Shot-gun proteomics

Why thousands of unidentified signals matter

den Ridder, Maxime; Daran-Lapujade, Pascale; Pabst, Martin

DOI
10.1093/femsyr/foz088

Publication date
2020

Document Version
Final published version

Published in
FEMS Yeast Research

FEMS Yeast Research, 20, 2020, foz088

doi: 10.1093/femsyr/foz088
Advance Access Publication Date: 20 December 2019
Minireview

MINIREVIEW

Shot-gun proteomics: why thousands of unidentified signals matter

Maxime den Ridder, Pascale Daran-Lapujade[†] and Martin Pabst[*,‡]

Delft University of Technology, Department of Biotechnology, van der Maasweg 9, 2629 HZ Delft, The Netherlands

[*]**Corresponding author:**Delft University of Technology, Department of Biotechnology, van der Maasweg 9, 2629 HZ Delft, The Netherlands. Tel: +31(0)15 27 86369; E-mail: m.pabst@tudelft.nl

One sentence summary: The authors review recent developments in yeast proteomics for the discovery and analysis of post-translational modifications, which are important regulators in many cellular processes of yeast.

Editor: John Morrissey
[†]Pascale Daran-Lapujade, http://orcid.org/0000-0002-4097-7831
[‡]Martin Pabst, http://orcid.org/0000-0001-9897-0723

ABSTRACT

Mass spectrometry-based proteomics has become a constitutional part of the multi-omics toolbox in yeast research, advancing fundamental knowledge of molecular processes and guiding decisions in strain and product developmental pipelines. Nevertheless, post-translational protein modifications (PTMs) continue to challenge the field of proteomics. PTMs are not directly encoded in the genome; therefore, they require a sensitive analysis of the proteome itself. In yeast, the relevance of post-translational regulators has already been established, such as for phosphorylation, which can directly affect the reaction rates of metabolic enzymes. Whereas, the selective analysis of single modifications has become a broadly employed technique, the sensitive analysis of a comprehensive set of modifications still remains a challenge. At the same time, a large number of fragmentation spectra in a typical shot-gun proteomics experiment remain unidentified. It has been estimated that a good proportion of those unidentified spectra originates from unexpected modifications or natural peptide variants. In this review, recent advancements in microbial proteomics for unrestricted protein modification discovery are reviewed, and recent research integrating this additional layer of information to elucidate protein interaction and regulation in yeast is briefly discussed.

Keywords: Mass spectrometry; post-translational modifications; unrestricted modification search; protein regulation; yeast proteomics

INTRODUCTION

Mass spectrometry (MS) is currently the most powerful technology for the identification, characterisation and quantification of complex mixtures of proteins (Aebersold and Mann 2003; Cox and Mann 2011; Oliveira and Sauer 2012). The performance of mass spectrometers has steadily improved over recent decades, providing a near-complete yeast proteome coverage using short chromatographic separation times and minimum amounts of sample (Hebert *et al.* 2014; Richards *et al.* 2015). The actual proteome is approximately 2 to 3 orders of magnitudes more complex than can be predicted from its genome (Walsh, Garneau-Tsodikova and Gatto 2005), in which diversification processes such as post-translational modifications (PTMs) contribute substantially. PTMs are enzyme-mediated covalent modifications (or cleavage products) introduced following biosynthesis of the amino acid backbone, thereby expanding the range of possible protein isoforms (proteoforms) and functions without the need for changing the genetic code itself (Smith and Kelleher 2013). For example, the central carbon

Received: 1 September 2019; **Accepted:** 19 December 2019

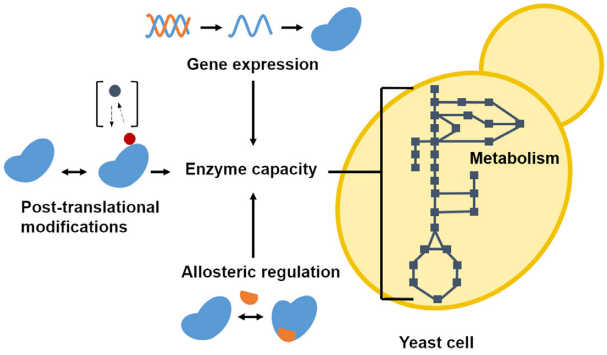

Figure 1. Overview of metabolic regulation mechanisms in a yeast cell. The metabolic flux in a cell depends on the capacities of the metabolic enzymes, which depend on enzyme abundances, allosteric regulation but also on the occurrence of covalent modifications on the protein itself, termed post-translational modifications (PTMs).

metabolism has been fine-tuned to exactly meet the requirements for building blocks and Gibbs free energy in conjunction with cell growth. Thus, when cells experience environmental changes, their metabolism immediately aims to adjust (Nielsen 2003; Daran-Lapujade *et al.* 2004, 2007). Apart from the relatively slow adjustment of enzyme abundances, cells utilise fast and dynamic routes such as allosteric regulation but also through the above-mentioned (reversible) protein side chain modifications of the enzyme itself (Daran-Lapujade *et al.* 2007; Gerosa and Sauer 2011) (Fig. 1). The number of different modification types reported across various species are in the range of hundreds (Creasy and Cottrell 2004; Oughtred *et al.* 2019), but most have been only observed at low frequency and low stoichiometry, and a potential biological relevance remains elusive (Swaney and Villén 2016).

PTMs observed for the yeast model *Saccharomyces cerevisiae* have recently been compiled in the database YAAM (Ledesma *et al.* 2018). Here, statistics on 12 experimentally confirmed modification types (ubiquitination, phosphorylation, acetylation, lipidation, oxidation, succinylation, glycosylation, methylation, sumoylation, nitration, disulfide bond formation and N-terminal acetylation) are described that are experimentally confirmed by MS, point mutation or functional evidence (Table 1). According to this database, more than 70% of the complete yeast proteome has been observed post-translationally modified (albeit under different growth and experimental conditions), demonstrating the importance of these regulators for cellular processes (Ledesma *et al.* 2018). Nevertheless, low frequent or difficult to analyse modifications, such as lipoylation (Rowland, Snowden and Cristea 2018; Baldi *et al.* 2019) are rarely addressed, of which some may still have been left unnoticed to date.

