






Cite this: *Chem. Commun.*, 2023, 59, 11484

# Considerations for defining +80 Da mass shifts in mass spectrometry-based proteomics: phosphorylation and beyond

Leonard A. Daly,  Christopher J. Clarke,  Allen Po,  Sally O. Oswald  and Claire E. Eyers \*

Post-translational modifications (PTMs) are ubiquitous and key to regulating protein function. Understanding the dynamics of individual PTMs and their biological roles requires robust characterisation. Mass spectrometry (MS) is the method of choice for the identification and quantification of protein modifications. This article focusses on the MS-based analysis of those covalent modifications that induce a mass shift of +80 Da, notably phosphorylation and sulfation, given the challenges associated with their discrimination and pinpointing the sites of modification on a polypeptide chain. Phosphorylation in particular is highly abundant, dynamic and can occur on numerous residues to invoke specific functions, hence robust characterisation is crucial to understanding biological relevance. Showcasing our work in the context of other developments in the field, we highlight approaches for enrichment and site localisation of phosphorylated (canonical and non-canonical) and sulfated peptides, as well as modification analysis in the context of intact proteins (top down proteomics) to explore combinatorial roles. Finally, we discuss the application of native ion-mobility MS to explore the effect of these PTMs on protein structure and ligand binding.

Received 16th June 2023,  
Accepted 21st August 2023

DOI: 10.1039/d3cc02909c

[rsc.li/chemcomm](http://rsc.li/chemcomm)

## Introduction

Post-translational modification (PTM) is a term used to describe the processing of proteins post synthesis, either through cleavage, covalent attachment of a functional group, or modification of an amino acid side chain. Over 400 biological PTMs

*Centre for Proteome Research, Department of Biochemistry and Systems Biology, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK. E-mail: CEyers@liverpool.ac.uk*



**Leonard A. Daly**

*His research interests include using MS to interrogate signalling protein functions, whether it be at the level of PTM signalling or interactome analyses.*

*Leonard A. Daly received his undergraduate degree in Biochemistry (2015) and PhD (University of Liverpool, 2019) using mass spectrometry (MS) to interrogate the hypoxia-induced PTM landscape of specific transcription factors. As a postdoctoral research associate working with Prof. Claire Eyers at Liverpool, his work has focused on the development and implementation of novel MS-based approaches for studying under-appreciated PTMs. His*



**Christopher J. Clarke**

*Christopher J. Clarke obtained his PhD from the University of Liverpool in 2014, following an MRes programme at the University of Edinburgh (2009) and a degree in Medical Biochemistry from the University of Manchester prior to this (2008). He completed several postdoctoral positions before joining the lab of Prof. Eyers, where his interests lie in developing MS techniques to study and understand the biological roles of non-canonical phosphorylation.*



have so far been described,<sup>1</sup> with our knowledge of the diversity of protein modifications continuing to expand. PTMs serve to regulate protein function in a dynamic and often reversible manner and are typically enzyme-driven processes. Covalent PTMs, which will be the focus of this article, are highly diverse, ranging from appendment of relatively small chemical moieties (such as phosphate) to the attachment of complex, whole (poly)protein chains such as ubiquitin. Alone, or in combination, PTMs are essential for the regulation of protein function across all organisms, allowing a rapid physiological response to a change in environment, and negating the need for energetically demanding and comparatively slow *de novo* synthesis. PTMs also diversify the available functional proteome, generating millions of (potentially functionally different) protein species from a defined number of genes. While the human genome only encodes for ~21 000 genes, it has been estimated that in excess of 6 000 000 proteoforms (expressed gene products with different combinations of single nucleotide polymorphisms and/or PTMs) can be generated (Fig. 1).<sup>2–4</sup>

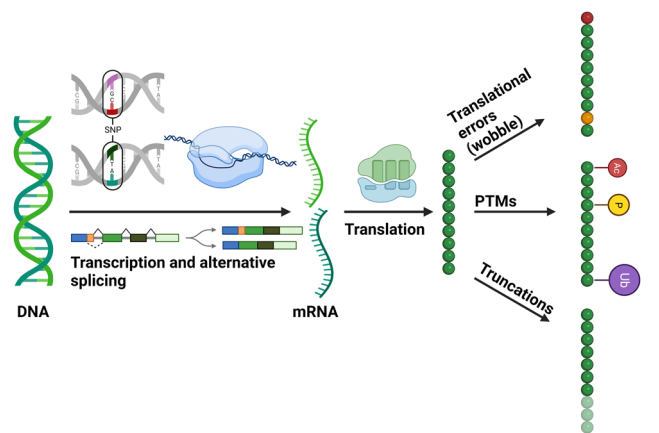
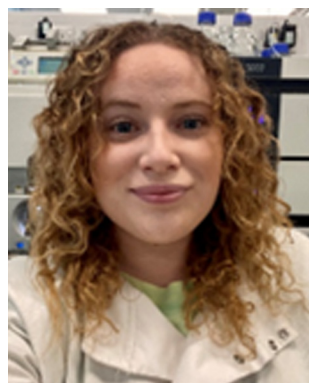


Fig. 1 Diagrammatic representation of cellular mechanisms that contribute to the generation of different proteoforms. Figure created with BioRender.com.



Allen Po

Allen Po received his BSc in Biochemistry/Chemistry from the University of California, San Diego in 2019. In 2020, he started his PhD with Prof. Claire Evers at the University of Liverpool. He is currently in the last year of his PhD developing top-down proteomics methods for the purpose of studying complex proteoforms and their populations.



Sally O. Oswald

Sally O. Oswald received her Cancer MRes at Newcastle University (2019). In early 2023 she obtained her PhD at the University of Liverpool working with Prof. Claire Evers focusing on neuroblastoma protein interactions and PTMs occurring in normoxia and hypoxia. She has stayed within the Evers group as a postdoctoral research associate. Her research interests include Cys PTMs and mass spectrometry applications to cancer research.



Claire E. Evers

Claire E. Evers completed her PhD at the University of Dundee (with Prof. Sir Philip Cohen, 2002) having graduated from the University of Bristol with a degree in Biochemistry (1998). After postdoctoral research at the University of Colorado, Boulder and the University of Manchester, she started her group at the University of Manchester in 2007. She is now Professor of Biological Mass Spectrometry and Director of the

Centre for Proteome Research at the University of Liverpool. Her research focusses on the development and application of mass spectrometry-based methods to explore the effects of protein post-translational modifications for which she was awarded the 2021 RSC Jeremy Knowles award.



typically requires validated site-specific antibodies or Edman degradation of a radiolabelled product) and benefits from being (semi)quantitative, thus allowing PTM dynamics in a changing environment to be evaluated. MS has thus revolutionised the interrogation of the proteome and PTM-mediated signalling cascades.

For the last 15 years, our group has developed MS-based analytical strategies for the identification and quantification of PTMs with a view to improving our understanding of signalling biology. In this feature article, we discuss some of our recent developments placing them in the broader context of the field, focussing in particular on those PTMs that induce a +80 Da mass shift – phosphorylation and sulfation. We also explore avenues for potential future investigation to improve our understanding of these covalent modifications.

## From single protein analysis to high-throughput characterisation

One of the most common PTMs is protein phosphorylation, namely the reversible and covalent addition of a phosphate group ( $\text{HPO}_3^-$ ) to the side chain of amino acids, predominantly (albeit not exclusively) the alcohol-containing side chain of Ser, Thr and Tyr residues.<sup>12–18</sup> Certainly, the vast majority of studies investigating the role of phosphorylation to date have focused on these three residues and have been key to advancing our understanding of the critical roles that this PTM plays in mammalian cell signalling (and their dysregulation in disease). There are a number of comprehensive reviews covering this topic<sup>19,20</sup> and will not be the primary focus of this article.

Our contributions to understanding the biological roles of specific protein phosphorylation (and other PTMs) events involve: characterising and exploring the effects of novel sites on numerous protein kinases, including the cell cycle-regulated enzymes Aurora A and PLK4,<sup>21–26</sup> quantifying dynamic site-specific phosphorylation with demonstrable effects on catalytic activity. We have also explored the roles of specific phosphosites on differential dimerization and DNA binding ability of transcription factors such as the NF- $\kappa$ B subunit RelA and hypoxia-inducible factor (HIF).<sup>27–29</sup> Of note, in-depth characterisation of proteins (by our group and others) inevitably reveals that a protein is typically decorated by multiple sites and types of PTM, often with numerous co-occurring sites of phosphorylation. Current estimates suggest that a third of human proteins are phosphorylated at any one time, and that ~90% of all proteins are capable of being phosphorylated at some point in their life-cycle. This is supported by the fact that 17.8k of the ~20k proteins in the human UniProt (reviewed) database are defined as being phosphorylated in PhosphoSite Plus (PSP).<sup>13,30–32</sup> Consequently, there has been a move away from focused investigations of specific proteins of interest to large-scale HTP phosphorylation site mapping to interrogate signalling networks.

The rise of HTP phosphopeptide characterisation studies (phosphoproteomics) over the last 20 years means that databases

such as PSP and Peptide Atlas (PA) are awash with ‘identified’ sites. However, variance in the confidence with which these phosphorylation sites have been identified means that many of these sites are unlikely to be real (either due to mislocalisation or poor spectral matching); while PSP currently lists over 130k unique phosphosites, PA details 164k sites, of which only 60% are shared between both databases.<sup>32</sup> Considering multiple observations in HTP studies, as well as orthogonal features such as evolutionary conservation, contextual information and functional enrichment analysis, we conservatively estimate that only ~82k of these are likely to be real,<sup>32</sup> hence databases of phosphorylation sites (and other PTMs) should be used with a degree of caution.

There is no doubt that we still have a long way to go to fully define the human phosphoproteome, let alone the functional significance or regulation of specific sites. Moreover, taking a wider perspective, deciphering phosphorylation sites in other species remains a substantial challenge. Beyond incomplete proteome databases, signalling in plants, fungi and prokaryotes is extensively regulated through phosphorylation of Asp and His residues as part of the two-component systems.<sup>33</sup> Numerous other amino acids can also be covalently modified by phosphate,<sup>16–18,34</sup> significantly expanding the search space complexity, and thus the rate of mis-localisation.

## The need for phosphopeptide enrichment

Protein phosphorylation, like other types of PTM, is seldom stoichiometric. Consequently, any analytical strategy looking to characterise PTMs must consider their comparatively low abundance compared with non-modified peptides generated from complex biological samples. Thus, methods have been developed to enrich a specific type of PTM, by invoking PTM-specific biochemical properties, prior to HTP MS analysis. In the case of phosphopeptides, the low  $\text{pK}_a$  value of phosphate (primary  $\text{pK}_a$  of phosphate monoesters: ~2) versus acidic aspartic acid and glutamic acid side chains ( $\text{pK}_a$ : ~3.5 and ~4.2 respectively)<sup>35</sup> is often used to facilitate the efficient and (relatively) specific enrichment of phosphopeptides by manipulating pH.

