Chapter 3
Simplifying the Proteome: Analytical Strategies for Improving Peak Capacity

Lee A. Gethings and Joanne B. Connolly

Abstract The diversity of biological samples and dynamic range of analytes being analyzed can prove to be an analytical challenge and is particularly prevalent to proteomic studies. Maximizing the peak capacity of the workflow employed can extend the dynamic range and increase identification rates. The focus of this chapter is to present means of achieving this for various analytical techniques such as liquid chromatography, mass spectrometry, and ion mobility. A combination of these methods can be used as part of a data-independent acquisition strategy, thereby limiting issues such as chimericy when analyzing regions of extreme analyte density.

3.1 Introduction

The term "proteomics" encompasses the large-scale qualitative and quantitative study of proteins in a complex biological sample. The potential to understand the nature of a biological system using this information, in conjunction with other multi-omic data, is extensive. Current mass spectrometric techniques allow large-scale, high-throughput analyses for the identification, characterization, and quantification of the proteome, but can encounter major limitations in effectively mining complex biological matrices. A typical bottom-up proteomic experiment, for example, can result in areas of extreme analyte density during liquid chromatography (LC) separation (Fig. 3.1). These highly concentrated analyte regions are not only problematic from a chromatography aspect but can also be an issue with respect to the mass domain separation and detection.

Overcoming such challenges and increasing separation capabilities of analytical systems involve finding means of increasing peak capacity and dynamic range of

L.A. Gethings (✉) • J.B. Connolly
Waters Corporation, Stamford Avenue, Altrincham Road,
Wilmslow, SK9 4AX, Manchester, UK
e-mail: lee_gethings@waters.com

A.G. Woods and C.C. Darie (eds.), *Advancements of Mass Spectrometry in Biomedical Research*, Advances in Experimental Medicine and Biology 806,
DOI 10.1007/978-3-319-06068-2_3,

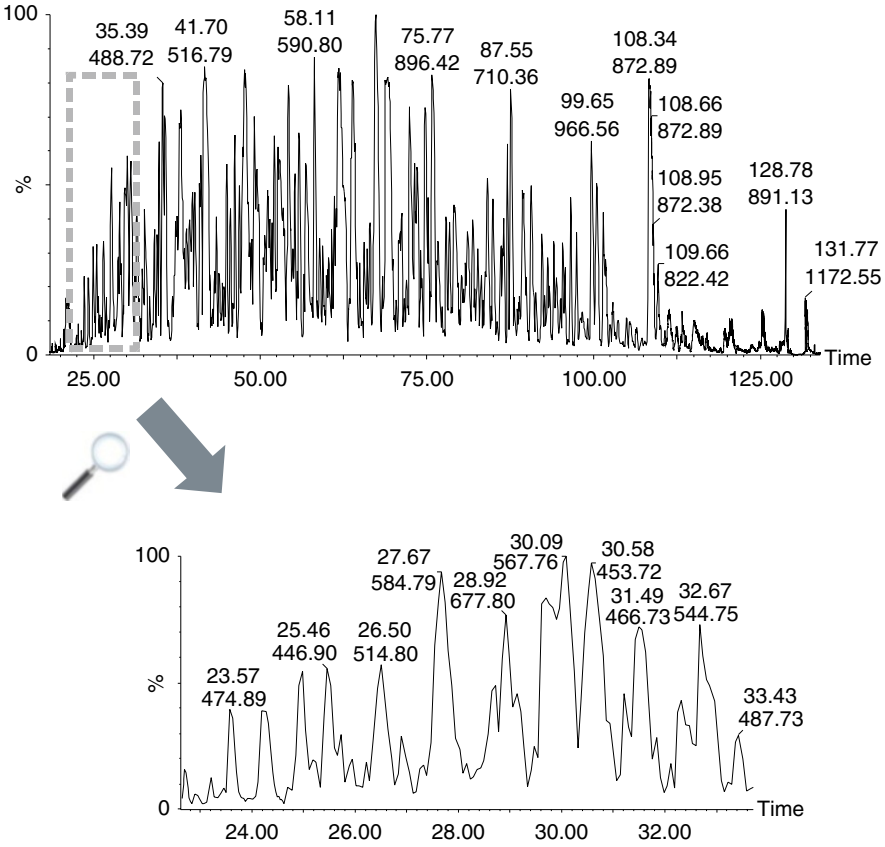

Fig. 3.1 Sample complexity and regions of analyte density represented by a 1,000 ng on-column loading of HeLa tryptic digest, separated over 150 min gradient. The lower figure shows the chromatogram expanded (23–33 min) with peak width half-height between 0.1 and 0.22 min

the workflow. Peak capacity is defined as the maximum number of components that can be separated to a specific resolution within a given separation space [1]. Throughout the course of this chapter, methodologies to increase system peak capacity will be introduced, including multidimensional chromatography and ion mobility (IM). Combined with a data-independent acquisition (DIA) strategy, these methods can provide enhanced sensitivity, specificity, and speed to biomarker discovery applications.

3.2 Expanding Peak Capacity from the LC Perspective

High performance liquid chromatography (HPLC) has and continues to be routinely used in combination with mass spectrometry (MS) for the separation of analytes. However, the performance of this technology has been augmented in recent years

with the introduction of ultra performance liquid chromatography (UPLC), making it possible to separate components from highly complex matrices with enhanced chromatographic resolution, sensitivity, and increased throughput. Innovative instrument design and advancements in column chemistries have contributed to these enhancements, particularly with the adoption of smaller column particle size for improved efficiency and pumping systems that can operate at elevated pressure, whilst maintaining the same selectivity and retention characteristics as HPLC equivalents [2]. Derivation of a mathematical expression which utilizes the van Deemter equation [3] illustrates how peak capacity is influenced by particle size, column length, diffusivity of the sample, linear velocity, and the length of gradient (Eq. 3.1). The terms used are defined as column length (L), equivalent theoretical plate height (H), analyte diffusion coefficient (D_m), packing material particle size (d_p), and empirical coefficients originating from the van Deemter equation (a, b, and c). The relationship between the logarithm of the capacity factor (k') and solvent composition is expressed as B, with Δc representing the difference in solvent composition over the course of the gradient. Finally, t_g/t_0 is defined as the gradient span.

$$n_{\mathrm{LC,gradient}} = 1 + \frac{1}{4}.\sqrt{\frac{L}{a.d_p + \frac{b.D_M}{u} + c.\frac{d_p^2}{D_M}.u}}.\frac{B.\Delta c}{B.\Delta c.\frac{t_0}{t_g} + 1} \tag{3.1}$$

If we consider the effect of reducing column particle size (d_p), transitioning from a 5 μm HPLC column packing to 1.7 μm stationary phase particles, efficiency and throughput can be improved up to a factor of three, but can generate backpressure increases by a factor of 27 [4]. Utilization of these high pressures results in enhanced separation efficiencies as represented by Fig. 3.2. Comparative chromatograms of glycogen phosphorylase B separated over the same length of gradient using HPLC and UPLC highlight increased column efficiency, resolution, and high throughput with UPLC. Sensitivity gains with UPLC show greater separation capability therefore limiting ion suppression and allow low intensity species to be readily identified. Factors other than d_p can influence the optimization of peak capacity from the LC platform. Three additional parameters, length of gradient, flow rate, and column length, also have roles to serve. An increased gradient length (t_g/t_0) provides higher peak capacities before eventually reaching a plateau where maximum peak capacity is attained. Within the scope of proteomics, gradients tend to routinely operate between 90 and 120 min depending on the complexity of sample being analyzed. Since longer gradients are employed with flow rates of less than 1 μL/min, the use of longer column lengths (L) to further increase peak capacity is viable. This is particularly important for the diverse range of biological samples discussed in the context of this review.

