

# A bespoke analytical workflow for the confident identification of sulfopeptides and their discrimination from phosphopeptides

Leonard A. Daly<sup>1,2</sup>, Dominic P. Byrne<sup>2</sup>, Simon Perkins<sup>3</sup>, Philip J Brownridge<sup>1</sup>, Euan McDonnell<sup>2,3</sup> Andrew R. Jones<sup>2,3</sup>, Patrick A. Eyers<sup>2</sup>, and Claire E. Eyers<sup>1,2\*</sup>.

<sup>1</sup> Centre for Proteome Research, Institute of Systems, Molecular & Integrative Biology, University of Liverpool, Crown Street, Liverpool, L69 7ZB, U.K.

<sup>2</sup> Department of Biochemistry & Systems Biology, Institute of Systems, Molecular & Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, U.K.

<sup>3</sup> Computational Biology Facility, Institute of Systems, Molecular & Integrative Biology, University of Liverpool, Crown Street, Liverpool, L69 7ZB, U.K.

\*Correspondence: Claire E. Eyers (ceyers@liverpool.ac.uk).

**ABSTRACT:** Protein tyrosine sulfation (sY) is a post-translational modification (PTM) catalysed by Golgi-resident Tyrosyl Protein SulfoTransferases (TPSTs). Information on protein tyrosine sulfation is currently limited to ~50 human proteins with only a handful having verified sites of sulfation. The contribution of this chemical moiety for the regulation of biological processes, both inside and outside the cell, remains poorly defined, in large part due to analytical limitations. Mass spectrometry-based proteomics is the method of choice for PTM analysis, but has yet to be applied for the systematic investigation and large-scale analysis of bio-molecular sulfation (constituting the ‘sulfome’), primarily due to issues associated with discrimination of sY- from phosphotyrosine (pY)-containing peptides. In this study, we developed a mass spectrometry (MS)-based workflow centred on the characterization of sY-peptides, incorporating optimised Zr<sup>4+</sup>-IMAC and TiO<sub>2</sub> enrichment strategies. Extensive characterization of a panel of sY- and pY-peptides using an array of MS fragmentation regimes (CID, HCD, EThC, ETcID, UVPD) highlights differences in the ability to generate site-determining product ions, which can be exploited to differentiate sulfated peptides from nominally isobaric phosphopeptides based on precursor ion neutral loss at low collision energy. Application of our analytical workflow to a HEK-293 cell extracellular secretome facilitated identification of 21 new sulfotyrosine-containing proteins, several of which we validate enzymatically using *in vitro* sulfation assays. This study demonstrates the applicability of our strategy for confident, high-throughput, ‘sulfomics’ studies, and reveals new sY interplay between enzymes relevant to both protein and glycan sulfation.

Protein phosphorylation involves the reversible, covalent addition of a phosphate group to the side chain of amino acids, predominantly Ser, Thr and Tyr, alongside 9 non-alcoholic amino acid substrates.<sup>2-4</sup> As a highly abundant post-translational modification (PTM), phosphorylation has been extensively investigated from both an analytical and functional perspective. It plays critical roles in almost all physiological processes, with an estimated ~90% of human proteins able to be phosphorylated *in vivo* (the phosphoproteome).<sup>5</sup> Phosphorylation is catalysed by protein kinases, and >500 enzymes encoded within the human genome have the ability to catalyse the transfer of the  $\gamma$ -phosphate group from adenosine triphosphate (ATP) to amino acid residues on substrates. Approximately 100 of these kinases are tyrosine specific,<sup>6</sup> while others can phosphorylate Ser/Thr and Tyr, or Ser/Thr alone. In contrast, protein sulfation, *i.e.* the covalent addition of sulfate which is transferred from the 3'-phosphoadenosine-5'-phosphosulfate (PAPS) co-factor, is poorly characterized analytically and is predicted to be considerably less abundant than phosphorylation. Furthermore, unlike phosphorylation, sulfation is believed to be relatively specific for tyrosine residues, being catalyzed in humans by two separate gene products that encode Tyrosyl-Protein-Sulfo-

Transferases (TPSTs), termed TPST1 and 2. Some 51 human proteins are annotated in UniProt as containing ‘sulfotyrosine’ (sY), but only 33 of these have been experimentally validated, compared with over 11,000 human ‘phosphoprotein’ entries (accessed May 2023).<sup>7</sup> Experimental and computational estimates based on site conservation and biological environment suggest that ~1-7% of all Tyr residues have the potential to be sulfated (in flies and mice),<sup>8,9</sup> which potentially makes sY a much more prevalent tyrosine-based PTM than often assumed.

sY was first identified over 50 years ago in fibrinogen/gastrin, and appears to be governed (minimally) by acidic consensus motifs in protein targets, which were first characterized biochemically in the early 1980's.<sup>10-13</sup> Sulfation results in biologically-relevant changes in protein:protein affinity, modulating host-pathogen interactions, chemotaxis, FGF7 signalling, proteolytic peptide processing and HIV entry *via* the viral chemokine co-receptor CCR5.<sup>14-22</sup> Mice lacking TPST1 have defects in body mass and retinal function, whereas TPST2-deficient mice exhibit hyperthyroidism and infertility.<sup>23</sup> In contrast, double TPST knockout mice have high perinatal mortality rates due to pulmonary asphyxia, and lack detectable sY, confirming that TPST1/2 are together rate-limiting for sY deposition *in vivo*.<sup>23</sup>

While TPST-dependent protein sulfation occurs in the Golgi and is believed to be an irreversible modification, sulfation has been most extensively identified on a variety of extracellular/secreted proteins. Interestingly, a recent study reported cytosolic sY of the tumour suppressor p53,<sup>24</sup> suggesting that the tyrosine 'sulfome' may be more extensive and diverse than previously predicted, revealing potential gaps in our current knowledge that are attributable to sub-optimal analytical capabilities for accurate high-throughput investigation of sY.

Phosphotyrosine (pY) and sY are chemically 'similar' with near isobaric mass (sY being 9.6 mDa lighter than pY; pY: 79.966331 Da and sY: 79.956815 Da), and are of comparable size and shape, containing at least one negative charge at physiological pH, formally pY = -2, sY = -1, Figure 1A. Despite these similarities, the best-established phosphopeptide enrichment protocols and gas-phase fragmentation strategies are highly unsuitable for sY-peptide isolation and analysis, stymieing isolation and MS-based investigation.<sup>25-29</sup> Currently, sY characterization is reliant on, and highly restricted by, low-throughput TPST screens *e.g.* on immunoprecipitated putative substrates and/or immunoblotting with monoclonal anti-sY antibodies,<sup>35</sup> S-radio-labelling or fluorescent-based enzyme assays.<sup>8, 17, 20, 30-38</sup> While such techniques are useful for confirming prevalent sY sites in simple mixtures, these approaches are restricted to known or predicted (acidic motif) sites of sulfation, which limits their application for discovery purposes. Hence, there is a clear and pressing need for the development of bespoke analytical pipelines that are compatible with high-throughput (HTP) technologies to identify & interrogate sulfomes.