There are two major phosphopeptide enrichment strategies currently in play: Metal Oxide Affinity Chromatography (MOAC) and Immobilised Metal ion Affinity Chromatography (IMAC), which have similar, but fundamentally different, approaches for phosphate capture.<sup>36</sup> MOAC, most commonly titanium dioxide, relies on the specific spatial properties of phosphate moieties and efficient bidentate hydrogen bonding to immobilised titania for enrichment.<sup>37–40</sup> IMAC, most commonly titanium<sup>4+</sup> and iron<sup>3+</sup> (although many other metal ions have been investigated), relies on the much lower  $\text{pK}_a$  of phosphate (relative to Asp and Glu) for the specific electrostatic interaction with immobilised metal ions for phosphate capture, with strategies such as methyl esterification or the inclusion of glycolic acid having been used to improve specificity.<sup>21,41–43</sup>



Both phosphorylation enrichment strategies encompass their own select benefits and caveats, however, a detailed discussion is beyond the scope of this review and has been presented elsewhere.<sup>36,44</sup>

## HTP phosphorylation analysis is more than just peptide identification

To explore the (patho)physiological roles of specific phospho-sites, confidence in phosphorylation site localisation is essential. The ability to localise phosphorylation sites on peptides during HTP MS/MS has been evaluated using a variety of different fragmentation regimes and software tools. Table 1 provides an overview of the advantages and disadvantages of different fragmentation strategies for the identification of phosphorylated (and sulfated) peptides. A detailed overview of phosphopeptide fragmentation mechanisms is outside the scope of this article, but has been reviewed by Potel *et al.*<sup>45</sup> The most common peptide ion fragmentation regimes are collision-induced, being either resonant collisional-induced dissociation (performed in ion traps, referred to as resonant CID) and higher energy, non-resonant CID as performed in a quadrupole or octopole-based collision cell (referred to as (beam-type) CID or HCD (in the family of instruments from Thermo Scientific)). However, it is well documented that phosphopeptides subjected to (resonant) CID undergo extensive neutral loss of the phosphate group ( $\Delta$  80 amu) with/without

the additional loss of H<sub>2</sub>O ( $\Delta$  98 amu), depending on the phosphorylated residue,<sup>12,46–48</sup> thus hampering the ability to pinpoint the site of phosphorylation on the peptide backbone.<sup>49–55</sup> Inevitably, site localisation becomes increasingly problematic as the number of potential phosphorylation sites on a given peptide increases. Extensive investigation has revealed that a variety of factors influence the degree of phosphopeptide neutral loss, primarily revolving around the number, and relative position, of basic residues. Specifically, if the charge state of a phosphopeptide ion is greater than the number of basic residues (Arg, Lys, His) resulting in a ‘mobile proton environment’ (MPE), there is less neutral loss.<sup>46,47,53–57</sup> Contributing to the understanding of the MPE effect on phosphopeptide fragmentation by CID, we demonstrated that promoting a MPE through removal of the C-terminal Arg/Lys residue (which acts as a proton ‘sink’) from tryptic phosphopeptides with carboxypeptidase-B (CBP-B) markedly reduced neutral loss and increased peptide backbone fragmentation, consequently improving both the number and confidence in phosphopeptide identifications.<sup>58</sup> While this study was interesting from a mechanistic perspective, the subsequent reduction in overall peptide ion charge state (reducing protein coverage due to an increase in singly protonated species) combined with the wider accessibility of different fragmentation regimes now available means that this approach likely has limited utility for HTP phosphoproteomics. However, the use of CBP-B for removal of C-terminal basic residues could be interesting to explore in the context of bigger peptides generated

**Table 1** Advantages and disadvantages of different fragmentation strategies for the mass spectrometric analysis of phosphopeptides

Fragmentation type	Advantages	Disadvantages
Resonant CID	<ul style="list-style-type: none"> <li>Fast</li> <li>No internal product ions – simpler spectra</li> </ul>	<ul style="list-style-type: none"> <li>Extensive neutral loss compromises localisation of labile PTMs (phosphorylation, sulfation)</li> </ul>
Beam-type CID/HCD	<ul style="list-style-type: none"> <li>Fast</li> <li>Good sequence coverage</li> </ul>	<ul style="list-style-type: none"> <li>Neutral loss can compromise localisation of labile PTMs (phosphorylation, sulfation)</li> </ul>
ETD	<ul style="list-style-type: none"> <li>Retention of labile PTMs (including phosphate/sulfate)</li> <li>Beneficial for peptides containing multiple sites of phosphorylation/sulfation</li> </ul>	<ul style="list-style-type: none"> <li>Slow</li> <li>Charge state dependency (not useful for peptides of <math>z &lt; 3</math>)</li> <li>Charge reduction, as opposed to fragmentation/dissociation decreases sensitivity of product ion generation</li> <li>Supplemental activation required to generate useful product ions (and overcome ETnoD)</li> </ul>
EThcD	<ul style="list-style-type: none"> <li>ETD-generated product ions improve site localisation</li> <li>Overcomes the charge-state dependency of ETD</li> <li>Multiple fragment ion types increases confidence in identification and site localisation</li> </ul>	<ul style="list-style-type: none"> <li>Slower (than ETD)</li> <li>Statistical considerations for confident identification/localisation given the generation of multiple product ion types</li> </ul>
UVPD	<ul style="list-style-type: none"> <li>Potentially tuneable for optimal fragmentation of target peptides (on non-commercial instrumentation)</li> <li>Multiple fragment ion types increases confidence in identification and site localisation</li> </ul>	<ul style="list-style-type: none"> <li>Slowest</li> <li>Extensive neutral loss compromises localisation of labile PTMs (phosphorylation, sulfation)</li> <li>Poor efficiency of fragmentation</li> <li>Statistical considerations for confident identification/localisation given the generation of multiple product ion types</li> <li>Optimal fragmentation of modified peptides at 193 nm which is not available on commercial instrumentation</li> </ul>





with LysC/ArgC. A larger scale study looking at over 34k publicly available CID-generated MS/MS spectra suggests that for peptide ions with a MPE, neutral loss propensity inversely correlated with phosphate distance (3–8 residues) from the N-terminus, and was reduced for peptides with acetylated N-termini, suggesting direct involvement of the N-terminal amine group in the loss of H<sub>3</sub>PO<sub>4</sub>.<sup>59</sup> Interestingly, the presence of a proline within the peptide appeared to reduce neutral loss irrespective of its position relative to the site of phosphorylation; this is likely due to the decreased flexibility of Pro-containing peptides and thus their reduced ability to form cyclic structures that facilitate charge solvation/bring the phosphate group into proximity with the mobile proton.<sup>59</sup>

Electron-mediated fragmentation regimes (electron transfer dissociation – ETD; electron capture dissociation – ECD) have gained popularity in phosphoproteomics pipelines due to their ability to overcome the extensive neutral loss of HPO<sub>3</sub>/H<sub>3</sub>PO<sub>4</sub> typically observed with collisional dissociation.<sup>60</sup> These strategies are particularly useful for accurate phosphosite localisation in instances where there are multiple potential sites of modification and as the number of sites of modification on a peptide increases. In contrast to CID/HCD, electron-driven fragmentation is not dependent on the re-distribution of thermal energy, rather spontaneously occurring after electron transfer (being described as non-ergodic), allowing labile covalent PTMs to remain attached to the amino acid side chain. However, a major deficit of such fragmentation approaches is that ions of charge state  $\leq 2+$  (the predominant species for tryptic phosphopeptides) is poor due to non-dissociative charge reduction which compromises the ability to generate fragment ions, a prerequisite for peptide sequence determination and PTM localisation. Dual fragmentation approaches that apply supplemental collisional dissociation of ETD products (EThcD/ETciD) largely overcome the charge state-dependent fragmentation deficiency of ETD, producing rich tandem mass spectra which contain a mixture of product ion types (b/y & c/z).<sup>61,62</sup> These approaches have been highly useful for characterising labile PTMs, with multiple benchmarking studies (including from our group) having evaluated the utility of EThcD as the optimal method for correct phosphorylation site localisation.<sup>49,63,64</sup> Our analyses revealed that the benefits of EThcD for phosphopeptide identification and low probability of site mis-assignment is dependent on the resolution at which MS2 data are acquired, as well as the algorithm used for data interrogation.<sup>49</sup> Importantly, while we observed a notable increase in the proportion of correctly localised phosphorylation sites for a synthetic phosphopeptide library in low resolution (ion trap; IT) experiments (using an Orbitrap Fusion instrument; 92% for EThcD *versus* 83% for HCD), there was a marked reduction in the number of peptide spectral matches (PSMs) (654 *versus* 1497 respectively) under the same conditions, revealing a ‘double-edged sword’ regarding the utility of EThcD for HTP phosphoproteomics. This reduction in PSMs (and consequently peptide identification rate) is a known downside to ET(hc/ci)D arising from the increased duty cycle of this sequential fragmentation regime. In contrast, there was

**Table 2** Proportion (percentage) of phosphosites localised with high confidence (<1% false localisation rate) with different fragmentation strategies (EThcD or HCD) using either high resolution (OT) or low resolution (IT) MS/MS data acquisition of a TiO<sub>2</sub>-enriched U2OS cell lysate. Data were interrogated using either Andromeda/PTM-score or Mascot/ptmRS. Information was extracted from ref. 49

Search algorithm		EThcD	HCD
Orbitrap (OT)	Andromeda/PTM-score	76	50
	Mascot/ptmRS	86	76
Ion trap (IT)	Andromeda/PTM-score	76	37
	Mascot/ptmRS	83	60

minimal benefit in phosphosite localisation confidence rates between HCD and EThcD for this phosphopeptide library at high resolution (94% *versus* 96% respectively), although the reduction in PSMs remained substantial (889 *versus* 417 respectively). We originally hypothesised that the relatively limited utility of EThcD observed in this benchmarking study may in part be due to the composition of the peptide library (comprised of 175 unique phosphopeptides, 191 phosphorylation sites), where the vast majority were observed as 2+ ion species. We therefore interrogated a phosphopeptide-enriched cell lysate sample, showing that EThcD identified a substantially higher proportion of peptides with high confidence site localisation than HCD, irrespective of either the mass analyser used for MS2 or the search engine employed (Table 2).

However, as observed with the peptide library, the slower EThcD duty cycle significantly compromised the total number of phosphopeptides identified (5733 *versus* 2413 for HCD and EThcD (Orbitrap MS2) respectively; equating to a reduction of ~58%). Consequently, a higher number of phosphopeptides with high site localisation confidence (*ptmRS* score > 99.0) were identified overall with HCD than EThcD (optimally using the orbitrap for MS2; 4337 *versus* 2078 respectively, an increase of 109%). These observations are somewhat more nuanced for ions of charge state  $\geq 3+$  where the MS2 acquisition strategy and search engine employed influence the choice of HCD or EThcD for optimal identification of confidently localised phosphorylation sites.

In an attempt to bypass the increased time requirements of dual fragmentation approaches, we explored a neutral loss-triggered regime that exploits phosphopeptide-derived loss of either 97.9763 Da and/or 79.9799 Da from the precursor ion in an initial HCD scan to trigger a second MS2 using EThcD.<sup>49</sup> Theoretically, this approach benefits from the speed of HCD to define phosphopeptide precursors and minimises the effect of reduced scan rates on co-enriched unmodified peptides, while benefitting from the improved PTM-site localisation confidence observed with EThcD. However, upon implementation, we noticed a 12% reduction in the number of phosphopeptides identified using this neutral loss triggering approach, with only a small increase (3%) in the rate of correct PTM-site localisation *cf.* HCD alone (3873 *versus* 4337 phosphopeptides identified using HCD OT NL EThcD *versus* HCD respectively).<sup>49</sup> Contrary to our expectations, we observed that ~90% of confident



phosphosite assignments were actually derived from the initial HCD scan rather than the secondary EThcD scan, concluding that our neutral-loss triggered EThcD method provided no additional benefit for this type of cell extract-derived phosphopeptide sample.<sup>49</sup>

Ultraviolet photodissociation (UVPD) is an alternative fragmentation method that has been successfully employed for characterising PTM-containing peptides, using high energy lasers at specific wavelengths to induce peptide fragmentation to generate all major types of product ions (a/x, b/y, c/z).<sup>65</sup> The major advantage of UVPD is its highly optimisable parameters, with custom instrumentation being able to modulate wavelength, frequency, energy and irradiation times, as required for the sequence and charge state needs of the analytes.<sup>65–70</sup> The Brodbelt and Heck groups have led developments in this area, demonstrating the benefits of phosphopeptide fragmentation with a (non-commercial) 193 nm UVPD laser. Their studies indicate higher phosphosite localisation confidence arising from the presence of ~20–45% more phosphate-retaining product ions (*i.e.* no neutral loss) compared with HCD. However, as seen with EThcD, the increase in site-localisation confidence was offset by a lower phosphopeptide identification rate (<20%).<sup>69,70</sup> Building on these investigations, we have investigated the effect of a commercial 213 nm UVPD set up (on the Fusion Lumos, Thermo Scientific) for fragmentation of a panel of phosphotyrosine (pTyr)-containing peptides. Our data broadly mirrored the phosphosite localisation confidence rates that were observed for the same analytes with HCD. However, we noted extremely poor fragmentation efficiency even with long irradiation times (~100 ms); UVPD fragmentation at 213 nm was thus deemed impractical for complex samples.<sup>71</sup> It also highlights the differences in spectral quality depending on laser set up.

The scientific interest in HTP phosphoproteomics studies has fuelled the development of multiple software packages for the interrogation of MS2 data that provide a measure of site localisation confidence, including *ptmRS*, AScore, PTM-score, LuciPhor and PTMprophet,<sup>13,72–75</sup> each of which uses a different statistical strategy to infer confidence in site assignment. Although developers of these algorithms might suggest that true benchmarking is hard to perform ‘fairly’, our group and others have indeed attempted to do this from a user perspective.<sup>49,72–74,76</sup> While there are caveats in terms of MS2 spectral resolution and fragmentation regime employed, in our hands *ptmRS*, in combination with the MASCOT search engine, overall provided the most confidently localised (and identified) phosphorylation sites when compared with AScore using either Andromeda or PEAKS as search engines.<sup>49,72</sup> Interestingly, the more recently developed PTMprophet appears to outperform our optimal Mascot/*ptmRS* search strategy with respect to the number of confidently localised PSMs when using the same synthetic phosphopeptide library to our studies.<sup>74</sup> A recent benchmarking study also encourages broad adoption of PTMprophet for phosphoproteomics data interrogation due to its ability to achieve higher sensitivity and quality.<sup>77</sup> An interesting aspect of all these tools, and one which

is consistently either not discussed or overlooked in almost all search algorithms, is the generation of chimeric spectra from isobaric phosphopeptide isoforms (where the same peptide sequence is phosphorylated on a different residue). Such peptides will likely co-elute and cannot be separated by *m/z*, resulting in co-isolation and generation of MS2 spectra that contain site determining ions from both species. This situation creates a paradoxical ‘coin toss’ scenario where evidence for confident localisation of both phosphosites within a single PSM results in statistical uncertainty. Consequently, site localisation algorithms typically report a 50% confidence for modification of each site, rather than reflecting the true 100% confidence that both sites are modified.