Multidimensional chromatography is a technique introduced as far back as 1953 [5], which described the use of two-dimensional (2D) paper chromatography and electrophoresis for peptide separation. It was soon adapted for 2D gel electrophoresis [6], but a number of drawbacks, such as the potential to modify proteins

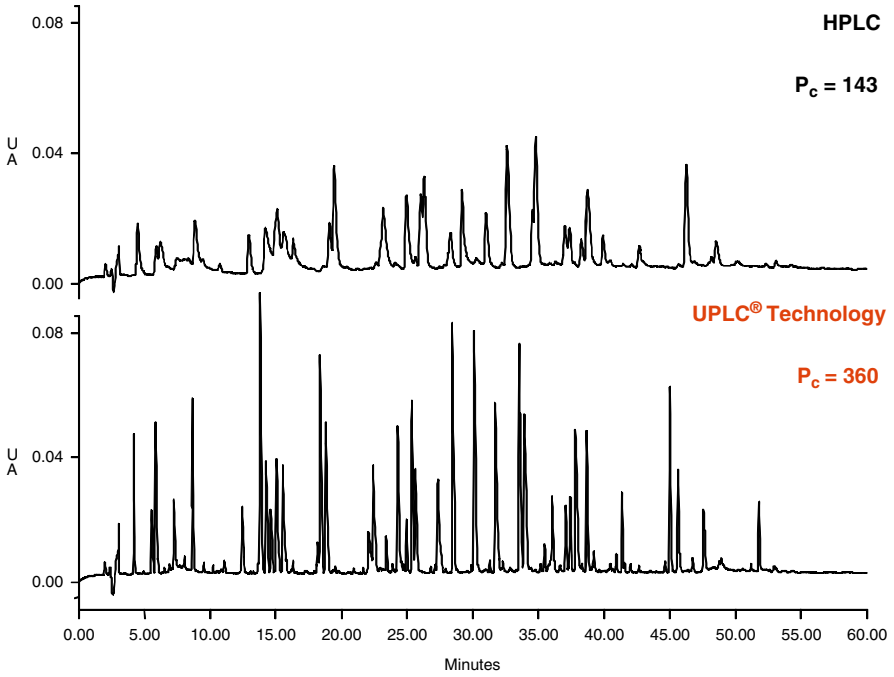

Fig. 3.2 Comparison of (**a**) HPLC (5 μm particles) with (**b**) UPLC (sub-2 μm particles) for the separation of a phosphorylase B tryptic digest

(i.e., free, unpolymerized acrylamide binding to the amino-terminal thereby preventing Edman degradation), technical reproducibility, and time constraints associated with performing gel analysis [7], make LC an attractive alternative. At the peptide level, as typically applied in bottom-up proteomic studies, 2D-LC provides high resolving power and ultimately increased peak capacity. Since the first and second dimensions have their individual peak capacities, $n_{c,1}$ and $n_{c,2}$, respectively, it is the product and not the sum of both [8, 9] that contributes to the overall peak capacity. This so-called product rule ($n_{c,2D}$) provides resolving powers that are not a true representation and often overestimate due to under-sampling effects of the first dimension. Accounting for the second-dimension separation cycle ($t_{c,2}$), the first-dimension peak capacity ($n_{c,1}$), and the first-dimension gradient time ($t_{g,1}$) [10], the influence of under-sampling on effective 2D-LC peak capacity shows no dependency on the first-dimension peak capacity (Eq. 3.2) with n_{2DLC} becoming less dependent on n_1.

$$n_{2DLC} = \frac{n_1 . n_2}{\sqrt{1 + 3.35 . \frac{t_{c,2} . n_{c,1}}{t_{g,1}}}} \tag{3.2}$$

In order to address the issue of under-sampling, an approximate model can be generated (Eq. 3.3), which assumes that the retention characteristics of the first and second dimensions are not correlated with little or no qualitative differences for non-orthogonal dimensions.

$$n_{2\mathrm{DLC}} \cong \frac{t_{g,1}.n_{c,2}}{1.83.t_{c,2}} \tag{3.3}$$

2D-LC experiments can be performed either off-line or on-line, depending on the configuration available. Off-line methods are often simpler, requiring fractions resulting from the first dimension to be collected prior to the second-dimension phase of separation, providing greater flexibility since the user is less restricted to the choice of eluent used in the first dimension. However, the risk of encountering peptide losses is increased. Implementing an on-line approach can work effectively, provided the selectivity between the two chemistries and compatibility of eluents are feasible. In most cases, the choice of the first-dimension conditions affords flexibility which can encompass a wide range of chemistries [11], such as strong cation exchange (SCX) [12–15], size exclusion chromatography (SEC) [16–19], immobilized metal ion affinity chromatography (IMAC) [20–24], or hydrophilic interaction chromatography (HILIC) [25–27]. The second dimension, however, tends to remain the same as that implemented for a 1D scenario (i.e., reversed phase). Of the examples provided, on-line SCX-RP is a common approach for peptide separation. The technique is partially orthogonal; however, distribution over multiple fractions is typically observed. An equally orthogonal technique, providing highly conserved peptide fractions and reducing the need for high salt concentration as is the case with SCX, is the use of a reversed phase–reversed phase (RP–RP) configuration [28]. Achieving RP–RP orthogonal separation is by differentiating on the basis of peptide interactions with the RP surface at differing charge states [29–32]. Considering acidic peptides, which are negatively charged at basic pH, this will result in early eluting fractions consisting of acidic peptides. Conversely for the second dimension at acidic pH, the acidic peptides eluting from the first dimension will now become "neutral" and hence switch their order of elution to becoming later eluting peptides. In the case of basic peptides the elution order is simply the reverse analogy described for acidic peptides. This scenario is clearly demonstrated in Fig. 3.3 which is a chromatogram of bovine haemoglobin tryptic digest. The chromatogram clearly shows orthogonal behavior of acidic and basic peptides across two dimensions operating at different pH. An RP–RP setup can be designed to comprise as many fractions as required. However, greater fractionation will compromise the analysis time and hence overall throughput, which also holds for off-line approaches. The higher peak capacity provided by RP–RP results in an increase of peptide and protein identifications [28, 33, 34]. To illustrate this, an example is shown in Fig. 3.4a, which represents comparative data from 1D, 2D 5-fraction, and 2D 10-fraction experiments for the bottom-up (2D) LC–DIA-MS analysis of cytosolic *Escherichia coli*. Sample loadings consisted of 750 ng (1D), 2.4 µg (2D 5-fraction), and 4.8 µg (2D 10-fraction). Confident peptides and proteins are shown, with significant gains