Recent advancements in mass spectrometry (MS)-based phosphoproteomics pipelines have facilitated interrogation of phosphorylation-dependent signalling networks to an unprecedented level of detail, consolidating it as an essential technique to explore phosphorylation-mediated signalling. An essential part of these analytical pipelines is the selective enrichment of phosphopeptides, based on inherent properties of the appended phosphate moiety, notably a formal net negative charge. Exploitation of standard enrichment strategies, which include titanium dioxide (TiO<sub>2</sub>), anion exchange and Immobilised Metal-ion Affinity Chromatography (IMAC) utilising various metal ions including: Ga<sup>3+</sup>, Ti<sup>4+</sup> and Fe<sup>3+</sup>, can aid analysis of select sY-peptide standards (typically 2-3 non-tryptic, synthetically sulfated peptides). However, these enrichment processes are reported to be highly inefficient.<sup>25-27</sup> Compounding this, the relatively low ionization efficiency of sulfopeptides and the high lability of the sulfoester bond during MS analysis compromises sensitivity: neutral loss occurs during collision-induced dissociation (and sometimes during electrospray ionization-MS in the absence of induced fragmentation) generating product ions equivalent to non-modified fragments, hampering sulfosite localization and discrimination from pY-containing peptides.<sup>28, 29, 39</sup>

Electron transfer dissociation (ETD) and related hybrid fragmentation strategies such as EThCD and ETcID (which use supplemental collisional dissociation), have proven useful for characterizing peptides that contain labile PTMs, notably phosphorylation.<sup>40-42</sup> However, sulfopeptides still remain highly prone to sulfate 'neutral loss'.<sup>28</sup> Direct implementation of electron transfer-mediated approaches in a high-throughput (positive ion mode) proteomics study thus do not appear suitable for sensitive sulfation analysis against the relatively small selection of sY peptides so far investigated.

Halim *et al.*<sup>43</sup> report the application of UVPD at 213 nm (as well as infrared multiphoton dissociation (IRMPD) and high-low photodissociation (HiLoPD)) for sulfopeptide characterisation. In line with the observations of others,<sup>39, 43, 44</sup> they demonstrate reduced stability of a singly protonated sulfopeptide compared with the doubly charged species. While their data indicated that UVPD at 193 nm (and to a lesser degree, 213 nm) provided almost complete sequence coverage for higher charge state sulfopeptide cations, they reported extensive sulfonate loss.

Negative ion mode MS offers potential advantages for sY-peptide characterization, given the acidic nature of both the sY residue and the side chains surrounding the site of modification reported in most substrates, with reports of greater ionization efficiency and less sulfonate neutral loss, akin to that observed for phosphopeptides.<sup>27, 39, 44</sup> In particular, negative ETD (NETD) and negative ion mode electron capture dissociation (niECD) show promise for the analysis and localization of peptide sulfation sites, with niECD generating spectra with complete sulfonate retention on the 5 non-tryptic sY-peptides investigated.<sup>44</sup> Negative ion UVPD also shows promise for localization of sY on peptides, having been used to analyze 5 non-tryptic peptides (4 in common with<sup>44</sup>). However, optimal characterisation utilises a custom 193 nm laser that is not currently available on commercial instrumentation.<sup>39</sup>

While there are reports of large-scale negative ion mode proteomics studies,<sup>45, 46</sup> negative ion mode is not routinely applied for HTP studies, in part due to a lack of tested/trained search engines and data analysis tools, as well as reduced commercially available options for fragmentation. Computational tools that are developed for automated interpretation of negative ion mode tandem mass spectra will need to take account of the different and more complex fragmentation pathways that arise *cf* positive ion mode during collision-induced dissociation, including multiple, and extensive, side chain neutral losses.<sup>47-52</sup>

In this paper, we present a proteomics pipeline that can be readily implemented using current commercial instrumentation and search algorithms that is optimised for the HTP identification of sY-containing peptides from both chemically-defined and cellular mixtures, in a manner that permits their discrimination from pY (and other) phosphate-containing peptides. As well as developing a sY-peptide enrichment strategy that employs acetic acid-based solutions with TiO<sub>2</sub> or IMAC-Zr<sup>4+</sup>, we substantially advance sY-peptide fragmentation studies. In-depth analysis of the largest panel of tryptic sulfopeptides (and phosphopeptide analogues) assembled to date using collision-based, electron-mediated and UVPD fragmentation strategies allowed us to optimise conditions for sY *versus* pY discrimination, exploiting the substantive differences in the energy required for sulfonate *versus* phosphate neutral loss. We also present, and validate, the first application of our sY specific LC-MS/MS-based pipeline for characterisation of the sulfome from a complex cellular mixture. The confident identification of sY-proteins includes 21 novel human sulfoproteins, several of which we validate as TPST1/2 substrates using *in vitro* sulfotransferase assays and MS-based tryptic peptide characterization.

## Experimental Section

### Reagents

Powdered chemical reagents were purchased from Merck. Custom peptides (unmodified and phosphorylated) were purchased from ThermoFisher as PEPOTEC light grade 2 purity. High performance liquid chromatography (HPLC) grade solvents for MS were purchased from ThermoFisher Scientific. All Eppendorf tubes used are ultra-high recovery Eppendorf tubes (STARLAB). PAPS (adenosine 3'-phosphate 5'-phosphosulfate, lithium salt hydrate) was purchased from Merck and stored at -80 °C to afford maximal stability.

### Purification of recombinant proteins

Human TPST1 (residues Lys43–Leu360) and TPST2 (residues Gly43–Leu359) (both lacking the transmembrane domains) were purified as recombinant proteins with an N-terminal 6xHis-tag and a 3C protease cleavage site using pOPINF (OPPF-UK) and refolded *in vitro* into a catalytically-active conformation, as previously described.<sup>31</sup> Lambda (λ) protein phosphatase was purified to homogeneity as previously described.<sup>53</sup> The tyrosine kinase EphA3, which autophosphorylates on Tyr when expressed in *E. coli*, comprised the kinase domain and the juxtamembrane region with an N-terminal 6xHis-tag, and was expressed in pLysS *E. coli* from pET28a LIC vector, and purified using Ni-NTA agarose and size exclusion chromatography.<sup>54</sup>

### In-vitro sulfation of peptide standards

Peptide sulfation assays were performed as described previously<sup>31</sup>. Briefly, 2.5 µg of each peptide in individual reactions were incubated at 20 °C for 18 h in the presence of 1 µg TPST1 and TPST2, and 0.5 mM of the co-factor PAPS, the source of the transferred sulfate. In vitro sulfated peptides were combined before undergoing stage-tip based strong cation exchange clean-up.<sup>55</sup> The eluent was aliquoted into 0.25 µg/peptide fractions and vacuum-dried by centrifugation prior to analysis.

### Protein phosphatase treatment

For cell-based experiments, a final concentration of 2 mM MnCl<sub>2</sub> and 1 mM dithiothreitol was added to the sample followed by addition of a 100:1 (w/w) ratio of peptide: purified λ phosphatase. Samples were incubated at 37 °C for 2 hr with constant shaking at 600 rpm.

### Enzymatic sTyr and pTyr peptide modification assays

A 5-FAM fluorophore (maximal excitation of 495 nm/maximal emission of 520 nm, detectable by a LED-induced fluorescence), was covalently coupled to the N-terminus of a custom peptide substrate containing a single modifiable Y site embedded in a canonical acidic motif (5-FAM-EDFED[Y]EDG-CO-NH<sub>2</sub>) which was purchased from Pepceuticals (Leicester, U.K). The peptide was enzymatically Tyr sulfated as described above, or enzymatically Tyr phosphorylated using the same buffer conditions but substituting TPST1 and 5 mM PAPS with the tyrosine kinase EphA3 and 5 mM Mg:ATP. The PerkinElmer Lab-Chip EZ II Reader system, 12-sipper chip and CR8 coating, and assay separation buffer, were all purchased from PerkinElmer and employed as reported previously.<sup>56,57</sup> Pressure and voltage settings were adjusted manually to afford optimal separation of modified (sY and pY) vs. unmodified peptides. Individual phosphatase assays were performed (in triplicate) in a 384-well plate in a volume of 80 µL using 2 µM peptide in the presence of the indicated quantity (µg) of λ phosphatase, 50 mM HEPES (pH

7.4), 0.015% (v/v) Brij-35, 5 mM MnCl<sub>2</sub> and 1 mM DTT. The degree of peptide sulfation/phosphorylation was directly calculated by the EZ Reader software by differentiating modified: unmodified phosphorylated or sulfated peptide peak ratios.