Incorporating ion mobility spectrometry (IMS) (drift tube IMS, FAIMS (high field asymmetric waveform IMS) or TIMS (trapped IMS)) in the LC-MS/MS phosphopeptide data acquisition pipeline has proven beneficial for the separation of such phosphopeptide isomers, improving their identification.<sup>78–80</sup> Indeed, the recent study by Oliinyk and Meier using dia-PASEF demonstrated a substantive improvement in the identification of positional phosphopeptide isomers by TIMS, reportedly resolving 86% of the positional isomer pairs.<sup>80</sup> Future improvement in IMS resolution, and routine incorporation of this conformation-based separation strategy into phosphoproteomics workflows could thus have a marked impact on the field.

## The challenges of non-canonical phosphorylation analysis

As discussed above, the field of phosphoproteomics has traditionally focused on the canonical phosphorylation of Ser, Thr and Tyr residues (referred to here as CPhos). However, another 6 amino acids can also be phosphorylated (referred to as non-canonical phosphorylation, NCPPhos) (Fig. 2). These can be grouped into three classes based on their phosphate bonding chemistry: phosphorothiolates – Cys, conjugated through an S–P bond; phosphoanhydrides – Asp and Glu, conjugated through the hydroxyl (OH) of the side chain carboxyl group; phosphoramidates – His, Lys and Arg, conjugated through an N–P bond.<sup>12,16–18</sup> Phosphohistidine (pHis) can also exist as two distinct phospho-isomers due to phosphorylation at either 1- or 3-position on the imidazole ring of the side chain (Fig. 2).<sup>17</sup> Although phosphorylation of tryptophan is also chemically feasible, there is currently no evidence of this having been observed in biologically derived material (either as a free amino acid, or in a protein).<sup>34</sup>

The biological relevance of NCPPhos in prokaryotes is well documented, notably phosphorylation of His and Asp which play key roles in two component signalling (TCS) relays that mediate intracellular responses to environmental changes through receptor His kinases (extensively reviewed in ref. 81). Although NCPPhos was first identified in mammalian cells over 60 years ago,<sup>82</sup> there has been relatively little investigation of these PTMs, hence their involvement in regulating mammalian signalling remains somewhat controversial. One reason for the



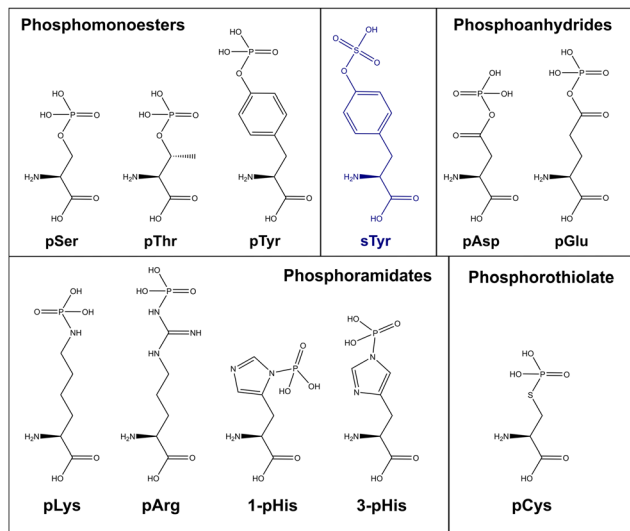


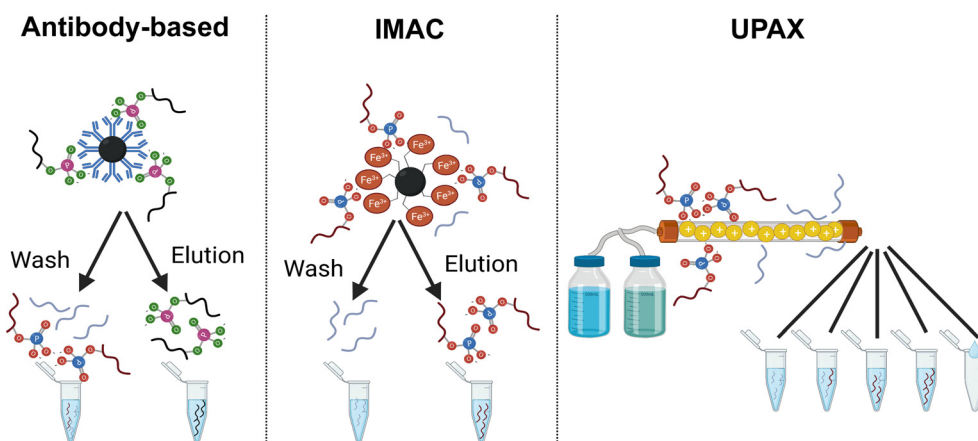
Fig. 2 Chemical structures of biologically relevant phosphorylated amino acids, and sulfotyrosine (sTyr). The atomic structure of all canonical (phosphomonoesters – Ser, Thr, Tyr) and non-canonical (grouped by chemical property); phosphoanhydride (pAsp, pGlu), phosphoramidates (pLys, pArg, 1-pHis, 3-pHis) and phosphorothiolates (pCys) alongside sulfotyrosine (sTyr). All models retrieved from Chemdraw 20.0.

limited information on NCPHos, despite the number and scale of CPhos datasets available, is that they are not generally deemed to be of relevance and are therefore not considered during analysis. Phosphoramidates and phosphoanhydrides are also much less stable than CPhos (particularly under low pH or high temperature conditions), compromising enrichment strategies and sensitivity of analysis.<sup>12,16–18</sup> Stability studies of pHis-containing peptides have shown that  $\text{pH} \leq \sim 2.5$

results in extensive hydrolysis, with  $t_{1/2}$  at pH 1 (conditions as used for *e.g.*,  $\text{TiO}_2$  enrichment) being around 15 min.<sup>12</sup> For perspective, phosphopeptide enrichment using standard protocols typically takes  $>1$  h, hence compromising the use of such methods for investigation of NCPHos. Consequently, three broad approaches for enrichment of NCPHos (or pHis) have been developed (Fig. 3): antibody based, a modified IMAC approach and, our approach – UPAX (Unbiased Phosphopeptide enrichment using strong Anion eXchange).<sup>12,83,84</sup>

Antibody-based immunoprecipitation (IP) is often preferred as they can be (i) highly selective and (ii) used at physiological pH, minimising NCPHos hydrolysis. However, their application in HTP studies relies on the generation of an antibody that recognises the NCPHos residue in a sequence-independent manner, with each NCPHos also requiring its own antibody. The groups of Muir and Hunter have developed generic antibodies suitable for IP of pHis, both as general reagents and for isoform specific recognition; a site-specific antibody has also been generated for pHis18 within histone H4.<sup>84–87</sup> Using their pHis isomer-specific antibodies, Fuhs *et al.* identified 786 pHis containing proteins from a human cell line extract using quantitative IP-based MS analysis. However, no specific pHis sites were defined.<sup>84</sup> No other NCPHos antibodies are currently available, hampering the broad utility of antibody-based strategies for NCPHos discovery.

Large-scale pHis analysis has also been achieved using a rapid (14 min)  $\text{Fe}^{3+}$ -IMAC-based enrichment strategy under relatively mild acidic conditions ( $\text{pH} = 2.3$ ), demonstrating the feasibility of such approaches.<sup>83,88,89</sup> This strategy successfully identified 276 sites of pHis on proteins from *E. coli*. However, application of this approach for the investigation of pHis sites from human cell extracts was met with relatively



Advantages	- Neutral pH - Potentially highly specific	- Phosphosite indiscriminate	- Neutral pH - Phosphosite indiscriminate
Disadvantages	- Few NCPHos antibodies available - NCPHos type-specific antibodies required	- Optimal enrichment requires acidic conditions which may not be conducive for optimal stability of some NCPHos	- Co-enrichment of acidic peptides (sample simplification versus enrichment)

Fig. 3 Strategies for the investigation of NCPHos. Table includes advantages and disadvantages of the three main enrichment strategies. Figure created in part with BioRender.com.



limited success: 83 acid-labile pHis sites were identified. Further filtering of the data based on the presence of a pHis immonium ion reduced this number to four.<sup>89</sup> This relatively stringent filtering criterion may explain the disparity with the numbers of mammalian pHis-containing proteins as reported by Fuhs *et al.*, and observed in studies from our group. Although Leijten *et al.* report that 15–25% of the pHis peptides enriched from *E. coli* display an intense pHis immonium ion (being present in 13% of their synthetic pHis peptides), our evaluation, both of synthetic pHis peptides and pHis peptides from human cell extracts, suggests that the pHis immonium ion at  $m/z$  190.04 is seldom observed with significant intensity.<sup>12,83,88,89</sup> These differences could be explained by the fact that the relative abundance of any pHis immonium ion is likely dependent on the position of the pHis within the peptide sequence, as well as the normalised collision energy used for HCD.

Building on the application of this  $\text{Fe}^{3+}$ -IMAC strategy for the enrichment of pHis-containing peptides, Lemeer and colleagues used this approach for large scale phosphoarginine (pArg) peptide enrichment from *Staphylococcus aureus*, permitting the identification of 470 ‘type I’ pArg sites (Andromeda localization probability > 0.75), the largest cohort of pArg sites defined to date in any species.<sup>90</sup> It is interesting that while this workflow could in theory be used for the interrogation of CPhos and NCPHos peptides (*i.e.*, considering all nine phosphorylated residues, assuming that acid labilities are similar), this is yet to be explored and it will be interesting to consider this approach in future.

Neutral pH IMAC-based systems, potentially capable of enriching all CPhos and NCPHos have also been reported. Whilst the utility of Phos-tag (an alkoxide-bridged dinuclear zinc(II) complex)<sup>91</sup> for NCPHos enrichment has not yet been investigated,  $\text{SiO}_2@DpaZn$  has recently been shown to enrich phosphoramidate-containing peptides (pArg, pLys, pHis),<sup>92</sup> albeit with co-enrichment of acidic peptides.

We have developed a strategy for the indiscriminate enrichment of NCPHos (and CPhos) along a similar vein, using (near) neutral pH to minimise phosphate hydrolysis. However, instead of using a bind/wash/elute bead-based approach, we developed a workflow exploiting strong anion exchange chromatography, termed UPAX.<sup>12</sup> While we also observe binding of acidic peptides with UPAX, chromatographic fractionation following salt gradient elution simplifies the components of each fraction for analysis, compensating for the presence of these ‘unwanted’ peptides.