Protein phosphorylation appears to be the most commonly observed PTM in yeast (Table 1), however, this may be heavily overestimated since many studies specifically focus on this type of modification. A similar trend was observed in the Uniprot database entries for the proteins of *S. cerevisiae*, in which also phosphorylation appears to be the most frequent type of modification in yeast (The Uniprot Consortium 2017). In addition to phosphorylation, other common modifications such as acetylation and methylation have been frequently described and were found involved in a variety of cellular processes. For example, in yeast glycolysis, the shuttling back and forth of Hexokinase 2 between the nucleus and cytoplasm is regulated by phosphorylation (Fernańdez-García *et al.* 2012). Furthermore, studies over the past years have demonstrated that

Table 1. Number of PTMs and protein targets described for *S. cerevisiae* as found in the YAAM database (http://yaam.ifc.unam.mx, November 2019), where frequencis may include sites captured more than once (Ledesma *et al.* 2018).

Modification	# Mods	# Proteins
Acetylation	10052	1814
Disulfide	264	79
Glycosylation	1972	424
Lipidation	183	128
Methylation	287	143
Nitration	16	15
N-terminal Acetylation	762	687
Oxidation	875	605
Phosphorylation	87739	3955
Succinylation	1754	570
Sumoylation	138	48
Ubiquitination	14883	2355
Coverage		
Estimated total modified proteins		4759
Modified proteins of the proteome (%)		approx. 70

phosphorylation seems to affect many more processes, such as cell signalling (Fresques *et al.* 2015), glycerol metabolism (Nakic *et al.* 2016), regulation of nucleotide and amino acid biosynthesis (Oliveira *et al.* 2015), regulation of the outgrowth of autophagosomal membranes in autophagy (Papinski *et al.* 2014) and DNA damage checkpoint signalling (Memisoglu *et al.* 2019). Moreover, lysine acetylation (and glutamine methylation) is an evolutionarily highly conserved modification, which regulates chromatin accessibility and, therefore, affects gene expression directly (Allfrey, Faulkner and Mirsky 1964; Tessarz *et al.* 2014; Lawrence *et al.* 2017). Recent studies on PTMs in yeast are summarised in Supplementary Table 1 (Supporting Information). For a more extensive reviews on post-translational modifications in yeast, in particular for phosphorylation, we refer readers to recent reviews from Oliveira and Sauer (2012), Tripodi *et al.* (2015) and Chen and Nielsen (2016). Nevertheless, the functionality of many modification sites and (unknown) modifications remain to be investigated to date.

In this review, recent advancements in proteomics for unrestricted discovery of protein modifications are summarised, as data analysis remains the major challenge in identifying modifications. Moreover, we discuss the critical trade-off between maximum proteome coverage and maximum sequence coverage in the context of a comprehensive functional characterisation of PTMs.

ADVANCES IN BIOINFORMATICS TOOLS FOR UNRESTRICTED PTM DISCOVERY

Currently, the most frequently employed approach in discovery proteomics is referred to as shotgun proteomics (Fig. 2). Thereby, (in a bottom-up approach) protein extracts are analysed following proteolytic digestion using liquid chromatography (LC) coupled to tandem mass spectrometry (MS/MS) (Aebersold and Mann 2003). After chromatographic separation of the proteolytic digest, peptides are analysed for (accurate) mass that further triggers automatic fragmentation to obtain peptide sequence information (Yates *et al.* 1993; Mann and Wilm 1994; Washburn, Wolters and Yates 2001; Zhang *et al.* 2013). The spectra are then matched against a predefined protein database, generally derived from public repositories or genome sequencing

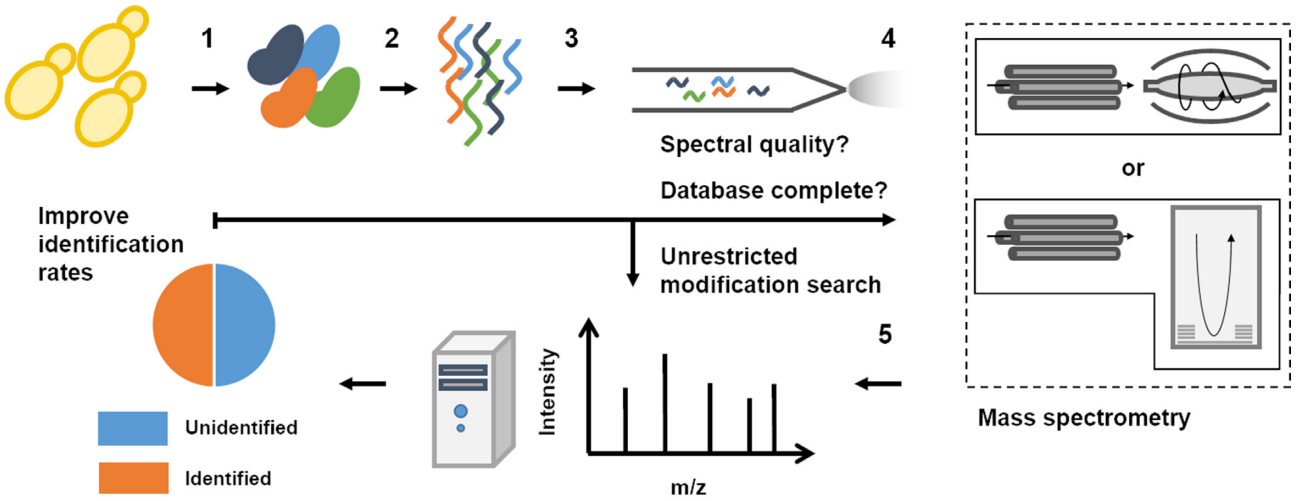

Figure 2. Protein identification by (a bottom-up) shot-gun proteomics experiment. The typical workflow consists of cell lysis, protein extraction from (yeast) cells and subsequent digestion into peptides using specific proteases (step 2). In step 3, the peptides are separated by liquid chromatography and further detected following electrospray ionisation (ESI) by data depended analysis, which automatically collects fragmentation spectra from (top intense) peptide signals, in step 4. In step 5, the fragmentation spectra (MS/MS) are commonly identified by database search approaches, matching fragmentation spectra to in-silico peptides derived from the target proteome database. Finally, the result of a typical shotgun proteomics experiment is depicted, in which roughly half of the spectra could not be assigned to a target protein sequence of the database. Similar identification rates have been reported in a survey reinvestigating hundreds of shot-gun proteomics experiments by Griss *et al.* (2016). To improve the identification rate, several approaches can be taken, which however should always start with confirming the quality of the aquired spectra. Further, modifications could be added to the database to increase database coverage. Finally, an unrestricted modification search could be performed to identify unexpected modifications.

itself. In common, cellular shotgun proteomics experiments create enormous amounts of sequencing spectra. For example, as much as 80 000 MS/MS spectra could be obtained for an 1 hour yeast proteome experiment (Hebert *et al.* 2014).

