDMS^e method on the Q-TOF instrument proved to be less performant in comparison to state-of-the-art

DDA experiments using a LTQ-Orbitrap Elite. This discrepancy in performance derived from several limitations, in both data processing and hardware performance. Overall, the HDMS^e approach did not offer the performance required for its application in a high-throughput chemical proteomics screening.

Regardless of the application, the field of proteomics will always rely on the improvement and development of technology that enable faster and more sensitive instruments as well as new biochemical assays. In respect to the requirements in shot-gun proteomics, the ideal mass spectrometer would be an instrument providing the following features: High resolution (e.g. > 50,000 for TMT 10plex[18]), mass accuracy < 1 ppm[19], a speed of >> 25 Hz[17] and a sensitivity high enough to see and successfully identify all eluting peptides generated in a proteomics experiment. For a full proteomic study, sensitivity spanning 5-6 orders of magnitude seems to be sufficient[1], whereas for post-translational modification analysis (PTM), an additional 1 to 2 orders would be needed[20]. The main limitation observed on the orthogonal acceleration TOF (oaTOF) instrument, in both DIA and DDA mode, was the sensitivity which was reduced by 100x when compared to the LTQ-Orbitrap Elite instrument. Both instruments were, at the time of comparison, the most sensitive machines offered by their respective vendors for bottom-up proteomics. The sensitivity of oaTOF instrument is a long-standing issue and originates from the comparatively low duty-cycle (approx. 10%), meaning that only every 10th ion entering the instrument reaches the detector. Thus, in chapter III, the performance issues concerning the oaTOF instrument (mainly sensitivity) was addressed by making use of the incorporated travelling wave ion mobility separation (TWIMS) device. The duty cycle enhancement using TWIMS has been previously described[21], however, the utility of this method was limited to a small *m/z* range. In the light of this problem, it was aimed to expand the utility of such a method to the full *m/z* range used for a proteomic experiment and implement it into a standard DDA acquisition approach. This resulted in a considerable enhancement of system performance, in terms of both speed and sensitivity. The developed method, called HD-DDA was, thereafter, made available on all commercial Synapt G2Si instruments from Waters Corporation and through alterations to the acquisition software an average speed of 10 Hz can now be achieved with the newest iteration.

As described in chapter II, the HDMS^e method was initially evaluated as part of a large-scale high throughput screening of 200 kinase inhibitors using a new generation of the kinobeads technology[16]. The goal of this screening was to analyse 200 kinase inhibitors currently in clinical development, including the already 30 FDA approved drugs. Among these drugs is the BCR-ABL inhibitor ponatinib (Chapter IV), approved since 2012[22] for the treatment of

chronic myeloid leukemia (CML). Ponatinib's potency, in contrast to other small molecule BCR-ABL inhibitors, is not affected by the gatekeeper mutation T315I in BCR-ABL[23], which is the primary mechanism of resistance against other BCR-ABL inhibitors. Ponatinib was of particular interest in this study since it was withdrawn from the market due to the occurrence of severe side effects[24-26]. The here presented selectivity profiling did not lead to a direct explanation of the occurring toxicity, nevertheless, it was possible to determine a set of new highly potent targets, including ZAK and MAPK14, of the drug. This helped to rationalize the previously observed inhibitory effect on inflammation processes by ponatinib[27]. These data allow a deeper understanding of the mode of action of ponatinib and perhaps, in the near future, assist in elucidating the processes leading to its toxicity and therefore lead to an improved therapeutic application. This large scale chemical proteomics study further will, besides the target space deconvolution of the included drugs, allow insight into the currently drug-able kinome[28] and help to find probes for biological research as well as to help medicinal chemists to focus on so far untargeted kinases.