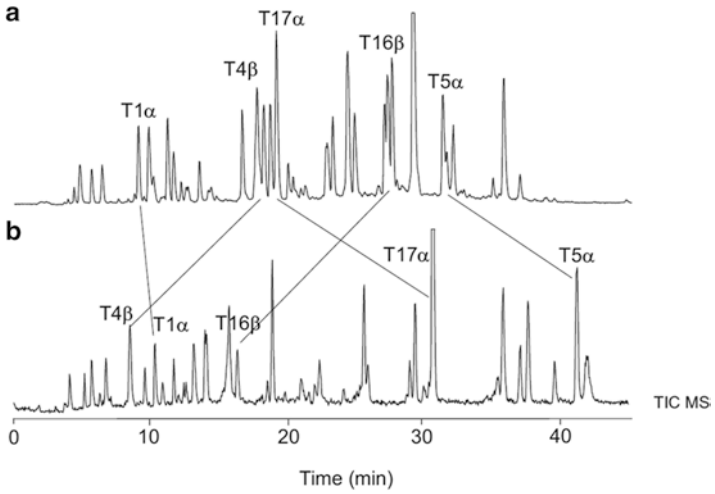

Fig. 3.3 The orthogonal nature of RP–RP demonstrated using a tryptic digest of bovine haemoglobin. The acidic (T16β, T4β), "neutral" (T1α), and basic (T17α, T5α) peptides clearly show a shift in retention order between pH 2.6 (**a**) and pH 10 (**b**) (Reprinted with permission from Gilar et. al., J. Sep. Sci., (2005), **28**, 1694–1703)

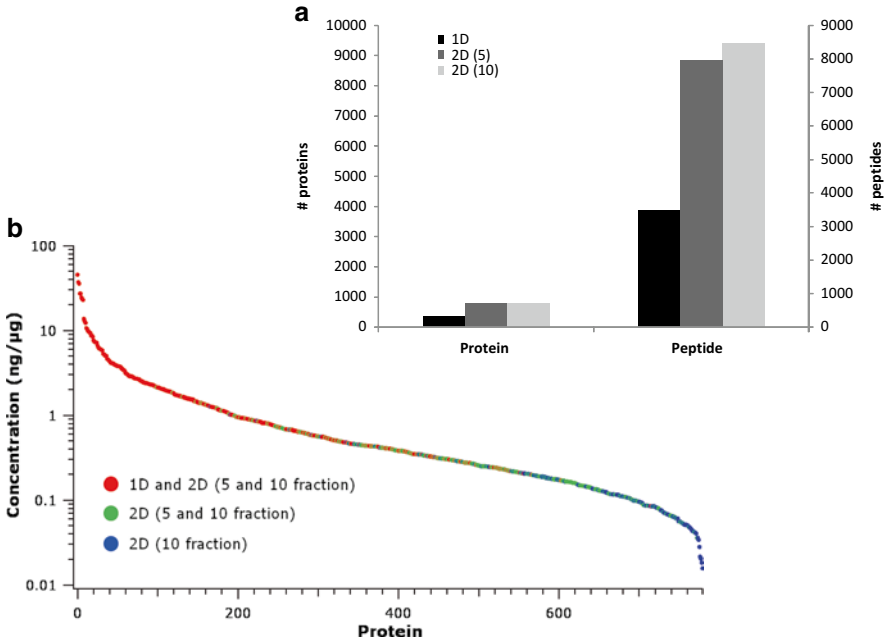

Fig. 3.4 Comparing 1D RP and 2D RP–RP (5- and 10-fraction) analysis of *Escherichia coli* shows (**a**) increased peptide and protein identifications with greater fractionation; (**b**) extended dynamic range gained with 2D RP–RP increasing the capability of identifying the less abundant peptides

achieved when 2D RP–RP is implemented. Transitioning from 1D to 2D (5-fraction) resulted in a 47 % increase in terms of protein identifications and an additional 11 % gain when fractionation is increased from 5 to 10 steps. This percentage gain can be accounted for with increased sample loadings and the ability to sample over a larger dynamic range (Fig. 3.4b).

3.3 Is Mass Resolution Alone Sufficient for Dealing with Sample Complexity?

Over recent years there have been significant gains in the mass resolution that can be achieved from mass spectrometers of various geometries. Mass resolution is an important parameter, providing a means of differentiating ions resulting from a complex sample. Resolution (R) is defined as a measure or capacity to distinguish ions of adjacent mass number, m and Δm, respectively (i.e., $R = m/\Delta m$). It is important however to differentiate between working resolution of an instrument and the resolution which can be obtained in practice, since acquisition speed and m/z are determining factors [35].

A range of high resolution mass analyzers are routinely used for biomarker discovery experiments including traps and quadrupole time-of-flight (Q-ToF) geometries. However, it is the geometry of a Q-ToF mass spectrometer which will be the discussed platform for the remainder of this chapter [36]. The geometry of a Q-ToF is a tandem version of the orthogonal accelerated ToF (oa-ToF), consisting of an MS1 quadrupole and collision region with reflecting oa-ToF MS as MS2. Data can be acquired in MS and MS/MS modes, providing high mass accuracy and sensitivity. These high levels of sensitivity are achieved compared with scanning instruments, since orthogonal geometry is applied for the detection of ions as opposed to sequential detection. A significant innovation to the original design saw the hybridization of ion mobility (IM) (Fig. 3.5) [37]. The principle of IM involves the separation of a packet of ions based on mobility differences as they drift through an inert gas under the influence of a weak electric field [38] and is a well-established technique for the structural analysis of proteins [39], but can also be applied to bottom-up proteomic applications [38]. Traditional IM experiments are typically performed using a drift tube platform, which achieves separation by applying a uniform, static electric field. In the case of IM-MS implemented within the geometry of a Q-ToF instrument, mobility separation is achieved by using an RF ion guide, termed travelling wave ion guide, to generate a travelling voltage wave, which has been thoroughly reviewed in the literature [40, 41]. Combining the travelling wave principle of the device with elevated inert gas pressures results in a proportion of traversing ions rolling back on the wave, whilst the average effect propels ions in the direction of the travelling wave. Ion species with low mobility traverse slower, whereas ions of higher mobility traverse faster, resulting in a shorter drift time. Comparative advantages associated with IM-MS compared to drift tube variants include higher sensitivity and sufficiently faster data acquisition. However, drift tubes still have a

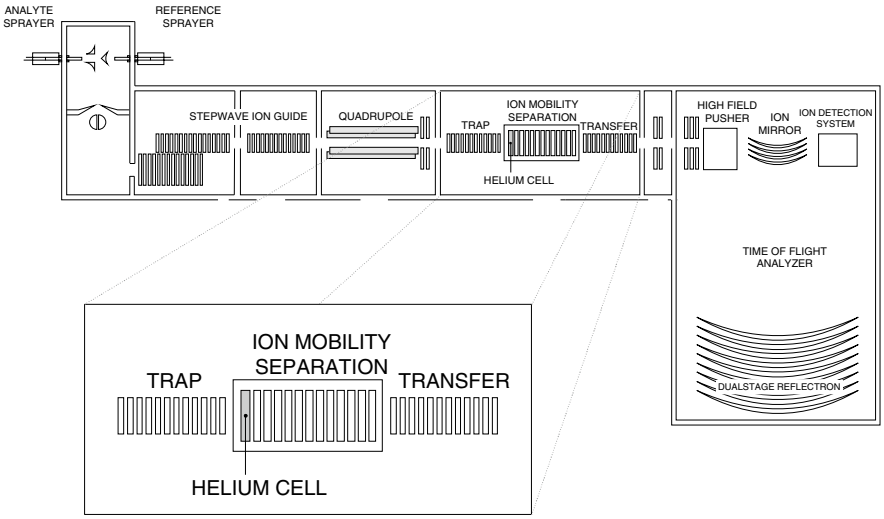

Fig. 3.5 A hybrid quadrupole-ion mobility-orthogonal acceleration time-of-flight mass spectrometer

distinct advantage in terms of the mobility resolution capabilities, which can be typically three orders of magnitude over that of current IM-MS [42].