State-of-the-art proteomics workhorses, such as the quadrupole Orbitrap mass spectrometer (Zubarev and Makarov 2013), are capable of acquiring spectra at high speed (such as >20 Hz), high resolving power and high mass accuracy fragmentation spectra, supporting the identification of several thousands of proteins in less than 1 hour of analysis time (Olsen *et al.* 2005; Kelstrup *et al.* 2018). Nevertheless, only a fraction of fragmentation spectra (on average less than half) is confidentially matched to the proteome database (Fig. 2) (Griss *et al.* 2016). It is estimated that a good proportion of those unidentified, but high-quality spectra, may originate from unexpected (including not considered) modifications (or sequence variants) not present in the database (Chick *et al.* 2015; Griss *et al.* 2016). Where the majority of modified peptides may be readily detected by the employed method, their confident identification, however, often fails. Peptides containing modifications could, for example, not only be shared by different proteins (protein inference), but several overlapping modified peptides could also share the same modified site (Langella *et al.* 2017).

Identification is commonly realised by a database matching algorithm, where the acquired fragmentation spectra are matched against in silico spectra from predefined protein sequence databases (Steen and Mann 2004). Modifications are thereby detected as mass deviations from the native peptide mass peak, and that, in the ideal case, can be further allocated to single amino acids (Fig. 3A).

However, identification of PTMs by the simple addition of mass increments (modifications) is restricted. First, it assumes prior knowledge of the modifications present in a sample. Second, consideration of multiple modifications leads to an exponential increase in search space, impacting on computational efforts and challenging common statistical parameters, resulting in increased false negative as well as false positive identifications (Ahrné, Muller and Lisacek 2010; Na, Bandeira and Paek 2012; Ma and Lam 2014). To overcome limitations of database-restricted approaches, alternative algorithms have been established, with the most common tools and latest developments being listed in Table 2. These open modification search tools do not require specifying the modification before analysis and can, therefore, also identify unexpected modifications. Current bioinformatics tools use different strategies such as multi-round search, de-novo sequencing, sequence-tagging or spectral library-based approaches. PeaksPTM, for example, is a multi-round search tool incorporated into the PEAKS proteomics software solution (Ma *et al.* 2003), enabling efficient searches for all modifications listed in the UNIMOD database, simultaneously (Han *et al.* 2011; Bateman 2019).

Following identification of the proteins present in a sample, modifications are searched using a 'one-PTM-per-peptide' limitation to avoid exponential growth of the search space. During this process, sample-common PTM types are identified, generating a finite set of relevant modifications used to search for peptides containing two or more PTMs (Han *et al.* 2011). Another multi-round tool—G-PTM-D—employs a similar strategy, but uses a mass-tolerant open search strategy in the first round (Li *et al.* 2017). A further advanced version was published by Solntsev *et al.* (2018) that increases the speed and accuracy. Alternatively, de novo sequencing derives the peptide sequence from a tandem mass spectrum without a protein database. This strategy is used by deNovoPTM to identify modified peptides. However, algorithms are computationally demanding, consider only a restricted number of modifications and require high-quality fragmentation spectra (He, Han and Man 2013). A more efficient de-novo PTM identification workflow, termed Open-pNovo, has been published by Yang *et al.* (2017).