The greater progress made in the field of MS, during the time period of this thesis, was regarding DDA method on hybrid Orbitrap instruments. At the start of the thesis, the fastest available instrument was the LTQ-Orbitrap Elite with approx. 5 Hz (HCD)[29]. Today, the state-of-the-art instruments are the Q-Exactive HF and the Fusion with >25 Hz acquisition speed[4, 30]. This reveals tremendous increase in system performance, however, yet not being enough for real comprehensiveness in proteomics studies. Although the HDMS^e approach has the discussed limitations, there is no doubt in the possible advantages it may have over a standard DDA experiment. Other DIA approaches like SWATH[31] are currently already marketed towards clinical research applications such as creating a digital map of patient's samples[32]. Although the SWATH approach is basing peptide and protein identification on previously acquired DDA data, in the light of this study, the question remains: if this step towards the clinics does not come too early? Especially if the underlying hardware issues are considered, which are similar to that of the here used Synapt G2Si instrument. Nevertheless, the combination of spectral libraries with DIA seems to be the most promising approach to circumvent the issues arising during the data processing and is one strategy that may enhance HDMS^e.

In respect of the aforementioned developments for DIA and DDA approaches, the prediction might be made that ultra-fast and highly sensitive DDA approaches seem to be a more promising prospect for reaching the necessary criteria for comprehensiveness in proteomics. Aside from the conceptual advantages, the DIA approaches lack state-of-the art performance.

Finally, developments such as those introduced in this thesis will lead to mass spectrometers enabling high throughput proteomic studies with before unreached level of sampling depth for various applications to well beyond the field of chemical proteomics.

ABBREVIATIONS

DDA	Data dependent acquisition
DIA	Data independent acquisition
FDA	Food and Drug Administration
HD-DDA	High definition data dependent acquisition (ion mobility enhanced DDA)
HDMS^e	Data independent acquisition using ion mobility separation (Waters, Corporation)
LC-ESI-MS/MS	Liquid-chromatography coupled to tandem mass spectrometry via an electrospray ionisation source
LTQ	Linear ion trap quadrupole
MS	Mass spectrometer
TMT	Tandem mass tag
TOF	Time-of-flight mass analyser
TWIMS	Travelling wave ion mobility separation
oaTOF	orthogonal acceleration TOF
Q-TOF	Quadrupole TOF

REFERENCES

1. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587 (2014).
2. Kim, M.S. et al. A draft map of the human proteome. *Nature* **509**, 575-581 (2014).
3. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
4. Hebert, A.S. et al. The One Hour Yeast Proteome. *Mol Cell Proteomics* (2013).
5. Hahne, H. et al. DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Methods* **10**, 989-991 (2013).
6. Pirmoradian, M. et al. Rapid and Deep Human Proteome Analysis by Single-dimension Shotgun Proteomics. *Mol Cell Proteomics* **12**, 3330-3338 (2013).
7. Bantscheff, M. & Kuster, B. Quantitative mass spectrometry in proteomics. *Anal Bioanal Chem* **404**, 937-938 (2012).
8. Bantscheff, M. & Drewes, G. Chemoproteomic approaches to drug target identification and drug profiling. *Bioorg Med Chem* **20**, 1973-1978 (2012).
9. Kruse, U., Bantscheff, M., Drewes, G. & Hopf, C. Chemical and pathway proteomics: powerful tools for oncology drug discovery and personalized health care. *Mol Cell Proteomics* **7**, 1887-1901 (2008).
10. Oda, Y. et al. Quantitative chemical proteomics for identifying candidate drug targets. *Anal Chem* **75**, 2159-2165 (2003).
11. Anastassiadis, T., Deacon, S.W., Devarajan, K., Ma, H. & Peterson, J.R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1039-1045 (2011).
12. Karaman, M.W. et al. A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol* **26**, 127-132 (2008).
13. Oppermann, F.S. et al. Large-scale proteomics analysis of the human kinome. *Mol Cell Proteomics* **8**, 1751-1764 (2009).
14. Daub, H. et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol Cell* **31**, 438-448 (2008).
15. Daub, H. Characterisation of kinase-selective inhibitors by chemical proteomics. *Biochim Biophys Acta* **1754**, 183-190 (2005).
16. Medard, G. et al. Optimized Chemical Proteomics Assay for Kinase Inhibitor Profiling. *J Proteome Res* (2015).
17. 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. *Journal of Proteome Research* **10**, 1785-1793 (2011).
18. Werner, T. et al. Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal Chem* **86**, 3594-3601 (2014).
19. Zubarev, R.A., Håkansson, P. & Sundqvist, B. Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements. *Analytical Chemistry* **68**, 4060-4063 (1996).
20. Ruprecht, B. et al. Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. *Mol Cell Proteomics* **14**, 205-215 (2015).
21. Giles, K. et al. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* **18**, 2401-2414 (2004).
22. FDA, Vol. 2014 (
23. O'Hare, T. et al. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell* **16**, 401-412 (2009).