Somewhat unsurprisingly, we (and others) have shown that collision-induced fragmentation techniques (CID/HCD) induce extensive neutral loss from NCPHos-containing peptides, compromising the ability to generate site determining ions. To increase confidence in site localisation, or at least the type of amino acid to which the phosphate group is covalently bound, there have been a number of studies seeking to define residue-specific product ion features. Oslund *et al.* showed that pHis peptides subjected to resonant CID in a linear ion trap yielded predominant neutral loss of 98 amu from the precursor ion,

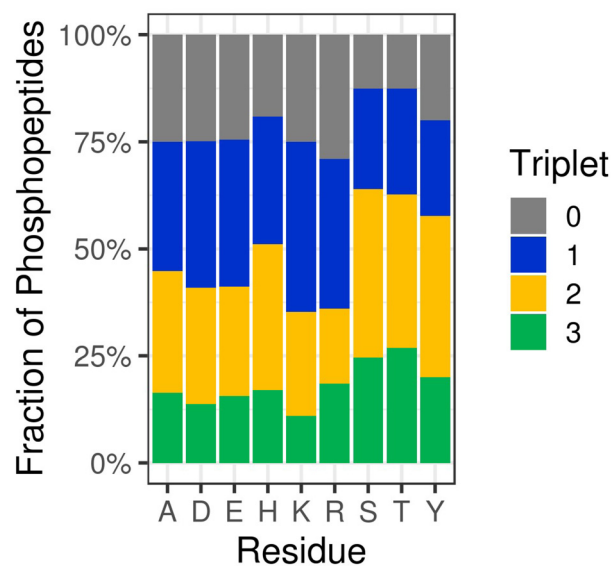


Fig. 4 HCD-induced triplet neutral loss from different types of phosphopeptides. Percentage of phosphopeptides exhibiting no neutral loss (triplet = 0) or neutral loss of any combination of 80, 98 and/or 116 Da (Triplet = 1, 2 or 3) as a function of the defined site of phosphorylation following HCD ( $ptmRS > 0.90$ ). Ala (A) was used as a decoy for site localisation analysis. Figure adapted from ref. 12 with permission. Copyright © 2019 Hardman *et al.* Published by EMBO Press, under a CC-BY 4.0 license.

and to a lesser extent, loss of either 80 amu or 116 amu (equating to loss of the phosphate group alone or with up to two water molecules), at the expense of backbone fragmentation.<sup>93</sup> Whilst peptide identification thus required multi-step activation (MSA), they were able to use this ‘triplet’ neutral loss feature ( $\Delta 80/98/116$  amu) to discriminate pHis-containing peptides from those phosphorylated on Ser, Thr or Tyr. To avoid the need for MSA for identification and explore whether this triplet neutral loss pattern was observed for other NCPHos, we quantified HCD-induced precursor neutral loss across our UPAX-generated NCPHos dataset ( $ptmRS \geq 0.90$ ).<sup>12</sup> Contrary to the strong triplet loss pattern observed with CID, we found that this neutral loss pattern was observed in less than 20% of pHis HCD spectra, with a similar proportion of spectra not showing any precursor ion neutral loss (Fig. 4). Of greater concern, spectra annotated as containing either CPhos or other NCPHos also yielded extensive triplet loss, indicating that for HCD, loss of 80, 98 and 116 amu from the precursor is not sufficient to define a phosphopeptide as being modified on His. As previously discussed, there is some disagreement in the field as to the prevalence of a pHis immonium ion that could be used for discrimination and positive identification of pHis. Although an in-depth analysis of immonium ion generation from NCPHos generally has yet to be performed, there are a small number of studies showing that immonium ions appear to be generated for pLys (at  $m/z$  164.047 or  $m/z$  181.074)<sup>64,94</sup> and pArg (at  $m/z$  237.0747 or  $m/z$  209.0798), albeit with relatively low intensity and in an inconsistent manner.<sup>90</sup> Hence, the use of immonium ions for CPhos or NCPHos discrimination currently





remains controversial (possibly with the exception of pTyr<sup>95</sup>). EThcD thus remains the fragmentation method of choice for NCPHos identification, where a slight hit in sensitivity is compensated for by increased confidence in the nature (site and type) of the phosphopeptide identified.<sup>12,64</sup>

An important aspect when studying variable PTMs by MS-based proteomics and is particularly evident when considering nine different potential sites of phosphorylation, is the combinatorial expansion of search space as the number of PTMs under consideration is increased. This not only increases the required computing power but crucially, reduces the statistical power for peptide identification and PTM site localisation. To better understand the effects of increased search space for phosphosite localisation when considering CPhos and NCPHos, we developed a false localisation rate (FLR) estimation strategy, using a non-permissible phosphorylated amino acid (phosphoalanine, pAla) to define the rate of 'random' assignment of a phosphate group for each phosphorylated residue.<sup>12,96</sup> This concept of using decoy amino acids to estimate site localisation confidence has since been expanded in a study that evaluated six different amino acids for their ability to act as suitable decoys for determining site localisation confidence. Using a variety of synthetic and complex biological sample-derived datasets, Ramsbottom *et al.* came to the conclusion that Ala (or potentially Leu) are the most appropriate decoys for reliable estimation of phosphosite FLR.<sup>96</sup>

Combining our UPAX technique for NCPHos (and CPhos) peptide enrichment from a U2OS osteosarcoma cell line extract with a neutral loss-triggered EThcD fragmentation strategy and our pAla decoy search approach, we were able to show that NCPHos is significantly more prevalent in mammalian cells than previously hypothesised, even when accounting for a relatively high (amino acid-dependent) false localisation rate.<sup>12</sup> We identified approximately double the number of pAsp residues than seen for pTyr (the least frequent CPhos site) under the same conditions, with the frequency of pHis, pGlu, pLys and pAsp being marginally lower than pTyr. Far fewer sites of pCys were identified in this dataset. We have since validated one of the pCys sites identified in this HTP analysis, demonstrating that pCys412 on human PINK1 appears to be involved in regulating PINK1-mediated phosphorylation of ubiquitin.<sup>97</sup> In a separate study, we have also shown that pCys90 on the transcription factor hypoxia-inducible factor (HIF)-1 $\alpha$  is induced in hypoxia, hypothesising a role for this PTM in heterodimer formation.<sup>27</sup> These data, combined with reports of a role for pCys as a Mg<sup>2+</sup> sensor in PRL (phosphatases of regenerating liver) proteins<sup>98</sup> speak to growing evidence of a role for pCys as a cell signalling regulatory element, beyond being 'just' a catalytic intermediate (as observed in *e.g.*, protein tyrosine phosphatase (PTP) 1B).<sup>99</sup>

A final point to note in considering confidence of identification and potential roles of CPhos and NCPHos is phosphate transfer; two component systems are known to operate through a phosphate relay mechanism from pHis to pAsp and there is evidence that such phosphate transfer can occur in the gas-phase during MS analysis. We have shown intermolecular

phosphate transfer within a pHis-containing peptide dimer during resonant CID, generating a doubly phosphorylated product ion from a singly phosphorylated precursor.<sup>100</sup> We believe that this is likely due to the nature of the system under investigation (high concentrations of a single peptide species) and the conditions used for fragmentation. Detailed investigation of a panel of synthetic pHis peptides subjected to HCD or EThcD failed to reveal any phosphate relocation.<sup>12</sup> However, gas-phase re-arrangement during both CID and ETD has also been reported for pLys and pArg-containing peptides under certain conditions.<sup>55,57,101,102</sup>

Gas-phase rearrangement of peptide ions is not a new phenomenon; numerous groups have explored this during CID, both at the peptide level,<sup>103–109</sup> and for CPhos.<sup>54</sup> Although some studies have reported CID-mediated transfer of a phosphate group to an unmodified hydroxyl side chain in synthetic peptides,<sup>54</sup> other studies on larger peptide cohorts have indicated that while this can be observed, the degree of phosphate transfer during CID is minimal and as such should not be of concern for HTP phosphoproteomics studies.<sup>110</sup> Indeed, the phosphoproteomics community has certainly taken this to heart and the idea of gas-phase phosphate transfer is seldom mentioned. To our knowledge, almost all reports of gas-phase phosphate transfer are following resonant CID, rather than higher energy CID/HCD or EThcD as is more typically used today. Thus, while phosphate 'jumping' may be worthy of consideration in the future, we do not currently believe this to be a major source of misidentification in CPhos/NCPHos analysis. However, whether phosphate transfer contributes to NCPHos in mammalian systems *in vivo* remains to be explored.

## Sulfation, the forgotten +80 Da modification

An important yet largely neglected PTM is protein sulfation. Although first identified nearly 60 years ago,<sup>111</sup> only ~50 human proteins have been identified as 'sulfated' in UniProt, with the site of sulfation having been defined experimentally in approximately half of these. Current dogma suggests that this modification occurs primarily on Tyr residues as they transit through the Golgi and is thus found primarily on secreted proteins and those localised to the plasma membrane, regulating protein–protein interactions. The absence of known protein sulfatases suggests that Tyr sulfation is not reversible and its occurrence on residues that may be modified by numerous other types of PTMs (*e.g.* phosphorylation, nitration, adenylation)<sup>112</sup> raises the possibility of site-specific PTM interplay, with sulfation preventing other regulatory modifications. Tyr sulfation is known to contribute to essential functions associated with host–pathogen interactions (*e.g.* HIV entry into cells), hormone/chemokine signalling, and pro-protein cleavage,<sup>113–122</sup> somewhat explaining the extensive phenotypic abnormalities observed by knockdown/out of individual Tyrosyl-Protein-Sulfo-Transferase (TSPT) isoforms (infertility, hyperthyroidism, retinal function) and perinatal death following knockdown of both isoforms.<sup>122</sup>



However, experimental and computational studies indicate that up to 7% of Tyr residues in mammalian proteins may be sulfated, potentially making it the most abundant Tyr-based PTM. Assuming this is true, we are lacking a wealth of detailed information regarding ‘the where, the why and the what’ of this modification.<sup>123,124</sup>

From an analytical and phosphoproteomics perspective, sulfation is particularly important given the nominally isobaric mass difference of ~9 ppm between sulfate and phosphate (Fig. 2). Consequently, it is possible to mis-identify sulfation as phosphorylation and *vice versa*, unless close attention is paid to the precursor *m/z* delta mass; it is also impossible to discriminate between these two PTMs with low resolution data.

Although sulfate and phosphate are both negatively charged, attempts to enrich sulfated peptides using well-established phosphopeptide enrichment strategies such as TiO<sub>2</sub>, anion exchange and IMAC (with Fe<sup>2+</sup>, Ga<sup>3+</sup> and Ti<sup>4+</sup> counterions) have so far not been very successful,<sup>125,126</sup> hampering HTP sulfopeptide analysis. In recognition of this pressing need, we have recently developed a ‘sulfoomics’ workflow. In-depth characterisation of 10 different IMAC counter-ions and TiO<sub>2</sub> with a range of loading/wash solutions against a panel of sTyr (and identical pTyr) peptides allowed us to define conditions for preferential enrichment of sulfopeptides over phosphopeptides.<sup>71</sup> In line with previous reports, we found that the recovery of sulfated peptides was minimal (<5%) with most of the IMAC counter ions investigated.<sup>125,126</sup> However, using TiO<sub>2</sub> or IMAC-Zr<sup>4+</sup> in an acetic acid-based solution (importantly lacking glycolic acid, a common additive in phosphopeptide enrichment pipelines) we were able to demonstrate semi-preferential enrichment of sTyr peptides in a background of unmodified and phosphorylated peptides.<sup>71</sup> As for CPhos and NCPHos-containing peptides, sulfated peptides exhibit extensive neutral loss, which is both more prevalent than seen for phosphopeptides, and occurs with both EThcD and HCD.<sup>71,127,128</sup> Site localisation of sulfated peptides is therefore a major challenge. To understand the scale of this neutral loss and to define conditions optimal for the generation of site-determining product ions, we performed comprehensive fragmentation analysis of a panel of sTyr and pTyr peptides, using resonant CID, HCD, ETcID, EThcD and UVPD, at a range of different settings (a total of 43 conditions).<sup>71</sup> These experiments revealed that any collisional dissociation resulted in near complete neutral loss of the sulfate moiety (80 amu) (Fig. 5), with a similar observation for UVPD. Although hybrid EThc/cID approaches showed some potential for sTyr localisation, with up to ~50% of product ions retaining the sulfate moiety, this was still much reduced compared with the equivalent phosphopeptide. Moreover, the reduction in the observed charge state of sulfopeptides *cf.* their phosphopeptide counterpart, meant that the efficiency of EThcD was peptide dependent.

Considering the CID/HCD-induced neutral loss propensity of sTyr, we evaluated the application of 10% NCE HCD for the discrimination of phospho- and sulfo-peptides. In contrast to the pTyr peptides where neutral loss (of either 80 or 98 Da) accounted for <1% of the total MS2 ion intensity, the

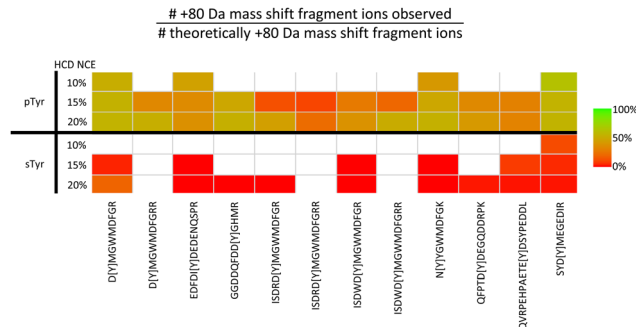


Fig. 5 Challenges associated with localisation of sulfotyrosine (sTyr) versus phosphotyrosine (pTyr) on equivalent peptides following HCD. Identical sTyr- and pTyr-containing peptides were subjected to low NCE (10%, 15% or 20%) HCD and the degree of product ion neutral loss quantified using the equation. Green indicates all theoretical fragment ions contain the PTM induced +80 Da mass shift. Red indicates all theoretical fragment ions lack the +80 Da mass shift. White – no software based identification. Adapted from ref. 71 with permission. Copyright © 2023 Daly *et al.*, under a CC-BY 4.0 license.

equivalent sTyr peptides resulted in a median of 85% neutral loss (80 Da) with no evidence of neutral loss of 98 Da. Therefore, we postulated that using a neutral loss triggering strategy, where observation of –80 Da from the precursor at 10% HCD triggered a second MS/MS scan at 32% HCD, could enable ‘on-the-fly’ sulfopeptide identification, albeit in the absence of site localisation. In combination with our bespoke acetic acid-based enrichment protocol, and incorporating a phosphatase treatment step, we have now characterised 21 sulfated proteins (27 sites of sTyr) from a HEK293 cell secretome sample, a 70% increase in the number of identified sulfation sites. We think this HTP strategy shows great promise for future application across a range of biological systems and, while undoubtedly advancing the field, believe that there is still room for improvement to advance our understanding of sulfation as a PTM.