### Titanium dioxide enrichment

Dried peptide standards were resuspended in the appropriate buffers to a concentration of 0.2 µg/µL by water bath sonication for 10 min. TiO<sub>2</sub> resin (GL-Sciences) was added to a final amount of 5:1 (w/w) TiO<sub>2</sub> resin : peptide and incubated at 25 °C with 1,500 rpm shaking for 30 min before centrifugation at 2,000 g for 1 min at room temperature and supernatant removed. TiO<sub>2</sub> resin-peptide complexes were washed 3x for 10 min with 1,500 rpm shaking in the same buffer, resuspending in an equal volume after centrifugation as before. TiO<sub>2</sub> resin was vacuum centrifuged for 15 min, prior to addition of an equal volume 5% ammonium hydroxide (in water) and shaking at 1,500 rpm for 10 min. Samples were centrifuged as before, supernatant collected and dried to completion under vacuum centrifugation.

### Labelling NTA-agarose coated magnetic beads

PureCube NTA MagBeads (Cube BioTech) were labelled according to the manufacturer's protocol, scaling to 100 µL resin slurry using magnetic support racks. Pre-labelled Ti<sup>4+</sup> and Zr<sup>4+</sup> PureCube MagBeads were washed following the labelling protocol. All beads were made to a 1% (w/v) stock and stored at 4 °C prior to use.

### Immobilized metal affinity chromatography enrichment

An adapted version of the TiO<sub>2</sub> enrichment protocol above was used. Alterations include: using 2.5:1 ratio of IMAC resin : peptide, and washing the required quantity of beads twice in resuspension buffer using a volume five-fold that of the bead slurry, prior to the addition to peptides.

### C18 stage-tip clean-up

sY-peptides for fragmentation studies, standard input samples and spiked into BSA/Casein digests were subjected to in-house C18 stage-tip (Empore Supelco 47 mm C18 extraction discs) clean-up prior to LC-MS/MS analysis. Three discs were packed per 200 µL tip and centrifuged at 5000 g for 5 min to pack the column. All centrifugation steps were at 4000 g for 2 min (or until all liquid has passed through the tip) at room temperature. Stage-tips were equilibrated by sequential washing with 200 µL of methanol, elution buffer (80% (v/v) ACN (acetonitrile) + 0.1% (v/v) TFA in water) and wash buffer (0.1% TFA in water) prior to loading peptide samples in 0.1% TFA. Flow through was re-applied before washing in 200 µL of wash buffer and eluting in 200 µL elution buffer. Elution was dried to completion by vacuum centrifugation.

### Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis

All dried peptides were resuspended in 3% (v/v) ACN + 0.1% TFA (in water) by water bath sonication for 10 min, followed by centrifugation at 13000 g at 4 °C for 10 min, and the cleared supernatant collected. For protein/peptide standards, peptides were separated by reversed-phase HPLC over a 30 min gradient using an Ultimate 3000 nano system (Dionex), as described in<sup>58</sup>. All data acquisition was performed using a Thermo Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific) using stated fragmentation conditions for 2+ to 5+ charge states

over a  $m/z$  range of 350-2000. MS1 spectra were acquired in the Orbitrap [120K resolution at 200  $m/z$ ], normalized automatic gain control (AGC) = 50%, maximum injection time = 50 ms and an intensity threshold for fragmentation = 2.5e<sup>4</sup>. MS2 spectra were acquired in the Orbitrap [30K resolution at 200  $m/z$ ], AGC target = normal and maximum injection time = dynamic. A dynamic exclusion window of 10 s was applied at a 10 ppm mass tolerance. For cell-based studies, peptides were separated and acquired as before with either of the following adaptions: 1) a 60 min gradient and higher-energy C-trap dissociation (HCD) set at 32% normalized collision energy (NCE). 2) For neutral loss triggered methods, HCD set at 10% NCE, MS2 acquired in the Orbitrap [15K resolution at 200  $m/z$ ] and targeted loss continuation trigger for  $m/z$  values equivalent to 1-3 sulfation PTMs (79.9568) at charge states 2+ to 5+ (25 ppm mass tolerance). Continuation of trigger was performed for neutral loss ions with at least 10% relative intensity for correct charge state assigned losses. Triggered scans were acquired in the Orbitrap [30K resolution at 200  $m/z$ ], HCD set to 32% NCE, AGC target = standard and maximum injection time = auto.

### Mass spectrometry data analysis

The panel of synthetic peptide standards and the products of immunoprecipitated heparan-sulfate 6-O-sulfotransferase 1/2/3 (H6ST-1/2/3) *in vitro* Tyr sulfation assays were analyzed using Proteome Discoverer 2.4 (Thermo Scientific) in conjunction with MASCOT<sup>59</sup> against either a custom database of all (12) peptides, BSA and Casein ( $\alpha$ S1,  $\alpha$ S2,  $\beta$ ) isoforms only, or Uni-Prot Human Reviewed database [updated May 2023] for the H6ST immunoprecipitation experiments. Data was imported into Skyline for calculating charge state distributions and retention time shifts based off  $m/z$  values from Proteome Discoverer 2.4 analysis. Secretome data was converted into MZML format using MSConvert<sup>60</sup> with peak picking “2-“ filter applied. For neutral loss datasets, an additional filtering parameter of “HCD energy 32” was applied. MZML datafiles were searched using PEAKS11 against the UniProt Human reviewed database (updated June 2022). All data was searched using trypsin (K/R, unless followed by P) with 2 miscleaves permitted and constant modifications = carbamidomethylation (C), variable modification = oxidation (M), sulfation (STY), phosphorylation (STY). Neutral loss triggered methods were additionally searched using semi-specific trypsin and the additional variable modification of deamidation (NQ). MS1 mass tolerance = 10 ppm, MS2 mass tolerance = 0.01 Da, instrument type = Orbitrap (Orbi-Orb), fragmentation = HCD, acquisition = DDA. All data was filtered to 1% FDR.

### Informatics and Localization data analysis

For localization data analysis and heatmap preparation, synthetic peptide MS data sets were searched in Comet (with parameters matched to the PEAKS11 search), which enables export of all PSMs, without any machine learning-based rescoring (as in PEAKS11) or score thresholding. Heat maps were produced in R (tidyverse, ggplot2), for every fragmentation mode (for phosphopeptides and sulfopeptides at multiple charge states). The heat map displays, for correctly identified PSMs, the proportion of identified count of fragment ions containing the known modification site, divided by the total possible count of theoretical observable ions containing the known modification site.