Separation by IM operates within a millisecond timescale which fits perfectly between LC and ToF, which operate at the second and microsecond scale, respectively. The three domains are not completely orthogonal; however, the resulting system peak capacity that can be typically achieved is between 10- and 25-fold. It has been illustrated how peak capacity can be determined from a chromatographic standpoint. A similar approach can also be adopted for an oa-ToF mass analyzer (Eq. 3.4) and ion mobility separation (Eq. 3.5).

$$n_{\text{oaToF}} = \frac{\frac{R_{\text{oa-ToF}}}{m}.W_m}{n_1} \tag{3.4}$$

$$n_{\text{TWIM}} = \int_{k_{\min}}^{k_{\max}} \frac{R_{\text{IM,max}}}{\sqrt{k.k_{\max}}}\,dk \tag{3.5}$$

The peak capacity contribution made by the MS dimension, as outlined by Eq. 3.4, is defined by the mass spectrometer resolution at full width half maximum ($R_{\text{oa-ToF}}$), monoisotopic mass (m), mass distribution (W_m), and the number of identified isotopic peaks within a spectrum (n_l). Conversely, IM also contributes additional peak capacity provided by the travelling wave device [43], describing the relationship between resolution and ion mobility, where k represents mobility and $R_{\text{IM,max}}$ is an empirically derived maximum resolution travelling wave mobility separation. Based on the assumption that all three domains are completely orthogonal a product

describing the entire system peak capacity would be as defined by Eq. 3.6. In practice, however, the degree of orthogonality is compromised as discussed earlier, since the analytes rely on having the same physiochemical properties.

$$n_{\text{2DLC,IMS,oaToF}} \sim n_{\text{2DLC}} . n_{\text{TWIM}} . n_{\text{oaToF}} . f \tag{3.6}$$

3.4 Integrating Ion Mobility into a Data-Independent Strategy as Means of Increasing System Peak Capacity

A previous chapter introduced the concept of data-independent acquisition (DIA) for acquiring bottom-up proteomic data sets. Combining DIA with the strategies discussed here to increase peak capacity results in a sophisticated workflow that utilizes ion mobility to provide IM-DIA-MS. The principle of this method works in a similar manner as DIA, but with the additional degree of separation and specificity that is afforded by IM. Based on the instrument geometry shown in Fig. 3.5, IM-DIA-MS operates whereby the quadrupole mass analyzer is non-resolving. The collision cell is located within the travelling wave ion guide region and when operating in DIA mode, it is the primary stacked ring ion guide (SRIG) which is used to induce fragmentation. However, for IM-DIA-MS it is the secondary collision SRIG located after the ion mobility SRIG which is utilized for fragmentation. The primary SRIG maintains a static CE ensuring precursors are IM separated prior to fragmentation within the secondary SRIG, resulting in fragment ions sharing the same drift time as their associated precursors. This builds additional specificity into the analysis, since drift time and chromatographic retention time can now be used to correlate fragment ions with their respective precursors. For situations where the analyte density is large (i.e., the midpoint of a chromatographic gradient) the opportunity for multiple precursors to be present in the collision cell at a single time event is highly probable. The co-isolation of precursors sharing similar *m/z* and retention time is a phenomenon termed as chimericy. Since multiple MS/MS fragments are generated from multiple precursors the identification rate is significantly hampered resulting in unidentified fragments and hence under-sampling of the proteome. It has been reported that for highly complex samples, chimeras may be as high as 50 % of total spectra [44]. Relying solely on chromatography and mass resolution is not adequate to counter the effect. The additional separation capabilities provided by IM though do provide an opportunity to increase selectivity in those areas of extreme analyte density. Implementing IM-DIA overcomes some of the challenges associated with data-dependent analysis (DDA), such as the most abundant peptides typically being sampled and chimeric effects. Thoughts of increased sampling speed and sensitivity of instrumentation alone are deemed as insufficient and alternatively merging high resolution with a form of multiplexing (i.e., ion mobility) is thought to be necessary for comprehensive proteomic analysis [45]. Product ions are tentatively associated to precursors by means of a pre-database searching step. Isotope and charge state

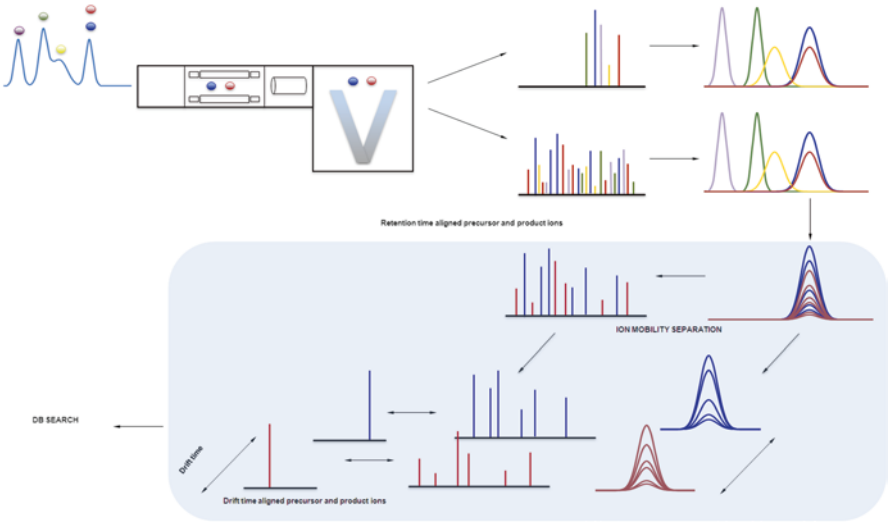

Fig. 3.6 Principle of IM-DIA showing retention time alignment for precursor and product ions is shown with additional drift time alignment for ion mobility workflows

information is collected for all precursors in addition to performing time and drift-alignment correlation (Fig. 3.6). Retention time-alignment correlation can be visualized by plotting the elution profiles of product ions from the elevated energy trace, resulting in identical elution profiles as their respective precursors. This chromatographic apex retention time principle forms the basis for associating precursor and product ions. The reasoning described for the alignment of chromatographic profiles can also be replicated with respect to the drift time domain, since fragment ions will share the same drift time as their associated precursor, thereby building additional specificity into the workflow.