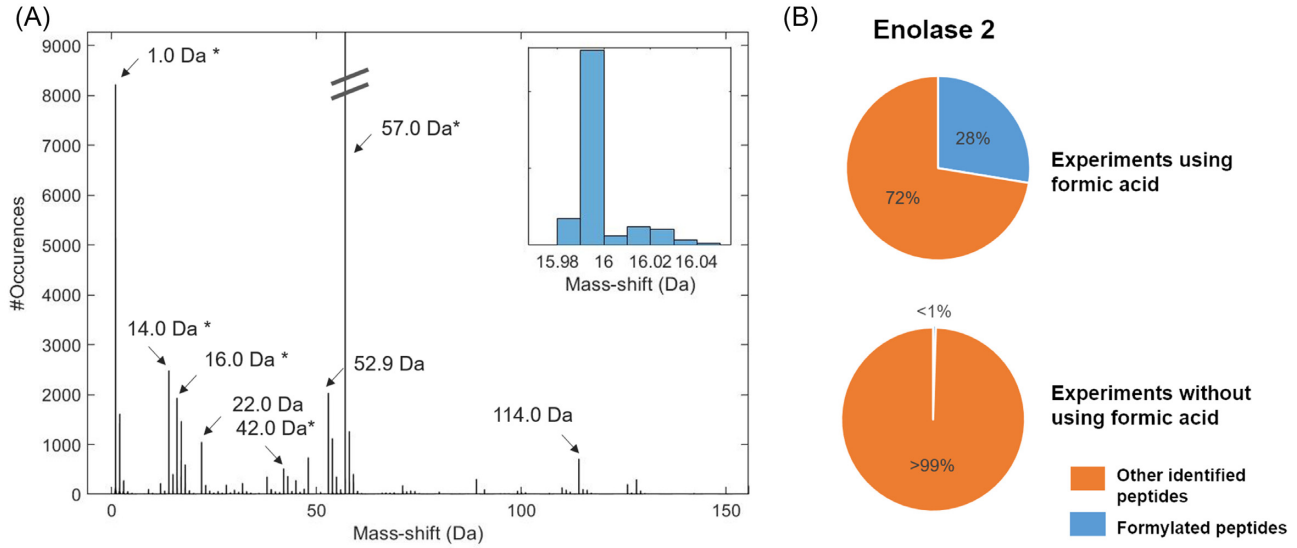

Figure 3. Peptide modification search. (**A**) Unrestricted modification search as obtained by using the Byonic software tool (Table 2). The observed mass shifts (between 0 and 150 Da) on tryptic peptides from a yeast shot-gun proteomics experiment are represented in a histogram graph. Commonly, many mass additions derive from artificial modifications introduced during sample preparation, while others are of natural origin. However, the distinction between both types is a delicate step during data processing and evaluation. Common mass shifts found are indicated by an arrow: deamidation (0.98 Da) ^{13}C isotope precursor selection (1.0 Da), methyl (14.0 Da), oxidation (15.98 Da), sodium adduct (22.0 Da), acetyl (42.0 Da), iron adduct (52.9 Da), carbamidomethyl (57.0 Da) and dicarbamidomethyl or ubiquitin (114.0 Da). Phosphorylation (80.0 Da) was observed only at low frequency in this search, even though it appears to be the most commonly observed PTM in yeast (Ledesma *et al.* 2018). However, this modification type usually occurs at low stoichiometry and, therefore, enrichment methods are commonly applied to inrease sensitivity, which was not performed in the experiment shown above. (*) Indicated mass shifts can also represent amino acid substitutions, which require case-by-case investigations. Inset: assignment of high-resolution mass shifts. Where high-resolution mass spectrometry can resolve 15.99 (oxidation) from a closely related addition at 16.02 Da. Other amino acid substitutions such as Ala→Ser show exactly the same composition and, therefore, mass shift. (**B**) The degree of formylation of Enolase 2 (*S. cerevisiae*) is quantified following a common shotgun proteomics experimental set-up with or without the use of formic acid as a solvent for protein solubilisation. The use of formic acid during sample handling introduced a considerable number of formylated peptide artefacts, as shown for Enolase 2.

Table 2. Recently developed open modification search tools for PTM discovery.

Software tool	(Main) Approach[a]	Fully unrestricted?	Download link	Reference
MSFragger	Error tolerant search	Yes	http://www.nesvilab.org/software	Kong *et al.* (2017)
DeNovoPTM	De novo sequencing	No	http://www.mybiosoftware.com/denovoptm-ms-based-peptide-identification-software-tool.html	He, Han and Man (2013)
Open-pNovo	De novo sequencing	No	http://pfind.ict.ac.cn/software/pNovo/index.html	Yang *et al.* (2017)
TagGraph	De novo sequencing/error tolerant search	Yes	http://sourceforge.net/projects/taggraph	Devabhaktuni *et al.* (2019)
ANN-Solo	Spectral library/error tolerant search	Yes	https://github.com/bittremieux/ANN-SoLo	Bittremieux *et al.* (2018)
SpecOMS	Spectral library/error tolerant search	Yes	https://github.com/matthieu-david/SpecOMS	David *et al.* (2017)
G-PTM-D	Multi-round/error tolerant search	Yes	https://github.com/smith-chem-wisc/gptmd	Li *et al.* (2017)
MetaMorpheus	Multi-round search	No	https://github.com/smith-chem-wisc/MetaMorpheus	Solntsev *et al.* (2018)
PeaksPTM[b]	Multi-round search	No	http://bioinfor.net/ptm	Han *et al.* (2011)
Byonic[b]	Various types, including sequence tagging/error tolerant search	Yes	www.proteinmetrics.com/products/byonic/	Bern, Kil and Becker (2012)
MODa	Sequence tagging/error tolerant search	Yes	https://omictools.com/moda-2-tool	Na, Bandeira and Paek (2012)
Open-pFind	Sequence tagging/multi-round/error tolerant	Yes	http://pfind.ict.ac.cn/software/pFind3	Chi *et al.* (2018)
PIPI	Sequence tagging/error tolerant search	Yes	http://bioinformatics.ust.hk/pipi.html	Yu, Li and Yu (2016)

[a]Many tools utilise hybrid-type approaches.
[b]Commercial platform(s).

The first developed unrestricted search tools included sequence-tag based approaches (Mann and Wilm 1994; Tabb, Saraf and Yates 2003; Tanner *et al.* 2005). Here, only de novo-sequence tags (extracted peptide sequence fragments of three to four amino acids) are required to find possible peptide matches from a sequence database. The differences between the expected and observed mass of the match are then assumed to be mutations or modifications. Sequence-tag-based approaches were then also the baseline for faster, further developed tools such as MODa, a 'multi-blind' spectral alignment algorithm (Na, Bandeira and Paek 2012). Many software tools further employ hybrid approaches to improve accuracy and speed. Open-pFind, for example, uses first a sequence-tag based approach followed by a restricted search in which modification types and protein sequence entries are set by semi-supervised machine learning (Chi *et al.* 2018). Kong *et al.* (2017) established a novel fragment-ion indexing method implemented in a database search tool termed MSFragger. This tool provided a substantial improvement in search speed and made unrestricted modification searches feasible for particularly large data sets. More recently, TagGraph was established by Devabhaktuni *et al.* (2019), which is an unrestricted de novo sequence-tag approach utilising a fast string-based search including a probabilistic validation model optimised for PTM assignments. Another commercial software package that enables advanced modification searches is Byonic (Bern, Kil and Becker 2012). This tool includes an option for Modification Fine Control that is also a fully unrestricted search approach for unanticipated modifications, termed Wildcard Search (Fig. 3A). An alternative peptide modification search strategy utilises spectral libraries. Thereby, identification of modified peptides are interpolated from the identification of unmodified reference spectra (Frewen *et al.* 2006; Lam *et al.* 2007; Zhang *et al.* 2011; Ma and Lam 2014; Griss 2016). This process results in improved accuracy and higher identification rates but is limited to peptides being present in the database. More recent developments here include the ANN-SoLo tool from Bittremieux *et al.* (2018) and the SpecOMS tool developed by David *et al.* (2017). It should be noted that most of the recently developed algorithms have not been applied in microbial or yeast proteomics to any significant degree, potentially due to difficulties in obtaining good false discovery rate (FDR) estimates and the risk of extensive false positive identifications (Fu and Qian. 2014).