### Cell culture, secretome precipitation and sample preparation

Adherent HEK-293 and HEK-293T cells were seeded at a density of  $\sim$ 1.75x10<sup>5</sup> cells/cm<sup>2</sup> in DMEM media supplemented with 10% (v/v) fetal calf serum, 1% (v/v) non-essential amino acids and 1% (v/v) penicillin/streptomycin and maintained at 37°C, 5% CO<sub>2</sub> until  $\sim$ 80% confluent ( $\sim$ 2x10<sup>8</sup> cells). Cells were washed 2x in 10 ml of phosphate buffered saline (PBS) before incubation for 18 hr in serum-free growth medium. The HEK-293 ‘secretome’ was collected from the medium and centrifuged at 200 g for 5 min and then 10,000 g for 5 min, preserving the cleared supernatant each time. Secreted proteins were captured by addition of 100:1 (v/v) secretome:Strataclean resin (Agilent Technologies) ( $\sim$ 2 ml) and incubated at room temperature with end over end rotation for 2 h. Strataclean resin-protein complexes were centrifuged at 500 g for 2 min and supernatant removed. Complexes were washed twice in an equal volume (relative to initial secretome) of PBS, resuspended in 1 ml PBS and transferred to a 1.5 ml centrifuge tube and washed (3x) 1 ml PBS. Proteins were eluted in 1 ml of 5% (w/v) SDS, 500 mM NaCl and 50 mM Tris (pH 8) at 80 °C for 10 min, with brief vortexing every 2 min, followed by centrifugation at 13,000 g for 10 min at room temperature. Protein content was determined by BCA assay before samples were reduced and alkylated with dithiothreitol and iodoacetamide<sup>58</sup>. A 1:1 mixture of magnetic hydrophobic and hydrophilic Seramag speedbeads (MERCK) were added at a 2:1 (w/w) ratio of beads : protein and ACN was added to a final concentration of 80% (v/v), and incubated at 25 °C, shaking (1,500 rpm) for 30 min. The supernatant was discarded and beads washed (3x) in 200  $\mu$ l of 100% ACN. Beads were dried by vacuum centrifugation for 10 min and resuspended by water bath sonication for 2 min in 1 ml 100 mM ammonium acetate pH 8. Trypsin Gold (Promega) was added at a 33:1 (w/w) protein:trypsin ratio and incubated overnight at 37 °C with 1,500 rpm shaking. The peptide containing supernatant was transferred into a fresh tube, and the remaining beads were then washed with a 2x volume of 8 M urea, 100 mM ammonium acetate pH 8 for 30 min with 1,500 rpm shaking, before the supernatants were combined. The pooled sample was incubated on ice for 30 min and centrifuged at 13,000 g for 10 min at 4 °C and cleared supernatant collected. Samples were  $\lambda$  phosphatase-treated (as described) prior to acidification with a final concentration of 0.5% (v/v) TFA, and incubated at 37 °C for 30 min followed by 4 °C incubation for 30 min. Samples were centrifuged at 13,000 g for 10 min at 4 °C, and the cleared supernatants collected. Samples were split 1:33:33:33 respectively, for total protein, agarose coated IMAC-Zr<sup>4+</sup>, BioResyn IMAC-Zr<sup>4+</sup> HP or TiO<sub>2</sub> sY enrichment protocols, prior to drying to completion by vacuum centrifugation.

### Functional enrichment analysis

DAVID Bioinformatics Resources [v6.8]<sup>61</sup> was used to understand cellular compartments (and biological processes) in the secretome enriched protein sample.

### Heparan-sulfate 6-O-sulfotransferase 1/2 immunoprecipitation and *in vitro* sulfation assay

The cytoplasmic regions of human HS6ST1 (37-410), HS6ST (228-605), and HS6ST3 (28-471) were cloned into pcDNA3 with a 3C-protease cleavable, N-terminal, tandem StrepTag for human cell expression. HEK-293T cells were transfected at  $\sim$ 40 % confluence using a 3:1 polyethylenimine (branched, average Mw  $\sim$ 25,000 Da; Merck) to DNA ratio

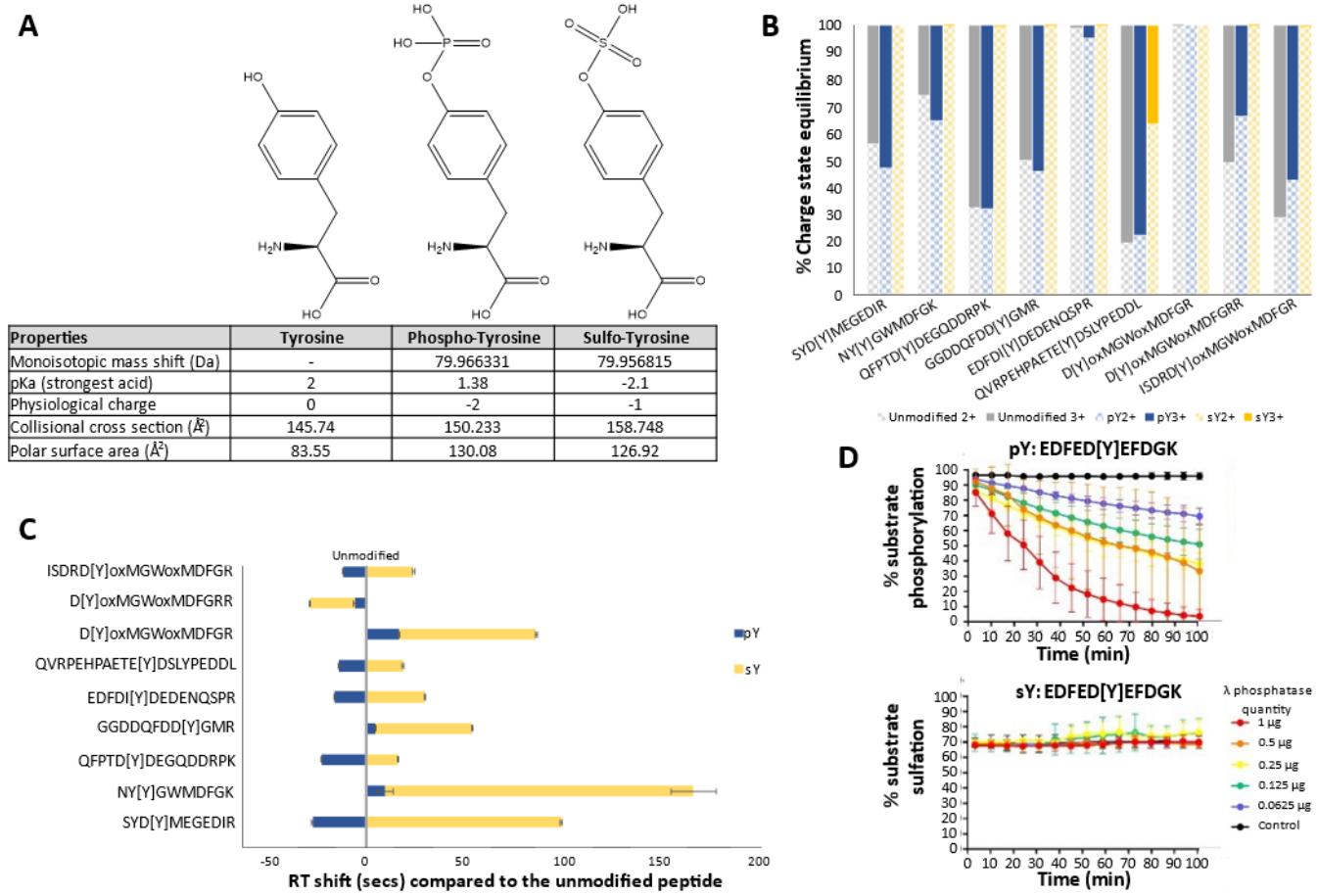
(30:10 µg, for a single 10-cm culture dish). Proteins (from 10x 10-cm culture dish) were pooled and harvested 48 h post transfection in lysis buffer (150 µl per dish) containing 50 mM tris-HCl (pH 7.4), 150 mM NaCl, 0.5% (v/v) Triton X-100, 2 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub> and 20% (v/v) glycerol and supplemented with protease and phosphatase inhibitors (Roche). Lysates were sonicated briefly on ice and clarified by centrifuged at 20,817g for 20 min at 4 °C, and the resulting supernatants were incubated with 15 µl Strep-Tactin sepharose resin (Thermo Fisher Scientific) for 3 h with gentle end over end mixing at 4 °C. Sepharose beads containing bound protein were collected and washed three times in 50 mM tris-HCl (pH 7.4) and 500 mM NaCl and then equilibrated in storage buffer (50 mM Tris-HCl (pH 7.4), 100 mM NaCl, and 5% (v/v) glycerol). HS6STs were then proteolytically eluted from the beads over a 2 h period using 3C protease (0.5 µg) at 4 °C with gentle agitation. Purified HS6STs (10 µl of the 35 µl elution volume) were sulfated with TPST1, TPST2 or both TPST1 and 2 (as described above) for 18 h at 20 °C.