Acquiring data sets using the methodologies outlined thus far is illustrated by a study involving the analysis of a *Rattus Norvegicus* exosome sub-proteome treated with either galactosamine (GalN) or lipopolysaccharide (LPS) culture media [46]. Comparative strategies were conducted involving 1D- and 2D (RP–RP)-LC in combination with DIA or IM-DIA. A three-way comparison of the results gained from 1D-LC–DIA-MS, 1D-LC–IM-DIA-MS, and 2D-LC–DIA-MS (Fig. 3.7) shows significant overlap with a large number of additional unique proteins identified with the implementation of IM. On average a 20 % increase is observed for the average number of peptides assigned to a protein suggesting the IM workflow to be more effective at sampling the proteome. Additional benefits include lower material consumption whilst increasing throughput and enhancing specificity of the identifications returned. The interaction between sensitivity and specificity has previously been addressed and has shown that increasing sensitivity alone does not result with increased identifications unless accompanied with additional specificity [46]. This is exemplified for the identified proteins of this study showing greater than twofold increase over the entire dynamic range (Fig. 3.8).

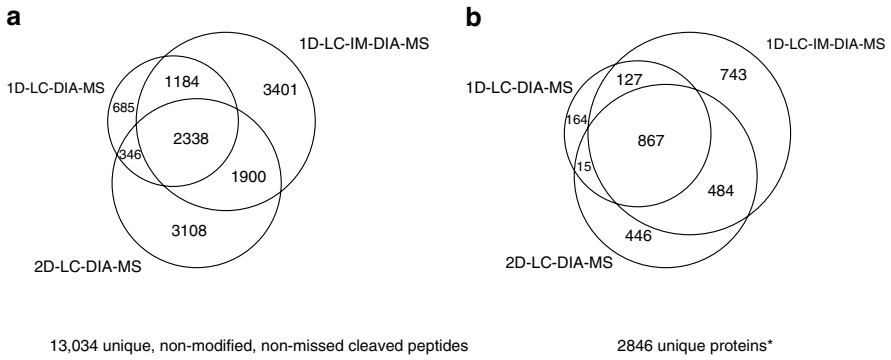

Fig. 3.7 A comparative evaluation of peptides (**a**) and proteins (**b**) identified from treated *Rattus Norvegicus* exosome samples. Both the peptide and protein results are subdivided into 1D (DIA), 1D (IM-DIA), and 2D (DIA) analyses. Identifications resulting from 1D were acquired at 10,000 FHWM resolution, whilst 2D was performed at 20,000 FHWM resolution (Rodriguez-Suárez et. al., Current Analytical Chemistry, 2013, 9, 199–211)

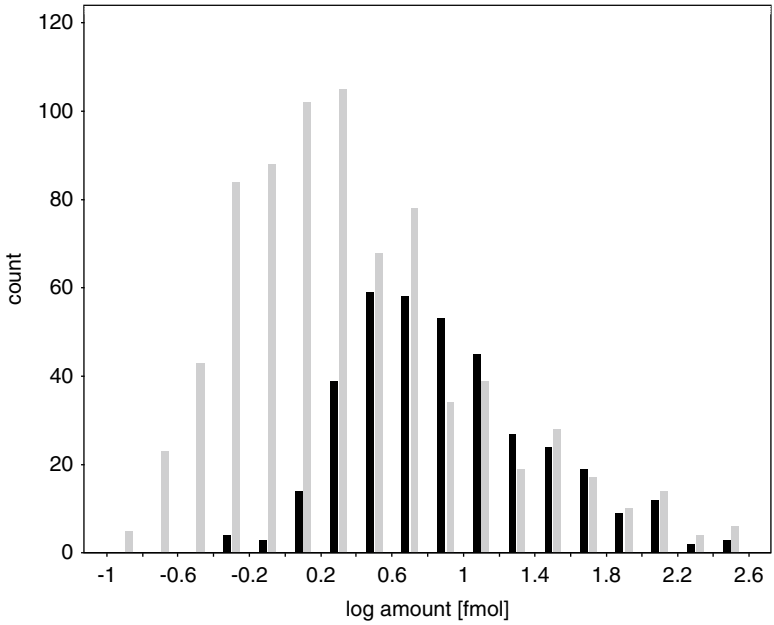

Fig. 3.8 Based on a 1 µg loading of control *Rattus Norvegicus* exosome, the normalized estimated molar amounts can be plotted to show a distribution expanding over two orders of magnitude of dynamic range. Data resulting from 1D-LC–DIA-MS (*black*) and 1D-LC–IM-DIA-MS (*gray*) is provided (Rodriguez- Suárez et. al., Current Analytical Chemistry, 2013, 9, 199–211)

The bulk of these discussions have focused on implementing analytical workflows, yet thought should also be given to the informatic strategy applied to such data sets. For example, the protein identification rate could be further improved by implementing the use of spectral libraries and/or DIA fragment ion repositories [47].

3.5 Informatics: Processing IM-DIA-MS Data Sets

The nature of DIA and IM-DIA strategies does not rely on instrument selection of the precursors as with DDA; the capacity to relate precursors with their fragment ions is therefore reliant on post-acquisition informatics using sophisticated algorithms for peak detection [48, 49], precursor–product ion correlation, and database searching [50].

3.5.1 Precursor–Product Ion Correlation

The tentative association of precursor and product ions as part of the pre-database search step is performed prior to the data being compiled into a list containing four dimensional attributes (ion mobility, retention time, *m*/*z*, and intensity) providing a multidimensional distribution. The presence of background noise can result in over-counting and therefore needs to be accounted for by applying a convolution filter to suppress the effect. Each of these deconvoluted measurements is referred to as accurate-mass, retention time (AMRT) components. For cases where multidimensional chromatography has been used, raw data files are processed individually before merging the processed outputs prior to database searching, ensuring that all peptides representing a single protein can be identified and quantified in a single event. Should peptides be shared between fractions, the summing process employed allows quantification.

3.5.2 Database Searching

A comprehensive overview discussing a search algorithm for qualitative identifications based on DIA data sets has been described in the literature [51]. Briefly, following peak detection and alignment of the low and elevated energy AMRTs a constructed list of precursor/product ion associations is interrogated for putative peptide identifications. A precursor/product subset list and a user-defined database are subjected to a pre-assessment search based on a set of physiochemical properties specific for tryptic peptides and proteins, which are applicable to both the liquid and gas phase dimensions. The database element consists of a decoy database (either randomized or reversed) being merged with the original user database, allowing a

false discovery rate and minimum protein score to be automatically determined. The search and score models are optimized using the previously mentioned physiochemical properties. The next stage of the algorithm search is divided into three steps. The first step queries the precursor/product ion entries against the protein database for sequences containing no tryptic-missed cleavages or variable modifications. A combination of precursor and fragment ion mass tolerances and accuracy of fit to the physiochemical properties provides a mechanism by which peptide identifications are scored and ranked. This process is continuously repeated until either the false-positive rate is exceeded or protein identifications no longer exceed the minimum score. The second step of the query specifically focuses on identifying modifications and nonspecific cleavages for protein entries resulting from the first pass as well as in-source fragments and precursor neutral losses. The final step of the query aligns any unidentified low and elevated energy AMRTs from the previous two steps before searching against the full database with no limitation on product ion intensity and restriction on the number of modifications per protein.