24. Deininger, M. The current status of ponatinib in the treatment of chronic myeloid leukemia. *Clinical advances in hematology & oncology : H&O* **12**, 329-331 (2014).
25. Nazha, A., Romo, C.G., Kantarjian, H. & Cortes, J. The clinical impact of ponatinib on the risk of bleeding in patients with chronic myeloid leukemia. *Haematologica* **98**, e131 (2013).
26. Dalzell, M.D. Ponatinib pulled off market over safety issues. *Managed care* **22**, 42-43 (2013).
27. Wong, J. et al. Small molecule kinase inhibitors block the ZAK-dependent inflammatory effects of doxorubicin. *Cancer Biol Ther* **14**, 56-63 (2013).
28. Knapp, S. et al. A public-private partnership to unlock the untargeted kinome. *Nat Chem Biol* **9**, 3-6 (2013).
29. Pachl, F., Ruprecht, B., Lemeer, S. & Kuster, B. Characterization of a high field Orbitrap mass spectrometer for proteome analysis. *Proteomics* **13**, 2552-2562 (2013).
30. Scheltema, R.A. 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* **13**, 3698-3708 (2014).
31. Gillet, L.C. 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* **11**, O111 016717 (2012).
32. Guo, T. et al. Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nature medicine* (2015).

LIST OF PUBLICATIONS

Helm, D., Vissers, J. P., Hughes, C. J., Hahne, H., Ruprecht, B., Pachl, F., Grzyb, A., Richardson, K., Wildgoose, J., Maier, S. K., Marx, H., Wilhelm, M., Becher, I., Lemeer, S., Bantscheff, M., Langridge, J. I., and Kuster, B. (2014) Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol Cell Proteomics* 13, 3709-3715

Helm, D., Klaeger, S.,Gohlke, B., Heinzlmeir, S., Quiao, H., Preissner, R. ,Medard, G., and Kuster, B., (2015) Chemical proteomics identifies Ponatinib as potent MAPK14 and ZAK inhibitr. manuscript in preparation.

Application note: **Helm, D.,** Hughes, C. J., Langridge, J. I., Richardson, K., Grzyb, A., Wildgoose, J., M., and Kuster, B. (2013) High Definition Multiple Reaction Monitoring: The Application of Quadrupole Ion Mobility Time-of-Flight Mass Spectrometry for Targeted Proteomics Studies. HUPO 12th Annual World Congress, Yokohama, Japan.

Medard, G., Pachl, F., Ruprecht, B., Klaeger, S., Heinzlmeir, S., **Helm, D.**, Qiao, H., Ku, X., Wilhelm, M., Kuehne, T., Wu, Z., Dittmann, A., Hopf, C., Kramer, K., and Kuster, B. (2015) Optimized Chemical Proteomics Assay for Kinase Inhibitor Profiling. *J Proteome Res*

Ku, X., Heinzlmeir, S., **Helm, D.,** Medard, G., and Kuster, B. (2014) New Affinity Probe Targeting VEGF Receptors for Kinase Inhibitor Selectivity Profiling by Chemical Proteomics. *J Proteome Res*

Hahne, H., Sobotzki, N., Nyberg, T., **Helm, D.**, Borodkin, V. S., van Aalten, D. M., Agnew, B., and Kuster, B. (2013) Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass spectrometry. *J Proteome Res* 12, 927-936