## Results and Discussion

### Sulfotyrosine and phosphotyrosine have distinct biochemical and analytical properties.

The isobaric masses of sY and pY moieties are extremely similar, but possess distinct chemical properties (Fig. 1A; <sup>62</sup>). At physiological pH (~7.4), sY carries a single net charge of -1, while pY has a net charge of -2. The primary (most acidic) pKa of sY is -2.1 compared with 1.38 for pY (as reported in the Human Metabolome database (<sup>63</sup>; [www.hmdb.ca](http://www.hmdb.ca))). Based on molecular dynamics simulations and experimental analysis, sY exhibits a marked reduction in the potential for hydrogen bonding interactions compared with pY.<sup>62</sup> This difference in electronegativity likely explains the previously reported reduced ionization efficiency of sY-peptides in positive ion mode and the lability of the sulfonate group during proton-driven collision-induced dissociation <sup>28, 29, 39</sup>.

To investigate how sY or pY affects peptide analysis, we synthesised a panel of 12 Tyr-containing peptides designed based upon known/putative sulfation sites from proteins. We also generated an analogous panel containing pY. The panel of peptides



**Figure 1: Physicochemical properties of phospho-tyrosine versus sulfo-tyrosine.** A) Chemical structure and properties of Tyrosine, Phospho-Tyrosine and Sulfo-Tyrosine. A table of the monoisotopic mass shift (Da) of the phosphate/sulfate moiety in comparison to Tyrosine, the pKa value of the strongest acidic group, net charge influence at physiological pH (~8), collisional cross section (Å) and the polar surface area charge is distributed across (Å<sup>2</sup>) are presented. Data were compiled from the 2022 Human Metabolome Database. B) Charge state distribution of our standard panel of peptides in an unmodified, phosphorylated or sulfated state. Modification site denoted by [Y], ox = oxidized residue. N = 10 replicates. C) Scaled retention time shifts of pY- and sY-peptides. Normalised retention time shifts (s) are plotted with reference to the retention time of the unmodified peptide counterpart. Modification site denoted by [Y], ox = oxidized residue. N = 10 replicates. D) Real-time λ phosphatase assay to study peptide dephosphorylation or desulfation. Identical fluorescent peptides containing either pY or sY were incubated with stated quantities of λ phosphatase and the level of modification reversal (dephosphorylation or desulfation) was determined over time (min) using a ratiometric real-time assay.

were enzymatically sulfated using TPST1 and  $2^{31}$  to generate a panel of tryptic sY-peptides for comparison against synthetically generated phosphopeptides modified on the same residue (Supp. Table 1, Fig. 1A). All peptide variants (either unmodified, pY- or sY-containing) were analysed under standard positive ion mode LC-MS/MS conditions, comparing charge states and retention time (Fig. 1B, C). Peptides identified in all forms (unmodified, pY and sY) across 10 replicates were included in these analyses (a total of 9 peptides). We identified little difference in the relative abundance of the 2+ and 3+ charge states for peptides when comparing the unmodified and pY forms (Fig. 1B). In marked contrast, and consistent with previous studies,<sup>39, 64</sup> we observed a significant reduction in the formation of 3+ sY-peptide ions, preferentially observing doubly protonated species; only 1 out of the 9 peptides in this set (QVRPEH<sup>PAETE</sup>[sY]DSL<sup>Y</sup>PEDDL) presented in its  $[M+3H]^{3+}$  form (likely due to the presence of internal Arg and His residues) with all others appearing almost exclusively as 2+ species. Singly protonated sY peptide ions, when present, were at substantially lower levels (<1% relative intensity compared to 2+ ions) suggesting that they are highly unstable.<sup>39, 43, 44</sup> Comparing LC retention times (RT) for sY-peptides *versus* their unmodified counterparts over the rapid 10 min LC gradient, we observe a clear increase (ranging from 16 s to 156 s *i.e.* a RT shift of between 0.8% and 4.3%) for 8 out of the 9 sY-peptides. Conversely, and as previously reported, pY-peptides exhibited faster LC elution times (ranging from 5 s to 26 s) (Fig 1C).<sup>65, 66</sup> Interestingly, D[Y]oxMGWoxMDFGRR was the only peptide where RT decreased due to sulfation, and indeed eluted earlier than the pY variant (-22 s for sY *versus* -5 s for pY). This peptide has the highest pI value of peptides in this panel (pI = 5.96, range= 3.62-5.83, Supp. Table 1) and contains 2 Arg residues, suggesting that there is not a direct relationship between pI and modification type, and hydrophobicity.

Given the comparatively high prevalence of phosphorylated proteins relative to known sulfoproteins in the human proteome, we explored whether enzymatic phosphatase pre-treatment could reduce the relative abundance of phosphopeptides and thereby i) minimize possible mis-identification of sY-peptides, particularly given the very small mass shift between the two moieties (~9.6 mDa), and ii) improve the sensitivity of sulfoproteomics pipelines by minimizing co-enrichment of sulfopeptides and phosphopeptides. Protein phosphatase treatment combined with MS has previously been used to investigate sY-peptide identification, discriminating from pY due to a lack of activity and thus an absence of mass shift post treatment.<sup>67, 68</sup> We confirmed the specificity for dephosphorylation by monitoring activity towards pY- and sY- forms of the same fluorescently labelled peptide substrate (EDFED[Y]EFDGK), using a phosphorylation assay<sup>31</sup> in reverse (Fig. 1D) similar to the procedure we developed for real-time desulfation analysis of glycans.<sup>69</sup> Irrespective of the quantity of protein phosphatase employed ( $\mu$ g) and under conditions with essentially complete dephosphorylation of the pY-peptide (>95%), no loss of the sY-containing peptide was observed, confirming that protein phosphatase pre-treatment is likely an effective means of reducing phosphopeptide content in biological samples, whilst preserving peptide sulfation prior to enrichment and MS analysis.