Given the significant increase in electronegativity of sY compared with pY (the primary pK<sub>a</sub> of sY is –2.1 compared with 1.38 for pY) (as reported in the Human Metabolome database;<sup>129</sup> <https://www.hmdb.ca>), we suggest that the most promising avenue for confident localisation of sulfation sites in the future is likely to be negative ion mode MS. Such an approach has already been shown to increase ionisation efficiency and be less susceptible to neutral loss.<sup>125,130,131</sup> However, HTP implementation would require development of suitable search algorithms for negative mode MS/MS data interrogation.

## Combinatorial PTM analysis by top-down proteomics

Unlike the peptide-based analyses detailed above, top-down proteomics (TDP) describes the characterisation of proteins as intact analytes.<sup>132,133</sup> Rather than defining each modified peptide (or site) as an individual entity, losing understanding of the fine-tuning interplay of PTM connectivity at the protein level, TDP aims to define the complexity and heterogeneity of the



various protein species that arise due to PTMs (covalent, chemical and cleavage), alternative splicing and single nucleotide polymorphisms (SNPs) (Fig. 1).<sup>134</sup> Each of these myriad of protein forms, or proteoforms, likely has a unique biological function; their characterisation is thus essential to understand the true biological complexity of protein function and the role that SNPs and PTMs have on fine-tuning these processes.<sup>2</sup>

Unlike peptide-based analysis, interrogation of intact proteins in theory overcomes issues associated with missing peptides (compromising complete PTM mapping of proteins) and enables characterisation of PTM interplay on a single molecule, something which is not possible when considering modified peptides from a protein as separate entities.<sup>135</sup> While bottom-up experiments benefit greatly from their ability to localise PTMs, crucial information on the overall proteoform population is lost. Given the prevalence of protein phosphorylation, its importance in biological regulation, and the priming function that specific phosphorylation events can play in mediating both other PTMs and biomolecular interactions, characterising the phosphorylation landscape (or indeed the PTM landscape) at the intact protein level is essential as we move to a genuine understanding of protein structure–function relationships. Hierarchical multi-site phosphorylation is an exquisite means of rapid and dynamic functional regulation which is not possible to define using peptide-based investigations.

Great strides have been made in TDP over the last quarter of a century. However, there remain substantial challenges, notably for accurate PTM site localisation on larger proteins (> ~30 kDa), and particularly when these modifications induce the same mass shift as is the case with multiple phosphorylation and/or sulfation events. High resolution mass spectrometers can be used to define the expressed gene product and identify the number and type of different PTMs based on MS1 spectra (mass determination) and a relatively limited number of MS2 product ions.<sup>136,137</sup> Unfortunately, inefficient fragmentation and data analysis challenges, combined with the need to separate proteoforms with small mass differences (and even smaller *m/z* differences) means that we are still far from being able to achieve genuine proteoform characterisation for the vast majority of the proteome.

That being said, TDP studies are now capable of identifying tens of thousands of proteoforms from the thousands of proteins expressed in human cells.<sup>3,137–139</sup> Yet these studies often fail to properly discriminate closely related proteoforms such as those that may differ by a single phosphate group, and are unable to differentiate those that are isobaric. Recent work in our group has focused on more targeted investigation of signalling proteins (> 45 kDa) with the aim not only of defining the proteoform landscape (the number and type of PTMs), but critically, pinpointing the sites of modification, the aim being to define the hierarchical and combinatorial phosphorylation-mediated regulation of protein kinases needed to correlate proteoform with cellular function.<sup>140</sup>

It should be emphasised that localisation of covalent modifications and SNPs proves particularly challenging for larger analytes and those that are heavily modified due to an

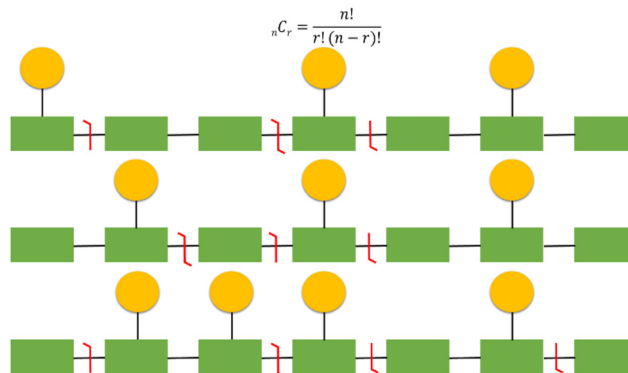


Fig. 6 Challenges of proteoform identification and the need for site specific fragmentation to differentiate isobaric proteoforms. Depicted are examples of three different proteoforms (amino acids in green, phospho-sites represented by yellow circles). Illustrating isobaric proteoforms (top 2) and proteoforms that differ by a single additional phosphate group (bottom 2). The equation can be used to calculate the number of potential proteoform combinations, where *n* is the number of potential phospho-sites and *r* is the number of observed events.

exponential increase in potential proteoform heterogeneity (Fig. 6). As a relatively simple but commonplace example, if a protein has 10 potential sites of phosphorylation and three phosphorylation or ‘action’ events, there are a theoretical 120 ( ${}_{10}C_3$ ) combinations. Multiple versions of the triply phosphorylated protein may indeed exist within the cell at a given time, but it would be reasonable to assume that not all 120 combinations are biologically feasible or relevant. However, in a situation where multiple combinations potentially exist, comprehensive and confidently assigned fragmentation data is necessary for unambiguous proteoform identification. Hence, as protein size, and/or the number of action sites increases, this becomes even more challenging both from an analytical and a computational perspective. Moreover, the increased number of product ions generated as a function of protein size means that ion current is split, resulting in decreased signal to noise and compromising sensitivity.<sup>141</sup>

To compound these issues, protein fragmentation is generally not uniform; product ions preferentially arise from the protein termini, with fragmentation in the middle being more sporadic due to inherent retention of some higher order structure even under denaturing conditions – sometimes referred to as the ‘spaghetti’ model. PTMs are thus harder to localise if they lie within these central, less accessible regions.<sup>142,143</sup> Current TDP search algorithms generally fail to highlight ambiguously localised PTMs, increasing the likelihood of misinterpretation in the absence of careful curation.

As with any analyses, the manner in which the MS instrumentation is operated is fundamental to the quality of the data generated and thus the ease with which data can be interpreted. For TDP, scan resolution at both MS1 and MS2 levels is a critical feature. While isotopic resolution at MS1 level is preferred to allow charge state determination, the time taken to achieve this can be problematic on a separation timescale, *i.e.* if MS measurement is coupled in-line with an LC system.



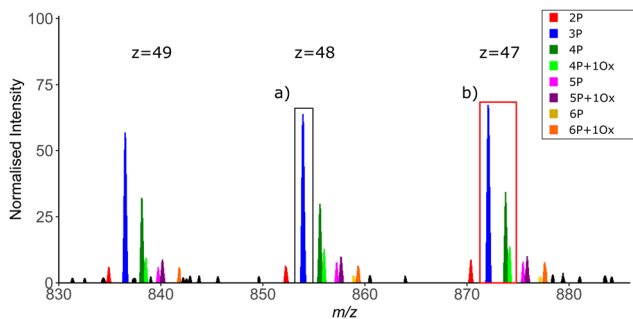


Fig. 7 Intact mass spectrum showing proteoform complexity across three charge state envelopes of an exemplar  $\sim 45$  kDa protein. Depending on the occurrence of related proteoforms and the isolation width, it may not be possible to generate proteoform specific MS2 spectra, instead generating chimeric spectra from co-isolated species. Compare isolation of a single 3-phosphate containing ion ((a) black box) versus co-isolation of the multiple proteoforms including oxidised and non-oxidised variants of the 4 phosphate-containing species ((b) red box).  $z$  = charge state; P = phosphorylation; O = methionine oxidation.

Alternatively, charge deconvolution based on observation of different precursor ion charge states means that isotopic resolution is not essential to determine mass (Fig. 7). However, the greater the resolution at MS1, the easier it is to differentiate closely related proteoforms of the same species at high charge state, minimising precursor co-isolation and thus the generation of chimeric MS2 spectra. The width of the isolation window for ion selection for MS2 is also important in this context to minimise the number of potential conflicting site-determining ions within a given MS spectrum (Fig. 7).

In contrast, high resolution at the MS2 level (resolving power  $> 50k$ ) is necessary to allow deconvolution and thus assignment of the multiple overlapping product ions of different charge state.<sup>144</sup> Averaging of multiple microscans for both MS1 and MS2 spectra can also improve the signal-to-noise ratio, albeit with a penalty on the overall duty time of the experiment.<sup>141</sup>

While advances in on-line proteoform separation strategies will undoubtedly improve site localisation capabilities in TDP (as proteoforms are introduced into the mass spectrometer at different times), the time taken for front end proteoform separation must be balanced against the requirement for extended duty cycles to maximise spectral quality.<sup>145–147</sup>

## Native IM-MS for the elucidation of PTM-induced conformational changes

While peptide and protein analysis under denaturing conditions can be used to elucidate information pertaining to the type and site of PTM (primary structure), it does not provide information on the effects of PTMs on higher order protein structure. X-ray crystallography and cryo-electron microscopy (cryo-EM) (and to a lesser extent NMR) remain the analytical methods of choice to study these effects given their ability for atomic level resolution. Although unable to yield such granular information, the relative ease of performing native ion mobility-MS (IM-MS) means that this technique has

become an important component of the structural biologists' toolbox.<sup>148</sup> Numerous groups, including ours, have demonstrated the power of IM-MS to reveal changes in higher order structure (including PTM-mediated conformational dynamics, stability and oligomerisation state) and explore ligand-induced changes to structure (including  $K_D$  determination).<sup>24,29,148–168</sup>

A number of studies have used IM-MS to investigate the effect of phosphorylation on structure, both at the peptide and the protein level.<sup>29,79,150,169,170</sup> Some of our examples of the broad utility of native IM-MS include investigations into the protein kinases Aurora A, the catalytic subunit of protein kinase A (PKA<sub>C</sub>), and the NF- $\kappa$ B transcription factor p50.<sup>29,150,171</sup> Building on our long-standing interest in the cell-cycle regulated protein kinase Aurora A, we used native IM-MS to define phosphorylation-induced protein conformational changes, determining that the weighted average collision cross section (CCS) for the phosphorylated, active version of this enzyme is larger than its non-phosphorylated, inactive (D274N) counterpart (CCS of 23.9 nm<sup>2</sup> compared with 22.3 nm<sup>2</sup> respectively). Furthermore, we demonstrated that the active (phosphorylated) protein existed as a smaller number of discrete conformers with smaller CCS distribution (CCSD), demonstrating that the phosphorylated active protein is more structurally constrained (less dynamic) (Fig. 8A).<sup>171</sup> In comparable studies, we have shown that phosphorylation (or Ser to Asp phosphomimetic mutation) increases the CCS of both PKA<sub>C</sub> and the NF- $\kappa$ B subunit p50 to a degree that exceeded the proportionally marginal increase in mass due to the PTM(s), while the effect on conformational flexibility (CCSD) appeared to be dependent on both the protein and the site of modification.<sup>29,150</sup> Interestingly, our investigations of PKA<sub>C</sub> showed conformational-dependent binding of PKI (the heat-stable inhibitor protein of PKA<sub>C</sub>), with preferential adduction to the more compact conformer and absence of binding to PKAc variants that lacked this specific conformer.<sup>150</sup>