Hahne, H., Pachl, F., Ruprecht, B., Maier, S. K., Klaeger, S., **Helm, D.**, Medard, G., Wilm, M., Lemeer, S., and Kuster, B. (2013) DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Methods* 10, 989-991

DANKSAGUNG

ACKNOWLEDGMENT

AGRADECIMENTOS

“Hier steh ich nun ich armer Tor! Und bin so klug als wie zuvor. Heisse Magister, heiße (bald) Doktor gar...“ (Faust I, J. W. von Goethe)

Dieses Zitat fast wohl am ehesten mein Gefühl zusammen, das mich nach fast 3,5 Jahren Doktorarbeit überkommt. Ich weiss sicher mehr als zuvor und bin sicher um einiges an Erfahrungen reicher, aber man fühlt es nicht. Es lässt sich nicht greifen, nicht riechen und auch nicht schmecken. Eine Möglichkeit jedoch diese Zeitspanne messbar und greifbar zu machen ist diese Doktorabeit.

Den Weg der letzten Jahre bis dieser Arbeit bin ich selbstverständlich nicht alleine gegangen. Viele Menschen aus meiner Familie, meinem Freundes- und Kollegenkreis haben mich in dieser Zeit begleitet und bei all diesen möchte mich an dieser Stelle herzlichst bedanken.

Zu aller erst möchte ich natürlich meinem Doktorvater Bernhard danken, der mir in den letzten Jahren immer mit Rat und Tat zur Seite stand und seinen “doktorväterlichen Pflichten“ nicht nur auf wissenschaftlicher Ebene mit bravur nachkam. Unvergessen ist sein erster Rat zu beginn meiner Arbeit: „*Don’t screw it up!*“

Desweiteren gilt man Dank Prof. Dr. Sieber, der sich bereit erklärt hat als Zweitprüfer diese Arbeit zu begutachten und natürlich auch Prof. Dr. Langosch, der freundlicherweise den Vorstiz meiner Prüfungskommission bekleidete.

Ein besonderes Dankeschön gilt allen Mitgliedern des Lehrstuhls. Sei es im „Maschinenraum“ oder im Bio- bzw. Chemielabor. Hierbei möchte ich besonders unsere Herzstücke Andrea, Michaela, Martina, Gabi, Silvia und natürlich Andi hervorheben die diesen ganzen Laden zusammenhalten. Darüberhinaus möchte ich mich bei den Leuten bedanken die meinen Weg hier am längsten, sei es ebenfalls als PhD students oder später PostDocs begleitet haben: Hannes, Guillaume, Ben, Chen, Steffie, Susan, Heiner, Huichao und Mathias. Ebenfalls danke an meine neugewonnen Zellengenossen und ihr

grosses Verständniss für meinen schwarzen Humor und das gesteigerte Level an Sarkasmus in meiner Umgebung: Charlotte und Peng. Bless them (the nice meaning!)

Further I more than appreciated the time I could spend at Waters Corporation in Manchester. An especially warm thanks goes here to Chris and Hans who taught me everything about the hardware and software I needed to know to get a good start into my project. Not unmentioned should be the on-site people such as Mathias, Harald, David and Max who were always here when something needed to be fixed. Also I want to thank Marc, who gave me great opportunities to speak at several occasions and to present my work.

Ein großes Dankeschön geht auch an meine engsten Freunde Mona, Steve und Sebastian die in sämtlichen Lebenslagen immer fuer mich da waren und – ich weiss – immer fuer mich da sein werden! Genauso wie meine Freunde aus Lauingen ohne die meine Jugend nicht einmal halb so lustig gewesen wäre!