SAMPLE PREPARATION ARTEFACT OR NATURAL PTM?

Whenever a peptide modification is detected, analysts have to make a decision regarding whether the observed modification is a genuine proteoform variant or whether it is an artefact introduced during the experiment. However, this is sometimes a very delicate process. Formylation, for example, is a natural histone modification (Jiang *et al.* 2007), which however can also be introduced during sample preparation when using formic acid-containing buffers to increase the solubility of hydrophobic peptides and aggregates (Fig. 3B) (Zheng and Doucette 2016). The same holds for carbamylation, which is frequently introduced when using buffers containing (high-molarity) urea (Kollipara and Zahedi 2013), or unspecific alkylation reactions introduced by extensive iodoacetamide treatment, broadly used during sulfhydryl alkylation reactions (Boja and Fales 2001; Müller and Winter 2017). Furthermore, chemically labile amino acid residues may undergo oxidation, deamidation, pyroglutamate formation, dehydration or metal ion adduct formation (Liu *et al.* 2005; Hao *et al.* 2011; Purwaha *et al.* 2014). Many of those chemically introduced modifications may, however, occur naturally in protein ageing processes and are, therefore, difficult to discriminate from sample preparation artefacts regardless of the method employed (Stadtman *et al.* 2005). In addition to preserving the native state of a peptide during sample preparation, chromatographic and ionisation properties during the analysis process also need to be chosen thoughtfully, particularly for modifications such as phosphorylation, which undergo rapid enzymatic hydrolysis and in-source fragmentation leading to neutral loss (Carapito *et al.* 2009). Where otherwise comparable protocols are used for cell lysis, protein extraction and proteolytic digestion compared to conventional discovery proteomics experiments, the analysis and interpretation of post-translational modifications requires additional careful considerations (Olsen and Mann 2013). Harsh conditions not only produce ambiguous identifications, but also induce mass spectrometric signal multiplications, reducing the discovery rates for both native and naturally modified peptides (Herbert *et al.* 2003; Grassl *et al.* 2009; Kollipara and Zahedi 2013).

MAXIMISE PROTEOME- OR SEQUENCE COVERAGE?

The proteome-wide discovery of post-translational modifications is challenged by factors such as sub-stoichiometric occurrence, competitive ionisation (sensitivity), computational limitations and lack of effective validation strategies. Thus, when a proteome-wide analysis of a certain type of modification is performed, peptide fractionation techniques are employed to reduce the number of signals and, therefore, increase sensitivity. Suitable methods are based on any physicochemical properties such as size, charge or hydrophilicity but also on affinity (Li *et al.* 2007; Han *et al.* 2008). Alternatively, targeted mass pre-selected subset of modifications has also been performed, but typically only for quantification purposes rather than discovery (Adachi, Narumi and Tomonaga 2016; Soares and Blackburn 2016; Arsova, Watt and Usadel 2018). For tools supporting quantitation, we refer to a recent review from Allmer (2012). However, over recent decades, it has become apparent that most proteins are modified by more than one modification at a time, and many modifications do not function in an isolated manner but seem instead to interact with modification sites from the same or other proteins, a process referred to as PTM cross-talk (Minguez *et al.* 2012; Guan *et al.* 2013; Swaney *et al.* 2013; Venne, Kollipara and Zahedi 2014; Huang *et al.* 2015). This process, however, is changing the view on how protein modifications are ideally investigated. For example, for the sake of improving sensitivity, an enrichment for a specific modification increases coverage for this particular modification tremendously, but it may abolish information essential for understanding a complete process.

Hence, to explore functional aspects and interactions, the analysis should aim to maximise protein sequence coverage of related pathways. In bottom-up proteomics, trypsin is predominantly used for digestion of the proteome due to its high specificity and ease of use (Tsiatsiani and Heck 2015). However, full sequence coverage is almost never achieved because digestion also generates peptides with sub-optimal length for MS detection (Swaney, Wenger and Coon 2010). To increase sequence coverage, multi-proteolytic digestion approaches have been proposed (Henriksen *et al.* 2012; Giansanti *et al.* 2016). Here, the proteome is subjected to digestion with multiple proteases

in parallel, resulting in complementary parts of the protein sequence and thus higher sequence coverage. The use of Trypsin in combination with LysC has, therefore, become common practice in the field of shotgun proteomics. Even though both proteases share lysine as a cleavage site, LysC/Trypsin digests were found most efficient to yield fully cleaved peptides (Glatter *et al.* 2012). Furthermore, the use of four alternative proteases (LysC, ArgC, AspN and GluC) in addition to trypsin led to nearly a three-fold improvement of sequence coverage for proteins at low abundances in yeast (Swaney, Wenger and Coon 2010).

In this context, prediction tools have been developed to support experiments designing a full protein sequence coverage. PTMselect is an example of such an open-source software tool, which simulates multi-enzyme digestion to tailor the optimal set of proteases for the discovery of global or targeted modification from any single or multiple proteins (Perchey *et al.* 2019). This approach allows sequence coverage to be achieved; however, it does not solve the sensitivity issues.

On the other hand, sensitivity for labile or very large modifications, such as phosphorylation and glycosylation (Grünwald-Gruber and Altmann 2019), could be increased using alternative fragmentation techniques such as electron transfer dissociation (ETD) (Wu *et al.* 2007; Elviri 2012; Wuhrer 2013).

Finally, using alternative bottom-up MS technologies to the commonly employed data-dependent acquisition (DDA), could lead to increased detection of modified peptides. PTMs are generally observed at low stoichiometry and, therefore, not selected for fragmentation using a data-dependent acquisition approach, in which only ions with highest intensity are chosen. To overcome this stochastic precursor ion selection, DIA methods could be employed (Röst *et al.* 2014). Here, all precursor ions are systematically fragmented in predefined retention time and precursor ion mass to charge (m/z) range. However, proper data analysis software tools should be utilised to correctly identify, localise and quantify the modifications (Meyer *et al.* 2017; Rosenberger *et al.* 2017). Moreover, these tools are important for discrimination of co-isolated modified peptide isoforms resulting from large precursor isolation windows (Rosenberger *et al.* 2017).