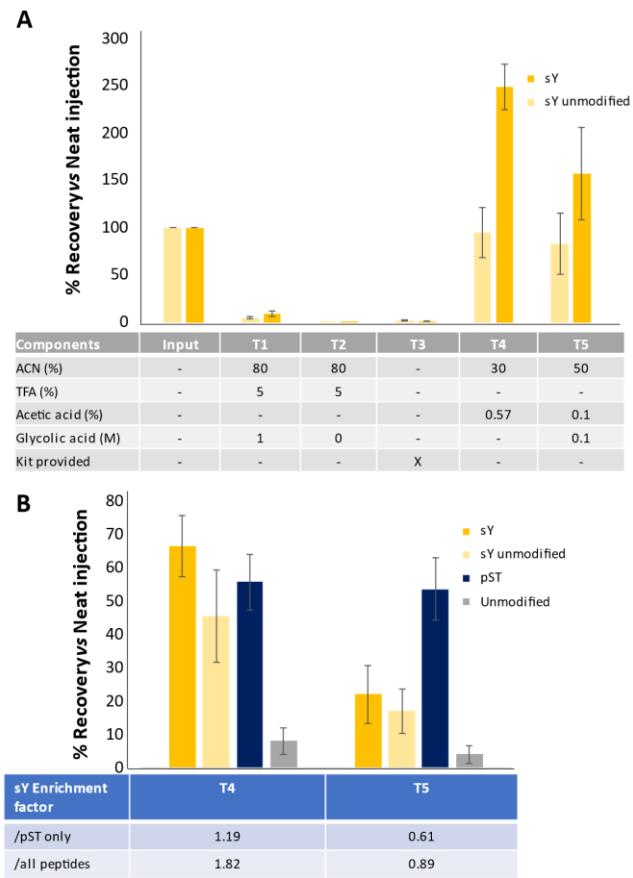
## Acetic acid-based solutions with $TiO_2$ or Zirconium<sup>4+</sup> IMAC can efficiently (and semi-preferentially) enrich sY-peptides

MS-based characterization of PTMs in complex mixtures requires enrichment of peptides (or proteins) to improve detection sensitivity and overcome the fact that PTM modified peptides usually represent a small proportion of the total peptide pool.<sup>2, 70-72</sup> The selectivity of phosphate-enrichment techniques is based on PTM-specific biochemical properties, or bespoke antibodies. Low pH Immobilised Metal ion Affinity Chromatography (IMAC) and  $TiO_2$  have thus become staples in phosphoproteomics pipelines.<sup>73-75</sup> Both  $TiO_2$  and IMAC-based phosphopeptide enrichment strategies exploit the relatively low pKa of the phosphate group (the primary pKa value of phosphate monoesters being ~2) compared with the side chains of other amino acids (the most comparable being Asp and Glu with pKa values of ~3.5 and 4.2 respectively<sup>76</sup>) to promote efficient and (relatively) specific binding. By maintaining a low pH, it is therefore possible to preferentially enrich for phosphopeptides as opposed to Asp/Glu-rich peptides with positively charged immobilized metal ions.

However, application of these approaches for the enrichment of sY-peptides is potentially complicated by two major factors; firstly, it is reported that sY undergoes acid-induced hydrolysis,<sup>77, 78</sup> which would reduce the recovery of sY using standard low pH enrichment conditions. However, we failed to detect acid-induced sulfate hydrolysis across our peptide panel, even after sample storage for a week at 4 °C in 0.1% (v/v) TFA (data not shown), in-line with other reports.<sup>27</sup> Second, the single hydroxyl group of sY and the size/orientation of the sulfonyl group (Fig. 1A) might prohibit the efficient capture of sY-peptides by  $TiO_2$ . Phosphopeptide enrichment with  $TiO_2$  relies on the spatial coordination and bidentate hydrogen bonding of the phosphate group to resin-chelated titanium ions,<sup>79-81</sup> the reduced hydrogen bonding capacity of sY-peptides may compromise  $TiO_2$  binding. Indeed, a previous report that used a commercial  $TiO_2$  kit reported no enrichment of two sY-peptides.<sup>27</sup> However, the binding and wash buffers used in this kit are proprietary, and buffer composition is known to play a significant role in the efficiency of such enrichment processes.<sup>82, 83</sup> Moreover, although sTyr antibodies exist, these are unreliable in our hands at immunoprecipitating proteins or peptides containing sTyr.

While  $Fe^{3+}$ -IMAC<sup>27</sup> and  $Ga^{3+}$ -IMAC<sup>26</sup> have been tested for the enrichment of sY-peptides, the relative efficiency has not been carefully evaluated. Other metal counter ions have also proven useful for phosphopeptide enrichment.<sup>84-86</sup> However, there has not been a comprehensive evaluation of the utility of different IMAC counter ions for sY-peptide enrichment. Previous attempts at sY peptide enrichment did not consider tryptic peptides, or fully explore the efficiency or selectivity of sY peptide enrichment as a function of binding and wash conditions. Given our desire to advance sTyr analysis in a variety of biological mixtures, we undertook a comprehensive quantitative evaluation of the ability of different immobilised media, including  $TiO_2$  and IMAC (comparing 10 different metal counter ions), in terms of specificity and efficiency for enriching tryptic sY-peptides (Figs. 2 & 3).

Initial evaluations focused on  $TiO_2$ -based enrichment using solutions typically employed for phosphopeptides, and sample loading conditions previously shown to enable binding of three (non-tryptic) sY-peptides<sup>87</sup> (Fig. 2). To rule out non-specific enrichment resulting from the increased hydrophobicity of the



**Figure 2: TiO<sub>2</sub> based enrichment of sY-peptides.** A) Solution optimisation for recovery of sY-peptides. The panel of sY- and unmodified peptides were enriched using standard TiO<sub>2</sub> protocols with different loading and wash solutions as stated and analysed by LC-MS. Recovery was determined by normalising to an equal amount of ‘unenriched’ input material subjected to C18 reverse phase clean-up. Recovery of the 8 sY- and 9 unmodified peptides is an average (mean +/- S.D.) of all equivalent (non-)modified peptides, including methionine oxidised variants which we have previously shown has little effect on relative signal intensity<sup>1</sup>. B) A mixture of trypsin digested BSA, casein- $\alpha$ S1, - $\alpha$ S2 and - $\beta$  and the sY-peptide panel were subjected to TiO<sub>2</sub> enrichment using T4 and T5 as specified in (A) and efficiency of enrichment of sY, pS/T and unmodified peptides determined. All samples were subjected to C18 clean-up prior to LC-MS/MS analysis. Recovery (mean +/- S.D.) was determined for 8 sY-, 9 unmodified synthetic peptides, as well as 9 pST peptides from casein, and 19 unmodified peptides from BSA and casein.

sY modification, sY peptide recovery rates for each condition were directly compared to the amount loaded following C18 reverse phase chromatography as a ‘clean-up’ step. Under standard TiO<sub>2</sub> phosphopeptide enrichment conditions (80% ACN, 5% TFA, 1 M Glycolic acid, T1, <sup>58</sup>), application of a modified loading buffer without glycolic acid (80% ACN, 5% TFA, T2) or using the commercially available Phos-TiO<sub>2</sub> spin-tip kit (GL Sciences, T3) we failed to observe efficient recovery or enrichment (<5%) of any sY-containing peptides (Fig. 2A). Little to no recovery was observed for non-modified peptides as would be expected. Reducing the acetonitrile content and/or replacing TFA with lower concentrations of acetic acid (30% ACN, 100 mM acetic acid: T4; 50% ACN, 0.1% acetic acid, 0.1 M glycolic acid: T5) improved recovery of all peptides with ~2.5-fold

and 1.5-fold enrichment of sY-peptides for conditions T4 and T5 respectively compared with the C18-desalting control (Fig. 2A).

To evaluate sY-enrichment in a mixture of tryptic synthetic sY peptides alongside phosphopeptides derived from purified BSA and  $\alpha/\beta$  casein (at molar excesses of 170x and 90x respectively), we compared the relative enrichment of sY- versus pST-peptides in these samples. From this peptide mixture we quantified recovery of sY peptides, unmodified peptides (lacking sTyr but still highly acidic), 9 phosphopeptides from casein, and 19 unmodified peptides (10 from BSA, and 9 from casein). Signal intensity was normalised to an equal load of non TiO<sub>2</sub>-enriched material, all samples having been subjected to C18 clean-up (Fig. 2B). sY enrichment factor was then determined by comparing the relative recovery of sY-peptides with respect to either all peptides observed, or phosphopeptides only.