Nun möchte ich meinen Dank an meine Familie richten – egal ob in Deutschland oder Brasilien. Jedoch habe ich das Gefühl, dass egal was ich sage hier nun schreibe es in keinster Weise meine Gefühl ausdrücken kann, aber ich geb mein Bestes: Ohne euch wäre ich nicht hier. In jeglicher Hinsicht. Danke an meine Eltern, die mich den gesamten Weg vom Kindergarten (leider nur teilweise fast) bis zu diesem Punkt begleiten konnten und mir immer den Rücken frei hielten, dass ich mich voll auf mein Studium konzentrieren konnte! *Um agradecimento especial à minha família no Brasil. Obrigada mãe, avós, Bigode e Gordo, pelo apoio e por ter vocês sempre comigo!*

Danke auch an meinen Onkel, der immer ein offenes Ohr für mich hat und immer für mich da ist. Sei es in es, um die letzten Entwicklungen beim FC Bayern zu erörtern oder im persönlichen Bereich. Wir werden immer zusammenhalten.

Quero, agora, agradecer à minha esposa. Bárbara, sou muito afortunado por tê-la em minha vida, pois, sem você, ela não seria a mesma. Não tenho como expressar toda a minha gratidão e todo o amor que sinto por você. Você é tudo para mim. Muito obrigado por estar sempre ao meu lado.

Dominic

APPENDIX

List of phosphopeptides used in Chapter II “Applicability of TWIMS for the separation of positional phospho peptide isomers”

Sequence	IPI acc. No.	Protein	# p sites	Mod. position	neutral mass SH Cys	z = 2+	z = 3+	Plate position	Avg. drift time [bin]	Avg. retention time
DSPGIPPSAGAHQLFR	IPI00329733	RPS6KA1	1	8	1728.7984	865.39921	577.26614	plate 2 / well A 4	69.5	17.18
DSPGIPPSANAHQLFR	IPI00329744	RPS6KA3	1	2	1785.8199	893.90994	596.2733	plate 2 / well B 4	71.2	16.96
ETTTSPKKYYLAEK	IPI00329713	TEC	1	3	1737.8226	869.91129	580.2742	plate 1 / well G 8	72.3	15.06
ETTTSPKKYYLAEK	IPI00329725	TEC	1	4	1737.8226	869.91129	580.2742	plate 1 / well H 8	70.8	14.97
FSDQAGPAIPTSNSYSK	IPI00329666	MARK2	1	15	1848.7931	925.39654	617.26436	plate 1 / well C 9	71.1	15.79
FSDQAGPAIPTSNSYSK	IPI00329678	MARK2	1	16	1848.7931	925.39654	617.26436	plate 1 / well D 9	71.5	16.24
LMTGDTYTAHAGAK	IPI00329488	ABL2	1	7	1515.6428	758.82142	506.21428	plate 1 / well A 3	63.8	14.45
LMTGDTYTAHAGAK	IPI00329719	ABL2	1	6	1515.6428	758.82142	506.21428	plate 1 / well H 2	66.2	15.13
MMSLSQSR	IPI00329681	RIPK2	1	5	1018.3977	510.19884	340.4659	plate 1 / well D12	43.6	15.87
MMSLSQSR	IPI00329693	RIPK2	1	3	1018.3977	510.19884	340.4659	plate 1 / well E12	41.9	15.95
SGGQRHSPLSQR	IPI00329760	TBKBP1	1	1	1388.631	695.31549	463.877	plate 2 / well C 9	53.6	14.02