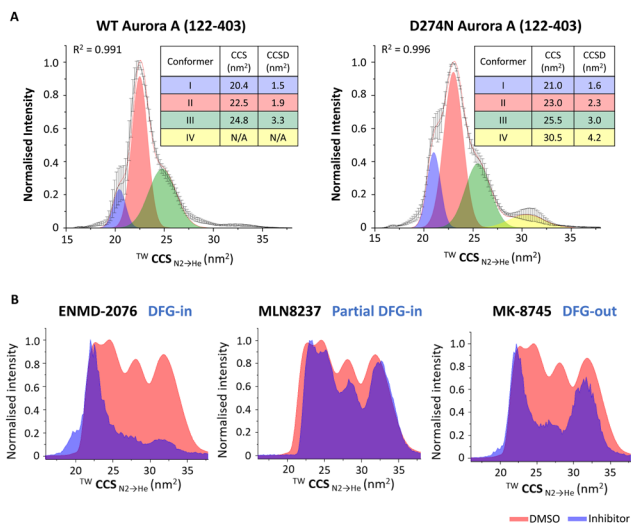
Further Aurora A investigations also showed the utility of collision-induced unfolding IM-MS to distinguish binding modes of ATP-competitive inhibitors, revealing that different inhibitors 'locked' Aurora A in distinct partially unfolded inhibitor-bound states (conformers), correlating with previously reported classifications of Aurora A inhibitors as 'DFG-in', 'DFG-out' or 'partial DFG-out' (DFG-inter) (Fig. 8B).<sup>171</sup>

In a similar vein, a study by Nshanian *et al.* provided insight as to the mechanism of inhibition of the microtubule-stabilising tau protein by the 'molecule tweezer' compound, CLR01. In addition to identifying the CLR01 binding sites using top-down electron capture dissociation, they also used IM-MS to explore the role of phosphorylation on inhibitor binding, demonstrating marked compaction of the inhibitor bound 4R-tau isoform in a manner that was dependent on its phosphorylation state.<sup>172</sup>

While IM-MS investigations aimed at exploring the effects of PTMs and/or ligand binding on protein structure are still relatively limited, such studies highlight the benefits of this strategy to unravel the disparate effects of PTMs on protein configuration and their interplay with molecular or exogenous ligands.







**Fig. 8** Exploring the effects of PTMs or ligand binding on the conformation of Aurora A (122–203). Native ion mobility mass spectra showing the  $^{TW}CCS_{N_2 \rightarrow He}$  for the  $[M + 11H]^{11+}$  species of (A) (left) wild-type (WT) phosphorylated active Aurora A (122–403) and (right) non-phosphorylated inactive D274N Aurora A (122–403), detailing the rotationally averaged collision cross section (CCS) and CCS distribution (CCSD) indicative of conformer flexibility for each conformational state. (B) WT Aurora A in the absence (red, DMSO vehicle) and presence (blue) of inhibitors: ENMD-2076 (which favours DFG-in mode); MLN8237 (partial DFG-out); MK-8745 (DFG-out), following collision-induced unfolding with 26 V collision energy. Figure adapted from ref. 24 with permission. Copyright © 2022 Tomlinson *et al.* Published by American Chemical Society, under a CC-BY 4.0 license.

## Final thoughts

The (apparently) never ending developments in the field of PTM analysis – separation strategies, technological solutions and computational algorithms, means that, as far as we have already come as a community in furthering our understanding of protein phosphorylation and other PTMs over the last century, there is plenty yet to explore. We are incredibly excited to see how continued analytical improvements advance our understanding of both phosphorylation and sulfation-driven biology. The scale of the challenge that we face means that as the community inevitably expands, and some analyses become more routine we must ensure that collectively we generate robust analytical data that is useful for advancing our knowledge on the effect of PTMs, alone or in combination.

## Author contributions

All authors contributed to writing – original draft preparation, review and editing of this manuscript. Claire E. Eyers: supervision, project administration; funding acquisition.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We acknowledge the talents and dedicated efforts of all students and staff, past and present, of the Eyers group. This work was supported by funding to C. E. E. from the Biotechnology and Biosciences Research Council (BBSRC: BB/X002780/1, BB/W00349X/1, BB/T018127/1, BB/S018514/1, BB/R000182/1, BB/M012557/1) and Northwest Cancer Research (CR1157).

## References

- S. Ramazi and J. Zahiri, *Database (Oxford)*, 2021, **2021**, baab012.
- R. Aebersold, J. N. Agar, I. J. Amster, M. S. Baker, C. R. Bertozzi, E. S. Boja, C. E. Costello, B. F. Cravatt, C. Fenselau, B. A. Garcia, Y. Ge, J. Gunawardena, R. C. Hendrickson, P. J. Hergenrother, C. G. Huber, A. R. Ivanov, O. N. Jensen, M. C. Jewett, N. L. Kelleher, L. L. Kiessling, N. J. Krogan, M. R. Larsen, J. A. Loo, R. R. Ogorzalek Loo, E. Lundberg, M. J. MacCoss, P. Mallick, V. K. Mootha, M. Mrksich, T. W. Muir, S. M. Patrie, J. J. Pesavento, S. J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S. A. Slavoff, L. M. Smith, M. P. Snyder, P. M. Thomas, M. Uhlén, J. E. Van Eyk, M. Vidal, D. R. Walt, F. M. White, E. R. Williams, T. Wohlschlagler, V. H. Wysocki, N. A. Yates, N. L. Young and B. Zhang, *Nat. Chem. Biol.*, 2018, **14**, 206–214.
- B. S. Drown, K. Jooß, R. D. Melani, C. Lloyd-Jones, J. M. Camarillo and N. L. Kelleher, *J. Proteome Res.*, 2022, **21**, 1299–1310.
- E. A. Ponomarenko, E. V. Poverennaya, E. V. Ilgisonis, M. A. Pyatnitskiy, A. T. Kopylov, V. G. Zgoda, A. V. Lisitsa and A. I. Archakov, *Int. J. Anal. Chem.*, 2016, **2016**(7436849), 7436849.
- C. E. Eyers and S. J. Gaskell, *Wiley Encycl. Chem. Biol.*, 2007, 1–34, DOI: [10.1002/9780470048672.webc469](https://doi.org/10.1002/9780470048672.webc469).
- J. M. Lee, H. M. Hammarén, M. M. Savitski and S. H. Baek, *Nat. Commun.*, 2023, **14**, 201.
- R. Nussinov, C. J. Tsai, F. Xin and P. Radivojac, *Trends Biochem. Sci.*, 2012, **37**, 447–455.
- P. Cohen, *Nat. Rev. Drug Discovery*, 2002, **1**, 309–315.
- H. Xu, Y. Wang, S. Lin, W. Deng, D. Peng, Q. Cui and Y. Xue, *Genomics, Proteomics Bioinf.*, 2018, **16**, 244–251.
- A. L. Darling and V. N. Uversky, *Front. Genet.*, 2018, **9**, 158.
- T. M. Karve and A. K. Cheema, *J. Amino Acids*, 2011, **2011**, 207691.
- G. Hardman, S. Perkins, P. J. Brownridge, C. J. Clarke, D. P. Byrne, A. E. Campbell, A. Kalyuzhnyy, A. Myall, P. A. Eyers and A. R. Jones, *EMBO J.*, 2019, **38**, e100847.
- J. V. Olsen, B. Blagoev, F. Gnäd, B. Macek, C. Kumar, P. Mortensen and M. Mann, *Cell*, 2006, **127**, 635–648.
- T. Hunter and B. M. Sefton, *Proc. Natl. Acad. Sci. U. S. A.*, 1980, **77**, 1311–1315.
- D. R. Stover, J. Caldwell, J. Marto, K. Root, J. Mestan, M. Stumm, O. Ornatsky, C. Orsi, N. Radosevic, L. Liao, D. Fabbro and M. F. Moran, *Clin. Proteomics*, 2004, **1**, 69–80.
- P. V. Attwood, P. G. Besant and M. J. Piggott, *Amino Acids*, 2011, **40**, 1035–1051.
- P. V. Attwood, M. J. Piggott, X. L. Zu and P. G. Besant, *Amino Acids*, 2007, **32**, 145–156.
- P. G. Besant, P. V. Attwood and M. J. Piggott, *Curr. Protein Pept. Sci.*, 2009, **10**, 536–550.
- S. J. Humphrey, D. E. James and M. Mann, *Trends Endocrinol. Metab.*, 2015, **26**, 676–687.
- P. P. Roux and P. Thibault, *Mol. Cell. Proteomics*, 2013, **12**, 3453–3464.
- C. E. Haydon, P. A. Eyers, L. D. Aveline-Wolf, K. A. Resing, J. L. Maller and N. G. Ahn, *Mol. Cell. Proteomics*, 2003, **2**, 1055–1067.
- L. Bury, P. A. Coelho, A. Simeone, S. Ferries, C. E. Eyers, P. A. Eyers, M. Zernicka-Goetz and D. M. Glover, *J. Cell Biol.*, 2017, **216**, 3571–3590.
- D. P. Byrne, S. Shrestha, M. Galler, M. Cao, L. A. Daly, A. E. Campbell, C. E. Eyers, E. A. Veal, N. Kannan and P. A. Eyers, *Sci. Signaling*, 2020, **13**, eaax2713.
- L. J. Tomlinson, M. Batchelor, J. Sarsby, D. P. Byrne, P. J. Brownridge, R. Bayliss, P. A. Eyers and C. E. Eyers, *J. Am. Soc. Mass Spectrom.*, 2022, **33**, 420–435.