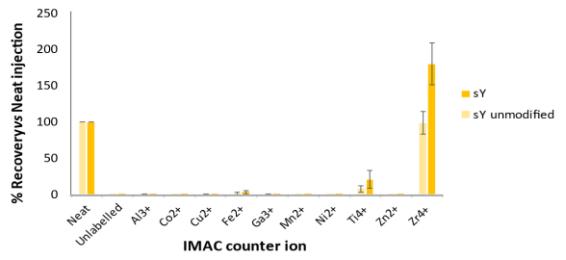
As previously observed for the synthetic peptides (Fig. 2A), sY peptide recovery from this mixture was much lower with the T5 condition than T4, while there was little variation ( $\pm$  <5%) in the relative recovery of either phosphorylated or non-phosphorylated peptides from BSA/casein. Consequently, a greater enrichment factor was observed for T4 than T5, whether normalised against all peptides (1.82 or 0.89 respectively), or just the phosphopeptide cohorts (1.19 or 0.61 respectively). While these results may be explained in part by the reduced acetonitrile content of T4, we hypothesise that the glycolic acid may be acting to reduce non-specific binding of the acidic residues in the sY peptides given the comparative recovery of the unmodified synthetic peptides. Thus, we were able to partially enrich for sY-peptides, and believe that this is primarily due to the high acidic content of TPST1/2-consensus containing peptides.

We next investigated sY-peptide enrichment using IMAC and a panel of 10 metal counter ions (Al<sup>3+</sup>, Co<sup>2+</sup>, Cu<sup>2+</sup>, Fe<sup>2+</sup>, Ga<sup>3+</sup>, Mn<sup>2+</sup>, Ni<sup>2+</sup>, Ti<sup>4+</sup>, Zn<sup>2+</sup> and Zr<sup>4+</sup>). Under relatively mild conditions (50% ACN, 0.1% TFA), Zr<sup>4+</sup> exhibited by far the most efficient recovery (~180 % cf C18 enrichment) across all sulfopeptide standards, with Ti<sup>4+</sup> being the only other counter ion capable of enriching multiple sulfopeptides, albeit rather inefficiently, with a recovery of ~20% (Fig. 3A). In contrast with previously published data which used Ga<sup>3+</sup> and Fe<sup>3+</sup>,<sup>26,27</sup> we did not observe sulfopeptide capture using Ga<sup>3+</sup>-IMAC and only minimal recovery (~3%) was seen with Fe<sup>3+</sup>-IMAC.

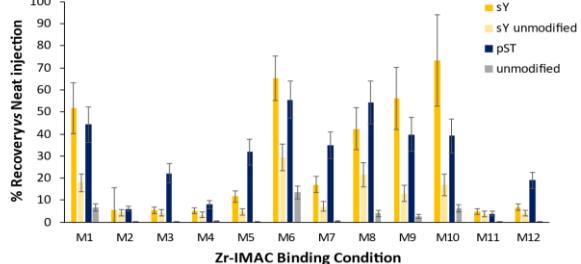
Given the comparatively high recovery of sY peptides with Zr<sup>4+</sup>-IMAC, we evaluated the effect of different binding conditions on the recovery and enrichment of sY peptides in a more complex BSA/casein/synthetic peptides mixture, as performed above for TiO<sub>2</sub>. Overall, we evaluated 12 different conditions, altering the concentration of ACN (50%, 80%), TFA (0, 0.1%, 5%), acetic acid (0 or 0.1%) and/or glycolic acid (0, 0.1 M, 1 M) (Fig. 3B). Optimal sulfopeptide recovery with Zr<sup>4+</sup>-IMAC was obtained in the presence of 80% ACN, 0.1 % acetic acid (M10) (~74%  $\pm$  20%, outperforming optimal TiO<sub>2</sub> conditions). Notably, the relative recovery of phosphopeptides (~39%) and sY-unmodified peptide (~17%) was also much reduced compared with optimal TiO<sub>2</sub>-enrichment conditions (Fig. 2B), resulting in enrichment factors of either 3.55 or 1.88 when considering either all peptides, or specificity with regard to phosphopeptide enrichment.

Increasing the acetonitrile concentration (from 50% to 80%) consistently increased the efficiency of sY-peptide recovery (cf M7/M8, M9/M10, M4/M11, M5/M12). However, the type and

A



B



Components	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
ACN (%)	50	80	80	80	80	80	50	80	50	80	50	50
TFA (%)	0.1	5	5	0.1	0.1	0.1	-	-	-	-	0.1	0.1
Acetic acid (%)	-	-	-	-	-	-	0.1	0.1	0.1	0.1	-	-
Glycolic acid (M)	-	1	0.1	1	0.1	-	0.1	0.1	-	-	1	0.1
sY Enrichment Factor	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
/pST only	1.17	0.89	0.24	0.66	0.37	1.17	0.49	0.78	1.41	1.88	1.20	0.35
/all peptides	2.25	1.53	0.61	1.34	0.95	1.99	1.2	1.59	3.05	3.55	1.82	0.88

**Figure 3: Development of an IMAC based enrichment protocol for sY-peptides.** A) Efficiency of recovery of the panel of sY- and unmodified peptides was evaluated for 10 IMAC counter-ions as indicated. Recovery was determined by normalising to an equal injection of ‘unenriched’ input material subjected to C18 reversed phase clean up. Recovery of the 8 sY- and 9 unmodified peptides is an average (mean  $\pm$  S.D.) of all equivalent (non-)modified peptides, including methionine oxidised variants which we have previously shown has little effect on relative signal intensity<sup>1</sup>. B) Optimisation of Zr<sup>4+</sup> recovery of sY-peptides. A mixture of trypsin digested BSA, casein- $\alpha$ S1/ $\alpha$ S2/β and the sY-peptide panel were subjected to enrichment with Zr<sup>4+</sup> IMAC using different loading solutions as indicated. All samples were subjected to C18 clean-up prior to LC-MS/MS analysis. Recovery (mean  $\pm$  S.D.) was determined for 8 sY-, 9 unmodified synthetic peptides, as well as 9 pST peptides from casein, and 19 unmodified peptides from BSA and casein.

concentration of acid had a greater effect on the efficiency of enrichment, with acetic acid being preferential for sY and TFA being preferential for pY-peptides (M7/M8, M9/M10 versus M4/M11, M5/M12). We also observed a reduction in sY enrichment factors with a decrease in glycolic acid concentration (from 1 M to 0.1 M), primarily due to an increase in the recovery of unwanted phosphopeptides (compare M2/M3, M4/M5, M11/M12). That being said, the overall recovery of sY peptides was substantially reduced in the presence of any concentration of glycolic acid, negating the potentially positive effect on co-enrichment of phosphopeptides (M4/M6, M1/M11, M7/M9, M8/M10) (Fig. 3B).