SGGQRHSPLSQR	IPI00329771	TBKBP1	1	7	1388.631	695.31549	463.877	plate 2 / well D 9	54.3	13.97
SNSTSSMSSGLPEQDR	IPI00329745	ABL2	1	3	1761.6876	881.84382	588.22922	plate 2 / well B 5	66.1	15.37
SNSTSSMSSGLPEQDR	IPI00329756	ABL2	1	4	1761.6876	881.84382	588.22922	plate 2 / well C 5	64.7	15.19
SNSTSSMSSGLPEQDR	IPI00329745	ABL2	1	3	1761.6876	881.84382	588.22922	plate 2 / well B 5	66.1	15.37
SNSTSSMSSGLPEQDR	IPI00329767	ABL2	1	5	1761.6876	881.84382	588.22922	plate 2 / well D 5	66.5	15.36
SNSTSSMSSGLPEQDR	IPI00329745	ABL2	1	3	1761.6876	881.84382	588.22922	plate 2 / well B 5	66.1	15.37
SNSTSSMSSGLPEQDR	IPI00329778	ABL2	1	6	1761.6876	881.84382	588.22922	plate 2 / well E 5	65.7	15.24
SNSTSSMSSGLPEQDR	IPI00329745	ABL2	1	3	1761.6876	881.84382	588.22922	plate 2 / well B 5	66.1	15.37
SNSTSSMSSGLPEQDR	IPI00329788	ABL2	1	8	1761.6876	881.84382	588.22922	plate 2 / well F 5	67.0	15.25
SNSTSSMSSGLPEQDR	IPI00329756	ABL2	1	4	1761.6876	881.84382	588.22922	plate 2 / well C 5	64.7	15.19
SNSTSSMSSGLPEQDR	IPI00329767	ABL2	1	5	1761.6876	881.84382	588.22922	plate 2 / well D 5	66.5	15.36
SNSTSSMSSGLPEQDR	IPI00329756	ABL2	1	4	1761.6876	881.84382	588.22922	plate 2 / well C 5	64.7	15.19
SNSTSSMSSGLPEQDR	IPI00329778	ABL2	1	6	1761.6876	881.84382	588.22922	plate 2 / well E 5	65.7	15.24
SNSTSSMSSGLPEQDR	IPI00329756	ABL2	1	4	1761.6876	881.84382	588.22922	plate 2 / well C 5	64.7	15.19
SNSTSSMSSGLPEQDR	IPI00329788	ABL2	1	8	1761.6876	881.84382	588.22922	plate 2 / well F 5	67.0	15.25
SNSTSSMSSGLPEQDR	IPI00329767	ABL2	1	5	1761.6876	881.84382	588.22922	plate 2 / well D 5	66.5	15.36
SNSTSSMSSGLPEQDR	IPI00329778	ABL2	1	6	1761.6876	881.84382	588.22922	plate 2 / well E 5	65.7	15.24

SNSTSSMSSGLPEQDR	IPI00329767	ABL2	1	5	1761.6876	881.84382	588.22922	plate 2 / well D 5	66.5	15.36
SNSTSSMSSGLPEQDR	IPI00329788	ABL2	1	8	1761.6876	881.84382	588.22922	plate 2 / well F 5	67.0	15.25
SNSTSSMSSGLPEQDR	IPI00329767	ABL2	1	5	1761.6876	881.84382	588.22922	plate 2 / well D 5	66.5	15.36
SNSTSSMSSGLPEQDR	IPI00329778	ABL2	1	6	1761.6876	881.84382	588.22922	plate 2 / well E 5	65.7	15.24
SQSDIFSR	IPI00329794	TESK2	1	1	1018.4121	510.20603	340.47069	plate 2 / well G 1	41.0	16.48
SQSDIFSR	IPI00329804	TESK2	1	3	1018.4121	510.20603	340.47069	plate 2 / well H 1	42.1	16.63
SRNSPLLER	IPI00329731	MARK2	1	4	1150.5495	576.27477	384.51652	plate 2 / well A 2	46.5	14.95
SRNSPLLER	IPI00329742	MARK2	1	1	1150.5495	576.27477	384.51652	plate 2 / well B 2	45.9	15.41
SRTPPSAPSQSR	IPI00329739	SRRM2	1	6	1349.6089	675.80443	450.86962	plate 2 / well A10	51.7	13.96
SRTPPSAPSQSR	IPI00329812	SRRM2	1	1	1349.6089	675.80443	450.86962	plate 2 / well H 9	51.2	14.02
SSSPTQYGLTK	IPI00329650	ZAK	1	2	1247.5435	624.77173	416.84782	plate 1 / well B 5	49.1	15.40
SSSPTQYGLTK	IPI00329709	ZAK	1	3	1247.5435	624.77173	416.84782	plate 1 / well G 4	48.8	15.39
SSSPTQYGLTK	IPI00329650	ZAK	1	2	1247.5435	624.77173	416.84782	plate 1 / well B 5	49.1	15.40
SSSPTQYGLTK	IPI00329721	ZAK	1	5	1247.5435	624.77173	416.84782	plate 1 / well H 4	48.6	15.51
SSSPTQYGLTK	IPI00329709	ZAK	1	3	1247.5435	624.77173	416.84782	plate 1 / well G 4	48.8	15.39
SSSPTQYGLTK	IPI00329721	ZAK	1	5	1247.5435	624.77173	416.84782	plate 1 / well H 4	48.6	15.51