- 25 Y. Tsuchiya, D. P. Byrne, S. G. Burgess, J. Bormann, J. Bakovic, Y. Huang, A. Zhyvolou, B. Y. K. Yu, S. Peak-Chew, T. Tran, F. Bellany, A. B. Tabor, A. E. Chan, L. Guruprasad, O. Garifulin, V. Filonenko, M. Vonderach, S. Ferries, C. E. Eyers, J. Carroll, M. Skehel, R. Bayliss, P. A. Eyers and I. Gout, *Redox. Biol.*, 2020, **28**, 101318.
- 26 D. P. Byrne, C. J. Clarke, P. J. Brownridge, A. Kalyuzhnyy, S. Perkins, A. Campbell, D. Mason, A. R. Jones, P. A. Eyers and C. E. Eyers, *Biochem. J.*, 2020, **477**, 2451–2475.
- 27 L. A. Daly, P. J. Brownridge, M. Batie, S. Rocha, V. See and C. E. Eyers, *Sci. Signaling*, 2021, **14**, eabf6685.
- 28 F. Lanucara, C. Lam, J. Mann, T. P. Monie, S. A. Colombo, S. W. Holman, J. Boyd, M. C. Dange, D. A. Mann, M. R. White and C. E. Eyers, *Open Biol.*, 2016, **6**, 3159–3175.
- 29 M. Vonderach, D. P. Byrne, P. E. Barran, P. A. Eyers and C. E. Eyers, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 128–138.
- 30 F. Ardito, M. Giuliani, D. Perrone, G. Troiano and L. Lo Muzio, *Int. J. Mol. Med.*, 2017, **40**, 271–280.
- 31 M. Mann and O. N. Jensen, *Nat. Biotechnol.*, 2003, **21**, 255–261.
- 32 A. Kalyuzhnyy, P. A. Eyers, C. E. Eyers, E. Bowler-Barnett, M. J. Martin, Z. Sun, E. W. Deutsch and A. R. Jones, *J. Proteome Res.*, 2022, **21**, 1510–1524.
- 33 D. Singh, P. Gupta, S. L. Singla-Pareek, K. H. M. Siddique and A. Pareek, *Curr. Genomics*, 2021, **22**, 59–74.
- 34 M. J. Piggott and P. V. Attwood, *Amino Acids*, 2017, **49**, 1309–1323.
- 35 G. R. Grimsley, J. M. Scholtz and C. N. Pace, *Protein Sci.*, 2009, **18**, 247–251.
- 36 T. Y. Low, M. A. Mohtar, P. Y. Lee, N. Omar, H. Zhou and M. Ye, *Mass Spectrom. Rev.*, 2021, **40**, 309–333.
- 37 P. A. Connor and A. J. McQuillan, *Langmuir*, 1999, **15**, 2916–2921.
- 38 M. R. Larsen, T. E. Thingholm, O. N. Jensen, P. Roepstorff and T. J. D. Jørgensen, *Mol. Cell. Proteomics*, 2005, **4**, 873–886.
- 39 A. Leitner, M. Sturm, O. Hudecz, M. Mazanek, J. H. Smått, M. Lindner, W. Lindner and K. Mechtler, *Anal. Chem.*, 2010, **82**, 2726–2733.
- 40 M. W. Pinkse, P. M. Uitto, M. J. Hilhorst, B. Ooms and A. J. Heck, *Anal. Chem.*, 2004, **76**, 3935–3943.
- 41 T. E. Thingholm and O. N. Jensen, *Methods Mol. Biol.*, 2009, **527**, 47–56, xi.
- 42 T. E. Thingholm and M. R. Larsen, *Methods Mol. Biol.*, 2016, **1355**, 123–133.
- 43 E. S. Simon, M. Young, A. Chan, Z. Q. Bao and P. C. Andrews, *Anal. Biochem.*, 2008, **377**, 234–242.
- 44 J. Urban, *Anal. Chim. Acta*, 2022, **1199**, 338857.
- 45 C. M. Potel, S. Lemeer and A. J. R. Heck, *Anal. Chem.*, 2019, **91**, 126–141.
- 46 A. Tholey, J. Reed and W. D. Lehmann, *J. Mass Spectrom.*, 1999, **34**, 117–123.
- 47 J. P. DeGnore and J. Qin, *J. Am. Soc. Mass Spectrom.*, 1998, **9**, 1175–1188.
- 48 S. A. Carr, M. J. Huddleston and R. S. Annan, *Anal. Biochem.*, 1996, **239**, 180–192.
- 49 S. Ferries, S. Perkins, P. J. Brownridge, A. Campbell, P. A. Eyers, A. R. Jones and C. E. Eyers, *J. Proteome Res.*, 2017, **16**, 3448–3459.
- 50 P. J. Boersema, S. Mohammed and A. J. Heck, *J. Mass Spectrom.*, 2009, **44**, 861–878.
- 51 A. Schlosser, R. Pipkorn, D. Bossemeyer and W. D. Lehmann, *Anal. Chem.*, 2001, **73**, 170–176.
- 52 W. D. Lehmann, R. Krüger, M. Salek, C.-W. Hung, F. Wolschin and W. Weckwerth, *J. Proteome Res.*, 2007, **6**, 2866–2873.
- 53 L. Cui, I. Yapici, B. Borhan and G. E. Reid, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 141–148.
- 54 A. M. Palumbo and G. E. Reid, *Anal. Chem.*, 2008, **80**, 9735–9747.
- 55 L. Cui and G. E. Reid, *Proteomics*, 2013, **13**, 964–973.
- 56 J. Villén, S. A. Beausoleil and S. P. Gygi, *Proteomics*, 2008, **8**, 4444–4452.
- 57 M. Rožman, *J. Mass Spectrom.*, 2011, **46**, 949–955.
- 58 F. Lanucara, D. Chi Hoo Lee and C. E. Eyers, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 214–225.
- 59 R. Brown, S. S. Stuart, S. Houel, N. G. Ahn and W. M. Old, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 1128–1142.
- 60 J. E. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz and D. F. Hunt, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9528–9533.
- 61 D. L. Swaney, G. C. McAlister, M. Wirtala, J. C. Schwartz, J. E. Syka and J. J. Coon, *Anal. Chem.*, 2007, **79**, 477–485.
- 62 C. K. Frese, A. F. M. Altaalar, H. van den Toorn, D. Nolting, J. Griep-Raming, A. J. R. Heck and S. Mohammed, *Anal. Chem.*, 2012, **84**, 9668–9673.
- 63 C. K. Frese, H. Zhou, T. Taus, A. F. Altaalar, K. Mechtler, A. J. Heck and S. Mohammed, *J. Proteome Res.*, 2013, **12**, 1520–1525.
- 64 M. Penkert, A. Hauser, R. Harmel, D. Fiedler, C. P. R. Hackenberger and E. Krause, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 1578–1585.
- 65 J. S. Brodbelt, L. J. Morrison and I. Santos, *Chem. Rev.*, 2020, **120**, 3328–3380.
- 66 M. S. Thompson, W. Cui and J. P. Reilly, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 1439–1452.
- 67 D. D. Holden, J. M. Pruet and J. S. Brodbelt, *Int. J. Mass Spectrom.*, 2015, **390**, 81–90.
- 68 W. Cui, M. S. Thompson and J. P. Reilly, *J. Am. Soc. Mass Spectrom.*, 2005, **16**, 1384–1398.
- 69 M. R. Robinson, J. M. Taliaferro, K. N. Dalby and J. S. Brodbelt, *J. Proteome Res.*, 2016, **15**, 2739–2748.
- 70 K. L. Fort, A. Dyachenko, C. M. Potel, E. Corradini, F. Marino, A. Barendregt, A. A. Makarov, R. A. Scheltema and A. J. R. Heck, *Anal. Chem.*, 2016, **88**, 2303–2310.
- 71 L. A. Daly, D. P. Byrne, S. Perkins, P. J. Brownridge, E. McDonnell, A. R. Jones, P. A. Eyers and C. E. Eyers, *bioRxiv*, 2023, DOI: [10.1101/2023.07.15.549150](https://doi.org/10.1101/2023.07.15.549150).
- 72 T. Taus, T. Köcher, P. Pichler, C. Paschke, A. Schmidt, C. Henrich and K. Mechtler, *J. Proteome Res.*, 2011, **10**, 5354–5362.
- 73 S. A. Beausoleil, J. Villén, S. A. Gerber, J. Rush and S. P. Gygi, *Nat. Biotechnol.*, 2006, **24**, 1285–1292.
- 74 D. D. Shteynberg, E. W. Deutsch, D. S. Campbell, M. R. Hoopmann, U. Kusebauch, D. Lee, L. Mendoza, M. K. Midha, Z. Sun, A. D. Whetton and R. L. Moritz, *J. Proteome Res.*, 2019, **18**, 4262–4272.
- 75 D. Fermin, S. J. Walmsley, A. C. Gingras, H. Choi and A. I. Nesvizhskii, *Mol. Cell. Proteomics*, 2013, **12**, 3409–3419.
- 76 M. M. Savitski, S. Lemeer, M. Boesche, M. Lang, T. Mathieson, M. Bantscheff and B. Kuster, *Mol. Cell. Proteomics*, 2011, **10**, M110.003830.
- 77 W. Jiang, B. Wen, K. Li, W. F. Zeng, F. da Veiga Leprevost, J. Moon, V. A. Petyuk, N. J. Edwards, T. Liu, A. I. Nesvizhskii and B. Zhang, *Mol. Cell. Proteomics*, 2021, **20**, 100171.
- 78 A. A. Shvartsburg, D. Singer, R. D. Smith and R. Hoffmann, *Anal. Chem.*, 2011, **83**, 5078–5085.
- 79 M. S. Glover, J. M. Dilger, M. D. Acton, R. J. Arnold, P. Radivojac and D. E. Clemmer, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 786–794.
- 80 D. Oliinyk and F. Meier, *Proteomics*, 2023, **23**, e2200032.
- 81 S. Tiwari, S. B. Jamal, S. S. Hassan, P. Carvalho, S. Almeida, D. Barh, P. Ghosh, A. Silva, T. L. P. Castro and V. Azevedo, *Front. Microbiol.*, 2017, **8**, 1878.
- 82 P. D. Boyer, M. Deluca, K. E. Ebner, D. E. Hultquist and J. B. Peter, *J. Biol. Chem.*, 1962, **237**, PC3306–PC3308.
- 83 C. M. Potel, M.-H. Lin, A. J. R. Heck and S. Lemeer, *Nat. Methods*, 2018, **15**, 187–190.
- 84 S. R. Fuhs, J. Meisenhelder, A. Aslanian, L. Ma, A. Zagorska, M. Stankova, A. Binnie, F. Al-Obeidi, J. Mauger, G. Lemke, J. R. Yates, 3rd and T. Hunter, *Cell*, 2015, **162**, 198–210.
- 85 J.-M. Kee, R. C. Oslund, A. D. Couvillon and T. W. Muir, *Org. Lett.*, 2015, **17**, 187–189.
- 86 J. M. Kee, R. C. Oslund, D. H. Perlman and T. W. Muir, *Nat. Chem. Biol.*, 2013, **9**, 416–421.
- 87 J. M. Kee, B. Villani, L. R. Carpenter and T. W. Muir, *J. Am. Chem. Soc.*, 2010, **132**, 14327–14329.
- 88 C. M. Potel, M.-H. Lin, N. Prust, H. W. P. van den Toorn, A. J. R. Heck and S. Lemeer, *Anal. Chem.*, 2019, **91**, 5542–5547.
- 89 N. M. Leijten, A. J. R. Heck and S. Lemeer, *Nat. Methods*, 2022, **19**, 827–828.
- 90 N. Prust, P. C. van Breugel and S. Lemeer, *Mol. Cell. Proteomics*, 2022, **21**, 100232.
- 91 E. Kinoshita, M. Takahashi, H. Takeda, M. Shiro and T. Koike, *Dalton Trans.*, 2004, 1189–1193, DOI: [10.1039/b400269e](https://doi.org/10.1039/b400269e).
- 92 W. He, H. Yingyun, J. Bo, S. Zhigang, Y. Huiming, Z. Baofeng, L. Zhen, Z. Lihua and Z. Yukui, *Talanta*, 2023, **251**, 123740.
- 93 R. C. Oslund, J. M. Kee, A. D. Couvillon, V. N. Bhatia, D. H. Perlman and T. W. Muir, *J. Am. Chem. Soc.*, 2014, **136**, 12899–12911.
- 94 X. Zhao, S. Fu, Y. Zhao and F. Ni, *Rapid Commun. Mass Spectrom.*, 2021, **35**, e9186.