3.6 Importance of Peak Capacity for Systems Biology: A Multi-omic Biomarker Case Study

Complex data sets as described can be readily generated from an analytical perspective. The inherent difficulty arises with the search results and interpreting these into a biological context that can provide insight of both physiology and disease state of the model being studied. Data sets can often be large and originate from not only proteomics but also other omic areas such as metabolomics, lipidomics, genomics, and transcriptomics to provide a comprehensive understanding of biochemical processes and ultimately provide potential biomarkers. Since there is no single analytical system or workflow that can generate information for all the omic areas, gaining access to data of a common format for searching and mapping pathways can prove to be troublesome. Figure 3.9 represents an example workflow detailing how multi-omic data sets are integrated and interrogated for system networks analysis [52].

A case study representing multi-omic data using techniques to improve peak capacity for potentially identifying disease biomarkers is described here. The study focuses on a rare genetic kidney disorder termed idiopathic nephritic syndrome (INS), affecting only a small percentage of paediatric patients. The condition arises from a faulty glomerulus, resulting in proteinuria additionally characterized by edema, hypoalbuminemia, and hyperlipidemia, as well as increased levels of cholesterol and triglyceride [53]. The study cohort consisted of urine samples from paediatric subjects, control, and INS diagnosed. Samples were purified and prepared appropriately for proteomic or metabolomic analysis [54, 55] prior to an experimental strategy which combined LC and DIA, with the proteomic experiments specifically utilizing IM-DIA-MS as a means of generating label-free data. In order to establish differences between the two groups, statistical analysis using multivariate methods of both data sets showed significant deviations between both cohorts.

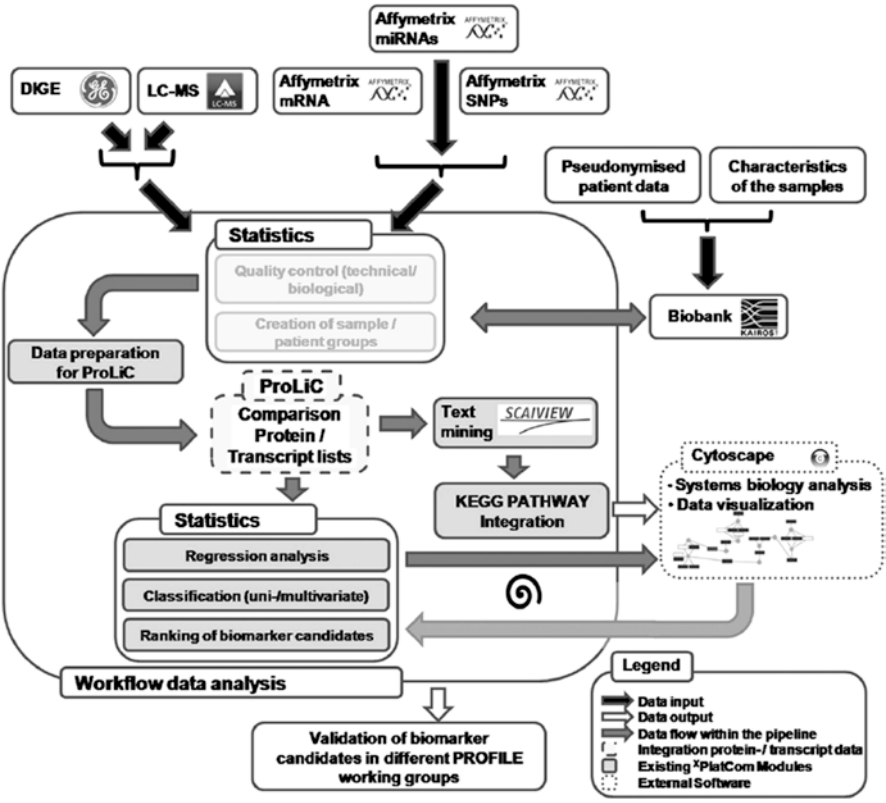

Fig. 3.9 Example pipeline for compiling and interrogating multi-omic data sets to pathway and network mapping studies (Kohl et. al., Biochim. Biophys. Acta, 2014, 1844, 52–62)

Identifying analytes responsible for such perturbation started by reviewing the protein data. In total more than 300 proteins were identified, with 80 % showing significant fold change (greater than two) and *p*-values of less than 5 % across all subjects. Hierarchical analysis (Fig. 3.10a) allows the visual identification of protein expression trends across conditions in addition to potential intra-subject variances. Regulated proteins showing distinct expression trends can be interrogated further as demonstrated with the example peptide associated to the prostaglandin receptor (Fig. 3.10b). The metabolite data was interpreted differently by means of constructing a loadings plot (Fig. 3.11) from previously derived OPLS-DA analysis. The loadings plot is constructed such that compounds contributing the greatest variance with the highest probability are target analytes for identification. In an effort to combine the data streams and understand their contributions in a biological context, pathway analysis was conducted to explore and visualize interactions and networks resulting from both omic data sets. The resultant network indicates relevant pathways such as chronic fatigue syndrome and neurological signs as major contributors.

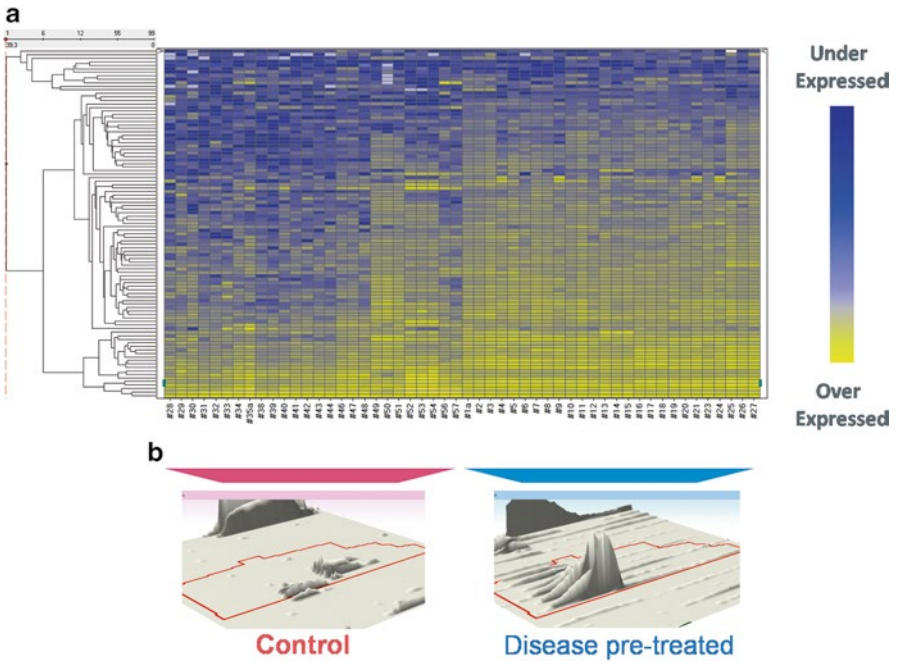

Fig. 3.10 (**a**) Hierarchical cluster analysis of statistically (ANOVA) significantly relevant regulated urinary proteins identified with 3 or more peptides and fold changes greater than 2. Subjects grouped within the pink banner are control, whilst disease pre-treated subjects correspond to the blue banner. The 3D montage images (**b**) are representative of prostaglandin receptor peptide TMLLQPAGSLGSYSYR