OUTLOOK

Research on post-translational modifications has advanced the understanding of protein phosphorylation in metabolic flux control and the understanding of modification cross-talk in yeast (Zahedi 2016; Chen, Wang and Nielsen 2017). However, many of the latest developments for the analysis, discovery and quantification of larger sets of post-translational modifications still challenge the field. Because many proteins undergo more than one modification at a time, a comprehensive exploration will require an examination beyond the most commonly investigated modifications such as phosphorylation and acetylation, by a simultaneous increase in protein sequence coverage.

A recent study on fission yeast by Telekawa *et al.* (2018) demonstrated the comprehensive characterisation of a protein complex following affinity purification. This work provided particularly high sequence coverage and gave insight on almost 40 modification sites of 3 different types of modifications within 1 complex. A similar study was performed by Šoštarić *et al.* (2019), who demonstrated the impact of acetylation and phosphorylation on subunit interaction in 3 large yeast complexes.

Considering that many modifications can influence binding affinities, modifications are often considered to be functionally associated (Duan and Walther 2015). A phosphoproteomics study in yeast illustrated that phosphorylated proteins engage in many more protein–protein interactions than their unmodified counterparts (Yachie *et al.* 2011). A better understanding of the impact of modifications on protein complex formation and on protein–protein or enzyme–substrate interactions may open effective intervention points and targets for engineering.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSYR online.

ACKNOWLEDGEMENTS

The authors are grateful to valuable discussions with our colleagues from the department of Biotechnology and acknowledge Carol de Ram for support in yeast proteomics.

FUNDING

This work was supported by a TU Delft StartUp fund.

Conflict of interest. None declared.

REFERENCES

Adachi J, Narumi R, Tomonaga T. Targeted phosphoproteome analysis using selected/multiple reaction monitoring (SRM/MRM). In: Reinders J (ed). *Proteomics in Systems Biology: Methods and Protocols*. New York, NY: Springer New York, 2016, 87–100.

Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;**422**:198–207.

Ahrné E, Muller M, Lisacek F. Unrestricted identification of modified proteins using MS/MS. *Proteomics* 2010;**10**:671–86.

Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci USA* 1964;**51**:786–94.

Allmer J. Existing bioinformatics tools for the quantitation of post-translational modifications. *Amino Acids* 2012;**42**:129–38.

Arsova B, Watt M, Usadel B. Monitoring of plant protein post-translational modifications using targeted proteomics. *Front Plant Sci* 2018;**9**:1–9.

Baldi N, Dykstra JC, Luttik MAH *et al.* Functional expression of a bacterial α-ketoglutarate dehydrogenase in the cytosol of *Saccharomyces cerevisiae*. *Metab Eng* 2019;**56**:190–7.

Bateman A. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;**47**:D506–15.

Bern M, Kil YJ, Becker C. Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics* 2012, DOI: 10.1002/0471250953.bi1320s40.

Bittremieux W, Meysman P, Noble WS *et al.* Fast open modification spectral library searching through approximate nearest neighbor indexing. *J Proteome Res* 2018;**17**:3463–74.

Boja ES, Fales HM. Overalkylation of a protein digest with iodoacetamide. *Anal Chem* 2001;**73**:3576–82.

Carapito C, Klemm C, Aebersold R *et al.* Systematic LC-MS analysis of labile post-translational modifications in complex mixtures. *J Proteome Res* 2009;**8**:2608–14.

Chen Y, Nielsen J. Flux control through protein phosphorylation in yeast. *FEMS Yeast Res* 2016;**16**:1–14.

Chen Y, Wang Y, Nielsen J. Systematic inference of functional phosphorylation events in yeast metabolism. *Bioinformatics* 2017;**33**:1995–2001.

Chi H, Liu C, Yang H *et al.* Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat Biotechnol* 2018, **36**, DOI: 10.1038/nbt.4236.

Chick JM, Kolippakkam D, Nusinow DP *et al.* A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat Biotechnol* 2015;**33**:743–9.

Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* 2011;**80**:273–99.

Creasy DM, Cottrell JS. Unimod: protein modifications for mass spectrometry. *Proteomics* 2004;**4**:1534–6.

Daran-Lapujade P, Jansen MLA, Daran JM *et al.* Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae*: a chemostat culture study. *J Biol Chem* 2004;**279**:9125–38.

Daran-Lapujade P, Rossell S, van Gulik WM *et al.* The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proc Natl Acad Sci* 2007;**104**:15753–8.

David M, Fertin G, Rogniaux H *et al.* SpecOMS: a full open modification search method performing all-to-all spectra comparisons within minutes. *J Proteome Res* 2017;**16**:3030–8.

Devabhaktuni A, Lin S, Zhang L *et al.* TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. *Nat Biotechnol* 2019;**37**:469–79.

Duan G, Walther D. The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput Biol* 2015;**11**:1–23.

Elviri L. ETD and ECD mass spectrometry fragmentation for the characterization of protein post translational modifications. *Tandem Mass Spectrometry – Applications and Principles* 2012, DOI: 10.5772/35277.

Fernańdez-García P, Peláez R, Herrero P *et al.* Phosphorylation of yeast hexokinase 2 regulates its nucleocytoplasmic shuttling. *J Biol Chem* 2012;**287**:42151–64.

Fresques T, Niles B, Aronova S *et al.* Regulation of ceramide synthase by casein kinase 2-dependent phosphorylation in *Saccharomyces cerevisiae*. *J Biol Chem* 2015;**290**:1395–403.

Frewen BE, Merrihew GE, Wu CC *et al.* Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal Chem* 2006;**78**:5678–84.

Fu Y, Qian X. Transferred subgroup false discovery rate for rare post-translational modifications detected by mass spectrometry. *Mol Cell Proteomics* 2014;**13**:1359–68.

Gerosa L, Sauer U. Regulation and control of metabolic fluxes in microbes. *Curr Opin Biotechnol* 2011;**22**:566–75.

Giansanti P, Tsiatsiani L, Low TY *et al.* Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* 2016;**11**:993–1006.

Glatter T, Ludwig C, Ahrné E *et al.* Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res* 2012;**11**:5145–56.