STFHAGQLR	IPI00329753	MARK2	1	2	1095.4862	548.74311	366.16207	plate 2 / well C 2	44.0	15.08
STFHAGQLR	IPI00329764	MARK2	1	1	1095.4862	548.74311	366.16207	plate 2 / well D 2	45.9	15.19
STLVLHDLLK	IPI00329662	BCR	1	2	1217.642	609.82102	406.88068	plate 1 / well C 5	52.7	18.89
STLVLHDLLK	IPI00329674	BCR	1	1	1217.642	609.82102	406.88068	plate 1 / well D 5	50.4	18.71
YMEDSTYYK	IPI00329699	PTK2	1	7	1278.4515	640.22577	427.15052	plate 1 / well F 6	54.7	15.06
YMEDSTYYK	IPI00329711	PTK2	1	8	1278.4515	640.22577	427.15052	plate 1 / well G 6	54.3	15.05
ETTTSPKKYYLAEK	IPI00329701	TEC	1	2	1737.8226	869.91129	580.2742	plate 1 / well F 8	72.3	14.95
ETTTSPKKYYLAEK	IPI00329713	TEC	1	3	1737.8226	869.91129	580.2742	plate 1 / well G 8	72.3	15.06
ETTTSPKKYYLAEK	IPI00329701	TEC	1	2	1737.8226	869.91129	580.2742	plate 1 / well F 8	72.3	14.95
ETTTSPKKYYLAEK	IPI00329725	TEC	1	4	1737.8226	869.91129	580.2742	plate 1 / well H 8	70.8	14.97
ILSDVTHSAVFGVPASK	IPI00329714	AAK1	1	6	1806.8916	904.44582	603.29722	plate 1 / well G 9	72.5	17.71
ILSDVTHSAVFGVPASK	IPI00329726	AAK1	1	8	1806.8916	904.44582	603.29722	plate 1 / well H 9	76.3	17.60
LRSADSENALSVQER	IPI00329791	MAP3K3	1	3	1753.7996	877.89978	585.59985	plate 2 / well F 8	65.5	15.27
LRSADSENALSVQER	IPI00329801	MAP3K3	1	6	1753.7996	877.89978	585.59985	plate 2 / well G 8	66.2	15.20
SSSFREMDGQPER	IPI00329750	ABL1	1	1	1604.629	803.3145	535.87633	plate 2 / well B10	58.4	15.20
SSSFREMDGQPER	IPI00329761	ABL1	1	2	1604.629	803.3145	535.87633	plate 2 / well C10	59.7	15.04

TGMGSGSAGKEGGPFK	IPI00329667	KIAA0528	1	5	1546.6486	774.32432	516.54955	plate 1 / well C10	61.0	15.22
TGMGSGSAGKEGGPFK	IPI00329679	KIAA0528	1	7	1546.6486	774.32432	516.54955	plate 1 / well D10	58.7	15.28
VSGRTSPPLLDR	IPI00329740	SRRM2	1	2	1376.6813	689.34064	459.89376	plate 2 / well A11	56.8	15.99
VSGRTSPPLLDR	IPI00329751	SRRM2	1	6	1376.6813	689.34064	459.89376	plate 2 / well B11	55.5	15.87
YIEDEDYYK	IPI00329675	PTK2B	1	7	1316.4849	659.24247	439.82831	plate 1 / well D 6	56.0	15.63
YIEDEDYYK	IPI00329687	PTK2B	1	8	1316.4849	659.24247	439.82831	plate 1 / well E 6	54.4	15.50