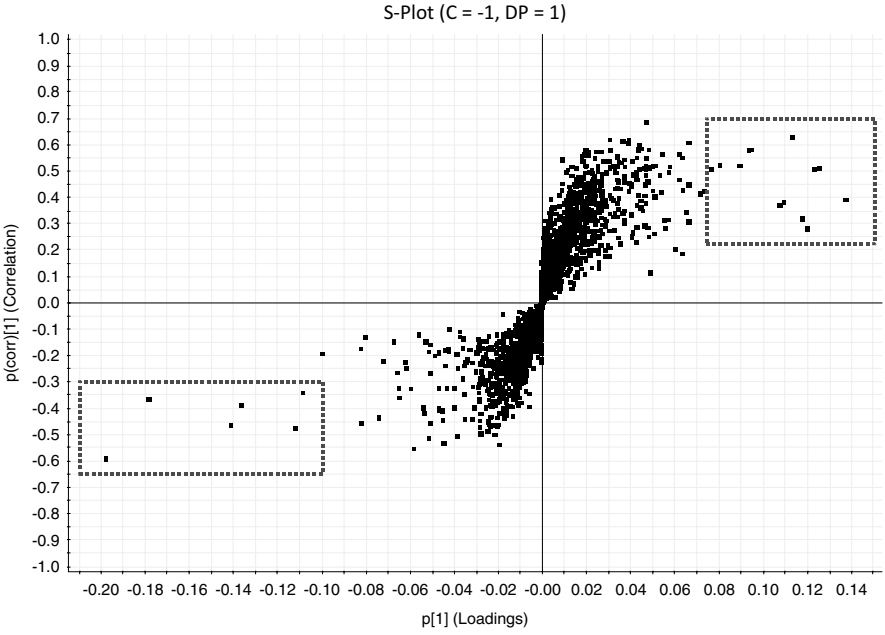

Fig. 3.11 Metabolite loadings plot resulting from OPLS-DA of disease pre-treated versus control subjects based in positive ion mode. Metabolites contributing the greatest variance such as hydroxyphenyl acetate and uridine are represented within the highlighted areas

3.7 Summary

The quest for identifying potential biomarkers is complicated and challenging with the majority of studies involving highly complex samples over wide dynamic ranges that require sophisticated analytical workflows to acquire, process, and biologically interpret data sets. The most interesting and often most significant proteins are those which are to be found at low abundance. From an LC–MS perspective, a number of potential strategies have been introduced to address how the proteome can be explored at greater depth, whilst limiting sample consumption and maximizing throughput. Multidimensional chromatography combined with ion mobility enabled DIA acquisition schema demonstrate how system peak capacity can be optimized. As technologies advance the volume of data increases in size and complexity; therefore, informatic requirements and capabilities should also be addressed. These combined efforts show not only increased peak capacity but also enhanced specificity, afforded by the chromatographic and drift time alignment of precursor and product ions, providing additional confidence to identifications.

Advancements towards finding relevant biomarkers are increasingly moving towards a system biology approach, whereby multi-omic data sets can provide a great insight into pathway information and the interaction of networks. Amalgamating techniques such as IM-enabled workflows with multi-omic biological pathway analysis has the potential to assist in our understanding of disease and its etiology for drug discovery.

Acknowledgements The authors would like to thank our collaborators who have provided figures and granted permission to use their collateral. In particular we would like to acknowledge Dr. Stefan Tenzer, Dr. Eva Rodriguez-Suárez, Dr. Sandra Kraljević Pavelić, Dr. Martin Gilar, Dr. John Shockcor, Kenneth Fountain, and Eric Grumbach. Finally Dr. Johannes P.C. Vissers is thanked for constructive comments during review.

References

1. Giddings JC (1984) Two-dimensional separations: concept and promise. Anal Chem 56: 1258A–1260A, 1262A, 1264A
2. Swartz ME (2007) Separation science & technology, vol 8. Academic, pp 145–147
3. Van Deemter JJ, Zuiderweg FJ, Klinkenberg A (1956) Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. Chem Eng Sci 5:271
4. Plumb R, Castro-Perez J, Granger J, Beattie I, Joncour K, Wright A (2004) Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 18:2331–2337
5. Levy AL, Chung D (1953) Two-dimensional chromatography of amino acids on buffered papers. Anal Chem 25:396–399
6. Honneger CG (1961) Dünnschicht-Ionophorese und Dünnschicht-Ionophorese-Chromatographie. Helv Chim Acta 44:173
7. Rabilloud T (2002) Two-dimensional gel electrophopresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics 2:3–10

8. Neue UD, Mazzeo JR (2001) A theoretical study of the optimization of gradients at elevated temperature. J Sep Sci 24:921–929
9. Karger BL, Snyder LR, Horvarth C (1973) An introduction to separation science. Wiley, New York
10. Li X, Stoll DR, Carr PW (2009) Equation for peak capacity estimation in two-dimensional liquid chromatography. Anal Chem 81:845–850
11. Dowell JA, Frost DC, Zhang J, Li L (2008) Comparison of two-dimensional fractionation techniques for shotgun proteomics. Anal Chem 80:6715–6723
12. Washburn MP, Wolters D, Yates JR (2001) Large-scale analysis of the yeast proteome by mułtidimensional protein identification technology. Nat Biotechnol 19:242–247
13. Wolters DA, Washburn MP, Yates JR (2001) An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem 73:5683–5690
14. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2:43–50
15. Vollmer M, Horth P, Nagele E (2004) Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. Anal Chem 76:5180–5185
16. Opiteck GJ, Jorgenson JW, Anderegg RJ (1997) Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. Anal Chem 69:2283–2291
17. Alvarez-Manilla G, James I, Guo Y, Warren NL, Orlando R, Pierce M (2006) Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. J Proteome Res 5:701–708
18. Preud'homme H, Far J, Gil-Casal S, Lobinski R (2012) Large-scale identification of selenium metabolites by online size-exclusion-reversed phase liquid chromatography with combined inductively coupled plasma (ICP-MS) and electrospray ionization linear trap-orbitrap mass spectrometry (ESI-MSn). Metallomics 4:422–432
19. Barqawi H, Ostas E, Liu B, Carpenter JF, Binder WH (2012) Multidimensional characterization of α, ω-telechelic poly(Є-caprolactone)s via online coupling of 2D chromatographic methods (LC/SEC) and ESI-TOF/MALDI-TOF-MS. Macromolecules 45:9779–9790
20. Albuquerque CP, Smolka MB, Payne SH, Bafna V, Eng J, Zhou H (2008) A multidimensional chromatography technology for in-depth phosphoproteome analysis. Mol Cell Proteomics 7:1389–1396
21. Palma SD, Zoumaro-Djayoon A, Peng M, Post H, Preisinger C, Munoz J, Heck AJR (2013) Finding the same needles in the haystack? A comparison of phosphotyrosine peptides enriched by immuno-affinity precipitation and metal-based affinity chromatography. J Proteomics 91:331–337
22. Frantzi M, Zoidakis J, Papadopulos T, Zürbig P, Katafigiotis J, Stravodimos K, Lazaris A, Giannopoulou I, Ploumidis A, Mischak H, Mullen W, Vlahou A (2013) IMAC fractionation in combination with LC-MS reveals H2B and NIF-1 peptides as potential bladder cancer biomarkers. J Proteome Res 12:3969–3979
23. Liu S, Hughes C, Lajoie G (2012) Recent advances and special considerations for the analysis of phosphorylated peptides by LC-ESI-MS/MS. Curr Anal Chem 8:35–42
24. Stasyk T, Huber LA (2012) Mapping in vivo signal transduction defects by phosphoproteomics. Trends Mol Med 18:43–51
25. Boersema PJ, Divecha N, Heck AJR, Mohammed S (2007) Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. J Proteome Res 6:937–946
26. Xie F, Smith RD, Shen Y (2012) Advanced proteomic liquid chromatography. J Chromatogr A 1261:78–90
27. Wang C, Yuan J, Wang Z, Huang L (2013) Separation of one-pot procedure released O-glycans as 1-phenyl-3-methyl-5-pyrazolone derivatives by hydrophilic interaction and reversed-phase liquid chromatography followed by identification using electrospray mass spectrometry and tandem mass spectrometry. J Chromatogr A 1274:107–117
28. Lau E, Lam MPY, Siu SO, Kong RPW, Chan WL, Zhou Z, Huang J, Lo C, Chu IK (2011) Combinatorial use of offline SCX and online RP-RP liquid chromatography for iTRAQ-based quantitative proteomics applications. Mol Biosyst 7:1399–1408