Grassl J, Westbrook JA, Robinson A *et al.* Preserving the yeast proteome from sample degradation. *Proteomics* 2009;**9**:4616–26.

Griss J. Spectral library searching in proteomics. *Proteomics* 2016;**16**:729–40.

Griss J, Perez-Riverol Y, Lewis S *et al.* Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nat Methods* 2016;**13**:651–6.

Grünwald-Gruber C, Altmann F. LC-MS analysis of (glyco-)proteins of *Pichia pastoris*. In: Gasser B, Mattanovich D . (eds). *Recombinant Protein Production in Yeast. Methods in Molecular Biology*. New York: Humana Press, 2019.

Guan X, Rastogi N, Parthun MR *et al.* Discovery of histone modification crosstalk networks by stable isotope labeling of amino acids in cell culture mass spectrometry (SILAC MS). *Mol Cell Proteomics* 2013;**12**:2048–59.

Han G, Ye M, Zhou H *et al.* Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics* 2008;**8**:1346–61.

Han X, He L, Xin L *et al.* PeaksPTM: mass spectrometry-based identification of peptides with unspecified modifications. *J Proteome Res* 2011;**10**:2930–6.

Hao P, Ren Y, Alpert AJ *et al.* Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Mol Cell Proteomics* 2011;**10**:O111.009381.

He L, Han X, Man B. De novo sequencing with limited number of post-translational modifications per peptide. *J Bioinform Comput Biol* 2013;**11**:1350007.

Hebert AS, Richards AL, Bailey DJ *et al.* The one hour yeast proteome. *Mol Cell Proteomics* 2014;**13**:339–47.

Henriksen P, Wagner SA, Weinert BT *et al.* Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in *Saccharomyces cerevisiae*. *Mol Cell Proteomics* 2012;**11**:1510–22.

Herbert B, Hopwood F, Oxley D *et al.* β-elimination: an unexpected artefact in proteome analysis. *Proteomics* 2003;**3**:826–31.

Huang Y, Xu B, Zhou X *et al.* Systematic characterization and prediction of post-translational modification cross-talk. *Mol Cell Proteomics* 2015;**14**:761–70.

Jiang T, Zhou X, Taghizadeh K *et al.* N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc Natl Acad Sci* 2007;**104**:60–5.

Kelstrup CD, Bekker-Jensen DB, Arrey TN *et al.* Performance evaluation of the Q Exactive HF-X for shotgun proteomics. *J Proteome Res* 2018;**17**:727–38.

Kollipara L, Zahedi RP. Protein carbamylation: in vivo modification or in vitro artefact? *Proteomics* 2013;**13**:941–4.

Kong AT, Leprevost F V, Avtonomov DM *et al.* MSFragger : ultrafast and comprehensive peptide identification in mass spectrometry – based proteomics. *Nat Methods* 2017;**14**:513–20.

Lam H, Deutsch EW, Eddes JS *et al.* Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* 2007;**7**:655–67.

Langella O, Valot B, Balliau T *et al.* X!TandemPipeline: a tool to manage sequence redundancy for protein inference and phosphosite identification. *J Proteome Res* 2017;**16**:494–503.

Lawrence MBD, Coutin N, Choi JK *et al.* Histone acetylation, not stoichiometry, regulates linker histone binding in *Saccharomyces cerevisiae*. *Genetics* 2017;**207**:347–55.

Ledesma L, Sandoval E, Cruz-martı U *et al.* YAAM: yeast amino acid modifications database. *Database* 2018;**2018**:1–8.

Li Q, Shortreed MR, Wenger CD *et al.* Global post-translational modification discovery. *J Proteome Res* 2017;**16**:1383–90.

Li X, Gerber SA, Rudner AD *et al.* Large-scale phosphorylation analysis of alpha-factor-arrested *Saccharomyces cerevisiae*. *J Proteome Res* 2007;**6**:1190–7.

Liu S, Zhang C, Campbell JL *et al.* Formation of phosphopeptide-metal ion complexes in liquid chromatography/electrospray