- 95 H. Steen, A. Pandey, J. S. Andersen and M. Mann, *Sci. STKE*, 2002, **2002**, pl16.
- 96 K. A. Ramsbottom, A. Prakash, Y. P. Riverol, O. M. Camacho, M.-J. Martin, J. A. Vizcaino, E. W. Deutsch and A. R. Jones, *J. Proteome Res.*, 2022, **21**, 1603–1615.
- 97 A. D. Waddell, H. Ojha, S. Agarwal, C. J. Clarke, A. Terriente-Felix, H. Zhou, P. Kakade, A. Knebel, A. M. Shaw, R. Gourlay, J. Varghese, R. F. Soares, R. Toth, T. Macartney, P. A. Eyers, N. Morrice, R. Bayliss, A. J. Whitworth, C. E. Eyers and M. M. K. Muqit, *bioRxiv*, 2023, DOI: [10.1101/2023.03.31.534916](https://doi.org/10.1101/2023.03.31.534916).
- 98 I. Gulerez, Y. Funato, H. Wu, M. Yang, G. Kozlov, H. Miki and K. Gehring, *EMBO Rep.*, 2016, **17**, 1890–1900.
- 99 A. D. B. Pannifer, A. J. Flint, N. K. Tonks and D. Barford, *J. Biol. Chem.*, 1998, **273**, 10454–10462.
- 100 M.-B. Gonzalez-Sanchez, F. Lanucara, G. E. Hardman and C. E. Eyers, *Int. J. Mass Spectrom.*, 2014, **367**, 28–34.
- 101 A. Schmidt, G. Ammerer and K. Mechtler, *Proteomics*, 2013, **13**, 945–954.
- 102 J. Bertran-Vicente, R. A. Serwa, M. Schümann, P. Schmieder, E. Krause and C. P. Hackenberger, *J. Am. Chem. Soc.*, 2014, **136**, 13622–13628.
- 103 A. G. Harrison, A. B. Young, C. Bleiholder, S. Suhai and B. Paizs, *J. Am. Chem. Soc.*, 2006, **128**, 10364–10365.
- 104 J. Yagüe, A. Parada, M. Ramos, S. Ogueta, A. Marina, F. Barahona, J. A. López de Castro and J. Vázquez, *Anal. Chem.*, 2003, **75**, 1524–1535.
- 105 C. Jia, Z. Wu, C. B. Lietz, Z. Liang, Q. Cui and L. Li, *Anal. Chem.*, 2014, **86**, 2917–2924.
- 106 S. Molesworth, S. Osburn and M. Van Stipdonk, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 2174–2181.
- 107 N. C. Polfer, B. C. Bohrer, M. D. Plasencia, B. Paizs and D. E. Clemmer, *J. Phys. Chem. A*, 2008, **112**, 1286–1293.
- 108 R. Chawner, S. W. Holman, S. J. Gaskell and C. E. Eyers, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 1927–1938.
- 109 R. Chawner, S. J. Gaskell and C. E. Eyers, *Rapid Commun. Mass Spectrom.*, 2012, **26**, 205–206.
- 110 M. Aguiar, W. Haas, S. A. Beausoleil, J. Rush and S. P. Gygi, *J. Proteome Res.*, 2010, **9**, 3103–3107.
- 111 H. Gregory, P. M. Hardy, D. S. Jones, G. W. Kenner and R. C. Sheppard, *Nature*, 1964, **204**, 931–933.
- 112 I. Griswold-Prenner, A. K. Kashyap, S. Mazhar, Z. W. Hall, H. Fazelinia and H. Ischiropoulos, *J. Biol. Chem.*, 2023, 105038, DOI: [10.1016/j.jbc.2023.105038](https://doi.org/10.1016/j.jbc.2023.105038).
- 113 J. H. Ippel, C. J. de Haas, A. Bunschoten, J. A. van Strijp, J. A. Kruijtzter, R. M. Liskamp and J. Kemmink, *J. Biol. Chem.*, 2009, **284**, 12363–12372.
- 114 C. Seibert, C. T. Veldkamp, F. C. Peterson, B. T. Chait, B. F. Volkman and T. P. Sakmar, *Biochemistry*, 2008, **47**, 11251–11262.
- 115 S. Costagliola, V. Panneels, M. Bonomi, J. Koch, M. C. Many, G. Smits and G. Vassart, *EMBO J.*, 2002, **21**, 504–513.
- 116 G. L. Hortin, *Blood*, 1990, **76**, 946–952.
- 117 A. Leyte, H. B. van Schijndel, C. Niehrs, W. B. Huttner, M. P. Verbeet, K. Mertens and J. A. van Mourik, *J. Biol. Chem.*, 1991, **266**, 740–746.
- 118 J. R. Bundgaard, J. Vuust and J. F. Rehfeld, *EMBO J.*, 1995, **14**, 3073–3079.
- 119 M. Farzan, T. Mirzabekov, P. Kolchinsky, R. Wyatt, M. Cayabyab, N. P. Gerard, C. Gerard, J. Sodroski and H. Choe, *Cell*, 1999, **96**, 667–676.
- 120 E. G. Cormier, M. Persuh, D. A. Thompson, S. W. Lin, T. P. Sakmar, W. C. Olson and T. Dragic, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 5762–5767.
- 121 S. D. Rodgers, R. T. Camphausen and D. A. Hammer, *Biophys. J.*, 2001, **81**, 2001–2009.
- 122 A. D. Westmuckett, A. J. Hoffhines, A. Borghei and K. L. Moore, *Gen. Comp. Endocrinol.*, 2008, **156**, 145–153.
- 123 K. L. Moore, *J. Biol. Chem.*, 2003, **278**, 24243–24246.
- 124 P. A. Baeuerle and W. B. Huttner, *J. Biol. Chem.*, 1985, **260**, 6434–6439.
- 125 A. L. Capriotti, A. Cerrato, A. Laganà, C. M. Montone, S. Piovesana, R. Zenezini Chiozzi and C. Cavaliere, *Anal. Chem.*, 2020, **92**, 7964–7971.
- 126 G. D. Balderama, E. P. Meneses, L. H. Orihuela, O. V. Hernández, R. C. Franco, V. P. Robles and C. V. F. Batista, *Rapid Commun. Mass Spectrom.*, 2011, **25**, 1017–1027.
- 127 Y. Yu, A. J. Hoffhines, K. L. Moore and J. A. Leary, *Nat. Methods*, 2007, **4**, 583–588.
- 128 M. Edelson-Averbukh, A. Shevchenko, R. Pipkorn and W. D. Lehmann, *J. Am. Soc. Mass Spectrom.*, 2011, **22**, 2256–2268.
- 129 D. S. Wishart, A. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S. Tian, B. L. Lee, M. Berjanskii, R. Mah, M. Yamamoto, J. Jovel, C. Torres-Calzada, M. Hiebert-Giesbrecht, V. W. Lui, D. Varshavi, D. Varshavi, D. Allen, D. Arndt, N. Khetarpal, A. Sivakumaran, K. Harford, S. Sanford, K. Yee, X. Cao, Z. Budinski, J. Liigand, L. Zhang, J. Zheng, R. Mandal, N. Karu, M. Dambrova, H. B. Schiöth, R. Greiner and V. Gautam, *Nucleic Acids Res.*, 2021, **50**, D622–D631.
- 130 M. R. Robinson, K. L. Moore and J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 1461–1471.
- 131 K. E. Hersberger and K. Håkansson, *Anal. Chem.*, 2012, **84**, 6370–6377.
- 132 G. E. Reid and S. A. McLuckey, *J. Mass Spectrom.*, 2002, **37**, 663–675.
- 133 N. L. Kelleher, *Anal. Chem.*, 2004, **76**, 197a–203a.
- 134 L. M. Smith, N. L. Kelleher, M. Linial, D. Goodlett, P. Langridge-Smith, Y. Ah Goo, G. Safford, L. C. Anderson, G. Kruppa, R. Zubarev, J. Rontree, J. Chamot-Rooke, J. Garavelli, A. Heck, J. Loo, D. Penque, M. Hornshaw, C. Hendrickson, L. Pasa-Tolic, C. Borchers, D. Chan, N. Young, J. Agar, C. Masselon, M. Gross, F. McLafferty, Y. Tsybin, Y. Ge, I. Sanders, J. Langridge, J. Whitelegge, A. Marshall and P. The Consortium for Top Down, *Nat. Methods*, 2013, **10**, 186–187.
- 135 K. R. Durbin, J. C. Tran, L. Zamdborg, S. M. Sweet, A. D. Catherman, J. E. Lee, M. Li, J. F. Kellie and N. L. Kelleher, *Proteomics*, 2010, **10**, 3589–3597.
- 136 E. N. McCool, T. Xu, W. Chen, N. C. Beller, S. M. Nolan, A. B. Hummon, X. Liu and L. Sun, *Sci. Adv.*, 2022, **8**, eabq6348.
- 137 R. D. Melani, V. R. Gerbasi, L. C. Anderson, J. W. Sikora, T. K. Toby, J. E. Hutton, D. S. Butcher, F. Negrão, H. S. Seckler, K. Srzentić, L. Fornelli, J. M. Camarillo, R. D. LeDuc, A. J. Cesnik, E. Lundberg, J. B. Greer, R. T. Fellers, M. T. Robey, C. J. DeHart, E. Forte, C. L. Hendrickson, S. E. Abbatiello, P. M. Thomas, A. I. Kokaji, J. Levitsky and N. L. Kelleher, *Science*, 2022, **375**, 411–418.
- 138 E. N. McCool, T. Xu, W. Chen, N. C. Beller, S. M. Nolan, A. B. Hummon, X. Liu and L. Sun, *Sci. Adv.*, 2022, **8**, eabq6348.
- 139 M. A. R. Hollas, M. T. Robey, R. T. Fellers, R. D. LeDuc, P. M. Thomas and N. L. Kelleher, *Nucleic Acids Res.*, 2022, **50**, D526–D533.
- 140 A. Po and C. E. Eyers, *J. Proteome Res.*, 2023, under revision.
- 141 P. D. Compton, L. Zamdborg, P. M. Thomas and N. L. Kelleher, *Anal. Chem.*, 2011, **83**, 6868–6874.
- 142 N. Siuti and N. L. Kelleher, *Nat. Methods*, 2007, **4**, 817–821.
- 143 F. Lanucara and C. E. Eyers, *Mass Spectrom. Rev.*, 2013, **32**, 27–42.
- 144 T. K. Toby, L. Fornelli and N. L. Kelleher, *Annu. Rev. Anal. Chem.*, 2016, **9**, 499–519.
- 145 J. A. Melby, D. S. Roberts, E. J. Larson, K. A. Brown, E. F. Bayne, S. Jin and Y. Ge, *J. Am. Soc. Mass Spectrom.*, 2021, **32**, 1278–1294.
- 146 P. V. Shliaha, S. Gibb, V. Gorshkov, M. S. Jespersen, G. R. Andersen, D. Bailey, J. Schwartz, S. Eliuk, V. Schwämmle and O. N. Jensen, *Anal. Chem.*, 2018, **90**, 12519–12526.
- 147 L. Sun, M. D. Knierman, G. Zhu and N. J. Dovichi, *Anal. Chem.*, 2013, **85**, 5989–5995.
- 148 F. Lanucara, S. W. Holman, C. J. Gray and C. E. Eyers, *Nat. Chem.*, 2014, **6**, 281–294.
- 149 C. E. Eyers, M. Vonderach, S. Ferries, K. Jeacock and P. A. Eyers, *Curr. Opin. Chem. Biol.*, 2018, **42**, 167–176.
- 150 D. P. Byrne, M. Vonderach, S. Ferries, P. J. Brownridge, C. E. Eyers and P. A. Eyers, *Biochem. J.*, 2016, **473**, 3159–3175.
- 151 H. Amijee, C. Bate, A. Williams, J. Virdee, R. Jeggo, D. Spanswick, D. I. C. Scopes, J. M. Treherne, S. Mazzitelli, R. Chawner, C. E. Eyers and A. J. Doig, *Biochemistry*, 2012, **51**, 8338–8352.
- 152 H. Hernández and C. V. Robinson, *Nat. Protoc.*, 2007, **2**, 715–726.
- 153 J. M. Daniel, G. McCombie, S. Wendt and R. Zenobi, *J. Am. Soc. Mass Spectrom.*, 2003, **14**, 442–448.
- 154 F. Debaene, A. Boeuf, E. Wagner-Rousset, O. Colas, D. Ayoub, N. Corvaia, A. Van Dorsselaer, A. Beck and S. Cianférani, *Anal. Chem.*, 2014, **86**, 10674–10683.
- 155 D. Cubrilovic, A. Biela, F. Sialaff, T. Steinmetzer, G. Klebe and R. Zenobi, *J. Am. Soc. Mass Spectrom.*, 2012, **23**, 1768–1777.
- 156 L. J. Tomlinson and C. E. Eyers, *Methods Mol. Biol.*, 2020, **2084**, 179–190.



- 157 B. T. Ruotolo, J. L. Benesch, A. M. Sandercock, S. J. Hyung and C. V. Robinson, *Nat. Protoc.*, 2008, **3**, 1139–1152.
- 158 C. Uetrecht, R. J. Rose, E. van Duijn, K. Lorenzen and A. J. Heck, *Chem. Soc. Rev.*, 2010, **39**, 1633–1655.
- 159 T. E. Angel, U. K. Aryal, S. M. Hengel, E. S. Baker, R. T. Kelly, E. W. Robinson and R. D. Smith, *Chem. Soc. Rev.*, 2012, **41**, 3912–3928.
- 160 S. Niu, J. N. Rabuck and B. T. Ruotolo, *Curr. Opin. Chem. Biol.*, 2013, **17**, 809–817.
- 161 E. R. Dickinson, E. Jurneczko, J. Nicholson, T. R. Hupp, J. Zawacka-Pankau, G. Selivanova and P. E. Barran, *Front. Mol. Biosci.*, 2015, **2**, 39.
- 162 P. Lossl, A. M. Brunner, F. Liu, A. C. Leney, M. Yamashita, R. A. Scheltema and A. J. Heck, *ACS Cent. Sci.*, 2016, **2**, 445–455.
- 163 B. Jovceviski, M. A. Kelly, J. A. Aquilina, J. L. P. Benesch and H. Ecroyd, *Anal. Chem.*, 2017, **89**, 13275–13282.
- 164 G. Ben-Nissan and M. Sharon, *Curr. Opin. Chem. Biol.*, 2018, **42**, 25–33.
- 165 R. Beveridge, L. G. Migas, R. W. Kriwacki and P. E. Barran, *Angew. Chem., Int. Ed.*, 2019, **58**, 3114–3118.
- 166 M. Nshanian, C. Lantz, P. Wongkongkathep, T. Schrader, F. G. Klärner, A. Blumke, C. Despres, M. Ehrmann, C. Smet-Nocca, G. Bitan and J. A. Loo, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 16–23.
- 167 J. W. McCabe, M. J. Hebert, M. Shirzadeh, C. S. Mallis, J. K. Denton, T. E. Walker and D. H. Russell, *Mass Spectrom. Rev.*, 2021, **40**, 280–305.
- 168 E. Christofi and P. Barran, *Chem. Rev.*, 2023, **123**, 2902–2949.
- 169 K. Thalassinou, M. Grabenauer, S. E. Slade, G. R. Hilton, M. T. Bowers and J. H. Scrivens, *Anal. Chem.*, 2009, **81**, 248–254.
- 170 K. Ogata, C. H. Chang and Y. Ishihama, *Mass Spectrom.*, 2021, **10**, A0093.
- 171 L. J. Tomlinson, M. Batchelor, J. Sarsby, D. P. Byrne, P. J. Brownridge, R. Bayliss, P. A. Eyers and C. E. Eyers, *J. Am. Soc. Mass Spectrom.*, 2022, **33**, 420–435.
- 172 M. Nshanian, C. Lantz, P. Wongkongkathep, T. Schrader, F.-G. Klärner, A. Blümke, C. Despres, M. Ehrmann, C. Smet-Nocca, G. Bitan and J. A. Loo, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 16–23.