29. Gilar M, Olivova P, Daly AE, Gebler JC (2005) Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem 77:6426–6434
30. Gilar M, Olivova P, Daly AE, Gebler JC (2005) Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. J Sep Sci 28:1694–1703
31. Francois I, Cabooter D, Sandra K, Lynen F, Desmet G, Sandra P (2009) Tryptic digest analysis by comprehensive reversed phasextwo reversed phase liquid chromatography (RP-LCx2RP-LC) at different pH's. J Sep Sci 32:1137–1144
32. Francois I, de Villiers A, Tienpont B, David F, Sandra P (2008) Comprehensive two-dimensional liquid chromatography applying two parallel columns in the second dimension. J Chromatogr A 1178:33–42
33. Siu SO, Lam MPY, Lau E, Kong RPW, Lee SMY, Chu IK (2011) Fully automated two-dimensional reversed-phase capillary liquid chromatography with online tandem mass spectrometry for shotgun proteomics. Proteomics 11:2308–2319
34. Zhou F, Cardoza JD, Ficarro SB, Adelment GO, Lazaro JB, Marto JA (2010) Online nanoflow RP-RP-MS reveals dynamics of multicomponent Ku complex in response to DNA damage. J Proteome Res 9:6242–6255
35. Scigelova M, Hornshaw M, Giannokulas A, Makarov A (2011) Fourier transform mass spectrometry. MCP 1–19. doi:10.1074/mcp.M111.009431
36. Morris HR, Paxton T, Dell A, Langhorn B, Berg M, Bordoli RS, Hoyes J, Bateman RH (1996) High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. Rapid Commun Mass Spectrom 10:889–896
37. Pringle SD, Giles K, Wildgoose JL, Williams JP, Slade SE, Thalassinos K, Bateman RH, Bowers MT, Scrivens JH (2007) An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. Int J Mass Spectrom 261:1–12
38. Mason EA, McDaniel EW (1973) The mobility and diffusion of ions in gases. Wiley, New York
39. Ruotolo BT, Giles K, Campuzano I, Sandercock AM, Bateman RH, Robinson CV (2005) Evidence for macromolecular protein rings in the absence of bulk water. Science 310:1658
40. Harvey SR, Macphee CE, Barran PE (2011) Ion mobility mass spectrometry for peptide analysis. Methods 54:454–461
41. Giles K, Pringle SD, Worhtington KR, Little D, Wildgoose JL, Bateman RH (2004) Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. Rapid Commun Mass Spectrom 18:2401–2414
42. Dugourd P, Hudgins RR, Clemmer DE, Jarrold MF (1997) High-resolution ion mobility measurements. Rev Sci Instrum 68:1122–1129
43. Shvartsburg AA, Smith RD (2008) Fundamentals of travelling wave ion mobility spectrometry. Anal Chem 80:9689–9699
44. Houel S, Abernathy R, Renganathan K, Meyer-Arendt K, Ahn NG, Old WM (2010) Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies. J Proteome Res 9:4152–4160
45. Michalski A, Cox J, Mann M (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. J Proteome Res 10:1785–1793
46. Rodriguez-Suarez E, Hughes C, Gethings L, Giles K, Wildgoose J, Stapels M, Fadgen KE, Geromanos SJ, Vissers JPC, Elortza F, Langridge JI (2012) An ion mobility assisted data independent LC-MS strategy for the analysis of complex biological samples. Curr Anal Chem 9:199–211
47. Geromanos SJ, Hughes C, Golick D, Ciavarini S, Gorenstein MV, Richardson K, Hoyes JB, Vissers JP, Langridge JI (2011) Simulating and validating proteomics data and search results. Proteomics 11:1189–1211

48. Thalassinos K, Vissers JP, Tenzer S, Levin Y, Thompson JW, Daniel D, Mann D, Delong MR, Moseley MA, America AH, Ottens AK, Cavey GS, Efstathiou G, Scrivens JH, Langridge JI, Germoanos SJ (2012) Design and application of a data-independent precursor and product ion repository. J Am Soc Mass Spectrom 23:1808–1820
49. Geromanos SJ, Vissers JPC, Silva JC, Dorschel CA, Li GZ, Gorenstein MV, Bateman RH, Langridge JI (2009) The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependent LC-MS/MS. Proteomics 9:1683–1695
50. Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, McKenna T, Nold MJ, Richardson K, Young P, Geromanos S (2005) Quantitative proteomic analysis by accurate mass retention time pairs. Anal Chem 77:2187–2200
51. Li GZ, Vissers JPC, Silva JC, Golick D, Gorenstein MV, Geromanos SJ (2009) 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
52. Kohl M, Megger DA, Trippler M, Meckel H, Ahrens M, Bracht T, Weber F, Hoffmann AC, Baba HA, Sitek B, Schlaak JF, Meyer HE, Stephan C, Eisenacher M (2014) A practical data processing workflow for multi-OMICS projects. Biochem Biophys Acta 1844:52–62
53. Chesney RW (1999) The idiopathic nephrotic syndrome. Curr Opin Pediatr 11:158–161
54. Vaezzadeh AR, Briscoe AC, Steen H, Lee RS (2010) One-step sample concentration, purification, and albumin depletion method for urinary proteomics. J Proteome Res 9:6082–6089
55. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, Holmes E, Nicholson JK (2010) Global metabolic profiling procedures for urine using UPLC-MS. Nat Protoc 5:1005–1018