mass spectrometry and their influence on phosphopeptide detection. *Rapid Commun Mass Spectrom* 2005;**19**:2747–56.
Ma B, Zhang K, Hendrie C *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 2003;**17**:2337–42.
Ma CWM, Lam H. Hunting for unexpected post-translational modifications by spectral library searching with tier-wise scoring. *J Proteome Res* 2014;**13**:2262–71.
Mann M, Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* 1994;**66**:4390–9.
Memisoglu G, Lanz MC, Eapen VV *et al.* Mec1ATR autophosphorylation and Ddc2ATRIP phosphorylation regulates DNA damage checkpoint signaling. *Cell Rep* 2019;**28**:1090–102.
Meyer JG, Mukkamalla S, Steen H *et al.* PIQED: automated identification and quantification of protein modifications from DIA-MS data. *Nat Methods* 2017;**14**:646–7.
Minguez P, Parca L, Diella F *et al.* Deciphering a global network of functionally associated post-translational modifications. *Mol Syst Biol* 2012;**8**:1–14.
Müller T, Winter D. Systematic evaluation of protein reduction and alkylation reveals massive unspecific side effects by iodine-containing reagents. *Mol Cell Proteomics* 2017;**16**:1173–87.
Na S, Bandeira N, Paek E. Fast multi-blind modification search through tandem mass spectrometry. *Mol Cell Proteomics* 2012;**11**:M111.010199.
Nakic ZR, Seisenbacher G, Posas F *et al.* Untargeted metabolomics unravels functionalities of phosphorylation sites in *Saccharomyces cerevisiae*. *BMC Syst Biol* 2016;**10**:1–15.
Nielsen J. It is all about metabolic fluxes. *J Bacteriol* 2003;**185**:7031–5.
Oliveira AP, Ludwig C, Zampieri M *et al.* Lysine methylation modulates the protein-protein interactions of yeast cytochrome C Cyc1p. *PLoS Genet* 2015;**15**:1–15.
Oliveira AP, Sauer U. The importance of post-translational modifications in regulating *Saccharomyces cerevisiae* metabolism. *FEMS Yeast Res* 2012;**12**:104–17.
Olsen JV, Godoy LMF De, Li G *et al.* Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 2005;**4**:2010–21.
Olsen JV, Mann M. Status of large-scale analysis of post- translational modifications by mass spectrometry. *Mol Cell Proteomics* 2013;**12**:3444–52.
Oughtred R, Stark C, Breitkreutz BJ *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res* 2019;**47**:D529–41.
Papinski D, Schuschnig M, Reiter W *et al.* Early steps in autophagy depend on direct phosphorylation of Atg9 by the Atg1 kinase. *Mol Cell* 2014;**53**:471–83.
Perchey RT, Tonini L, Tosolini M *et al.* PTMselect: optimization of protein modifications discovery by mass spectrometry. *Sci Rep* 2019;**9**:5–11.
Purwaha P, Silva LP, Hawke DH *et al.* An artifact in LC-MS/MS measurement of glutamine and glutamic acid: in-source cyclization to pyroglutamic acid. *Anal Chem* 2014;**86**:5633–7.
Richards AL, Hebert AS, Ulbrich A *et al.* One-hour proteome analysis in yeast. *Nat Protoc* 2015;**10**:701–14.
Rosenberger G, Liu Y, Röst HL *et al.* Inference and quantification of peptidoforms in large sample cohorts by SWATH-MS. *Nat Biotechnol* 2017;**35**:781–8.
Röst HL, Rosenberger G, Navarro P *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* 2014;**32**:219–23.
Rowland EA, Snowden CK, Cristea IM. Protein lipoylation: an evolutionarily conserved metabolic regulator of health and disease. *Curr Opin Chem Biol* 2018;**42**:76–85.
Smith LM, Kelleher NL. Proteoform: a single term describing protein complexity. *Nat Methods* 2013;**10**:186–7.
Soares NC, Blackburn JM. Mass spectrometry targeted assays as a tool to improve our understanding of post-translational modifications in pathogenic bacteria. *Front Microbiol* 2016;**7**:1–5.
Solntsev SK, Shortreed MR, Frey BL *et al.* Enhanced global post-translational modification discovery with MetaMorpheus. *J Proteome Res* 2018;**17**:1844–51.
Šoštarić N, O FJ, Giansanti P *et al.* Effects of acetylation and phosphorylation on subunit interactions in three large eukaryotic complexes. *Mol Cell Proteomics* 2019;**10**:1–63.
Stadtman ER, Van Remmen H, Richardson A *et al.* Methionine oxidation and aging. *Biochim Biophys Acta - Proteins Proteomics* 2005;**1703**:135–40.
Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 2004;**5**:699–711.
Swaney DL, Beltrao P, Starita L *et al.* Global analysis of phosphorylation and ubiquitylation crosstalk in protein degradation. *Nat Methods* 2013;**10**:676–82.
Swaney DL, Villén J. Proteomic analysis of protein posttranslational modifications by mass spectrometry. *Cold Spring Harb Protoc* 2016;**2016**:207–10.
Swaney DL, Wenger CD, Coon JJ. Value of using multiple proteases for large-scale mass spectrometry-based proteomics research articles. *J Proteome Res* 2010;**9**:1323–9.
Tabb DL, Saraf A, Yates JR. GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem* 2003;**75**:6415–21.
Tanner S, Shu H, Frank A *et al.* InsPecT: identification of post-translationally modified peptides from tandem mass spectra. *Anal Chem* 2005;**77**:4626–39.
Telekawa C, Boisvert FM, Bachand F. Proteomic profiling and functional characterization of post-translational modifications of the fission yeast RNA exosome. *Nucleic Acids Res* 2018;**46**:11169–83.
Tessarz P, Santos-Rosa H, Robson SC *et al.* Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* 2014;**505**:564–8.
The Uniprot Consortium. Uniprot: the universal protein knowledgebase. *Nucleic Acids Res* 2017;**45**:D158–69.
Tripodi F, Nicastro R, Reghellin V *et al.* Post-translational modifications on yeast carbon metabolism: regulatory mechanisms beyond transcriptional control. *Biochim Biophys Acta - Gen Subj* 2015;**1850**:620–7.
Tsiatsiani L, Heck AJR. Proteomics beyond trypsin. *FEBS J* 2015;**282**:2612–26.
Venne AS, Kollipara L, Zahedi RP. The next level of complexity: crosstalk of posttranslational modifications. *Proteomics* 2014;**14**:513–24.
Walsh CT, Garneau-Tsodikova S, Gatto GJ. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chemie - Int Ed* 2005;**44**:7342–72.
Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;**19**:242–7.
Wu SL, Hühmer AFR, Hao Z *et al.* On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with

post-translational modifications. *J Proteome Res* 2007;**6**:4230–44.

Wuhrer M. Glycomics using mass spectrometry. *Glycoconj J* 2013;**30**:11–22.

Yachie N, Saito R, Sugiyama N *et al*. Integrative features of the yeast phosphoproteome and protein-protein interaction map. *PLoS Comput Biol* 2011;**7**:e1001064.

Yang H, Chi H, Zhou WJ *et al*. Open-pNovo: de novo peptide sequencing with thousands of protein modifications. *J Proteome Res* 2017;**16**:645–54.

Yates JR, Speicher S, Griffin PR *et al*. Peptide mass maps: a highly informative approach to protein identification. *Anal Biochem* 1993;**214**:397–408.

Yu F, Li N, Yu W. PIPI : PTM-invariant peptide identification using coding method. *J Proteome Res* 2016;**15**:4435.

Zahedi RP. Joining forces: studying multiple post-translational modifications to understand dynamic disease mechanisms. *Expert Rev Proteomics* 2016;**13**:1055–7.

Zhang X, Li Y, Shao W *et al*. Understanding the improved sensitivity of spectral library searching over sequence database searching in proteomics data analysis. *Proteomics* 2011;**11**:1075–85.

Zhang Y, Fonslow BR, Shan B *et al*. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* 2013;**113**:2343–94.

Zheng S, Doucette AA. Preventing N- and O-formylation of proteins when incubated in concentrated formic acid. *Proteomics* 2016;**16**:1059–68.

Zubarev RA, Makarov A. Orbitrap mass spectrometry. *Anal Chem* 2013;**85**:5288–96.