### Challenges associated with accurate site localization of sulfotyrosine within tryptic peptides.

While there have been a number of studies exploring different strategies for sulfopeptide characterisation and site localisation (positive *versus* negative ion mode; fragmentation),<sup>27, 28, 39, 43, 44, 64, 67, 68, 88, 89</sup> the overall utility of these findings in terms of applicability for HTP proteomics studies remains unknown;

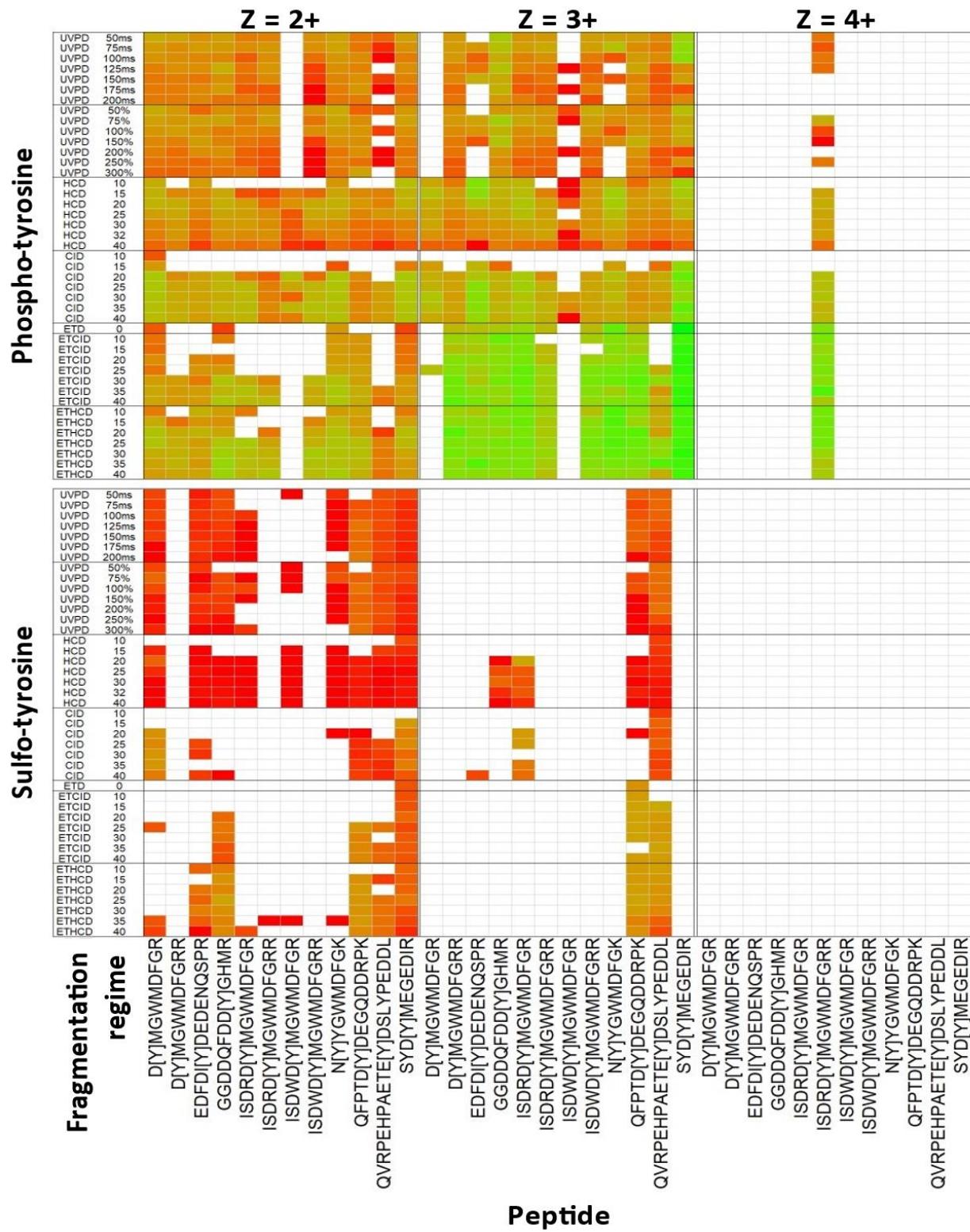
they either rely on instrumentation not commercially available, or focus on a very small number of non-tryptic peptides and are thus not representative of typical proteomics samples.

To better understand the ability to confidently localize sY sites in tryptic peptides (*i.e.* the generation of site-determining product ions), we characterised product ions generated from our synthetic panel of 12 sY tryptic peptides derived from known protein modifications using all potential fragmentation regimes available on a standard Fusion Lumos Tribrid instrument (Thermo Fisher), namely HCD, CID, ETD, EThcD, ETcID and UVPD. We also evaluated 12 analogous pY peptides, allowing us to compare site localisation confidence for both covalent modifications side-by-side. Considering the multiple settings available for each fragmentation regime, we thus investigated a total of 43 fragmentation conditions (Fig. 4). sY and pY peptide libraries were analysed by LC-MS/MS as separate pools, and searched with COMET to aid analysis. For each fragmentation condition, peptide and charge state, the tandem mass spectrum with the highest COMET score was subjected to further analysis, quantifying the number of observed product ions retaining the covalent modification as a function of the number of theoretically possible product ions (Fig. 4). Unambiguous site localisation ideally requires the generation of product ions that retain the covalent PTM. This is particularly important for sulfation, given the propensity for neutral loss of Δ80 Da resulting in ions that cannot be distinguished from those that would otherwise be unmodified. Data for ion types correlating with fragmentation around a particular residue (*e.g.* a<sub>4</sub>b<sub>4</sub>/y<sub>(n-4)</sub>, y<sub>8</sub>/y<sub>8</sub><sup>2+</sup>/z<sub>8</sub>) was compiled and treated as a single entity for the purpose of site localisation.

As shown in Fig. 4, there is stark disparity in the ability to localise the sites of modification on pY- versus sY- peptides, under all the conditions evaluated. Irrespective of the fragmentation strategy employed, and the energetics/time used, very few sulfonate-retaining product ions were observed for sY containing-peptides. ET(hc/ci)D yielded a higher proportion of sY- (and pY-) site determining ions than HCD for those peptides/conditions where product ions were observed, making electron-mediated fragmentation (EThcD at 25% NCE) optimal for localisation of the sY sites in this peptide panel. However, even for these peptides, the relative proportion of site-determining ions was substantially lower than for the pY-peptide equivalents and it is likely that the reduction in sY-peptide ion charge state (Fig. 1B, Fig. 4) prohibited efficient fragmentation (and thus identification) by EThcD/ETcID.

In contrast with HCD, very few sulfopeptides were identified following CID. Where they were identified, there was almost complete neutral loss of sulfonate, with the exceptions of D[Y]MGWMDFGR and SYD[Y]MEGEDIR. While HCD permitted sulfopeptide identification (for 9/12 peptides), the complete loss of 80 Da, even at low NCE (15%), meant that no site determining ions were observed.

Our phosphopeptide data contrasts significantly with that observed for the equivalent sulfopeptides, with varying degrees of phosphate retention seen for all peptides and charge states across all fragmentation strategies employed (Fig. 4). For doubly protonated pY-peptide ions, overall phosphate retention was highest with CID\_35, or EThcD\_25, while the equivalent 3+ ions benefitted substantially from ET(hc/ci)D, a finding well supported for phosphopeptides in previous studies.<sup>42, 58</sup> Both fragmentation regimes generally outperformed HCD in terms of the relative proportion of phosphopeptide-retaining product



**Figure 4: Comparison of site determining ions for sY and pY peptide panel using different fragmentation regimes.** The panel of 12 sY- and pY-peptides pools were subjected to LC-MS/MS on the Fusion Lumos Tribrid mass spectrometer (Thermo Fisher) using fragmentation conditions as detailed. Ranging normalised collision energies (NCE) were applied for CID, HCD, ETcID and EThcD as denoted by 'X' (where X = NCE %). ETD component was charge state calibrated. UVPD activation time was either calibrated to molecular weight (%) or manually-set (ms). Data were searched with COMET and the highest scoring PSM selected for further investigation. Heatmap shows log2 ratio of the number of observed MS2 product ions containing the known modification mass shift ( $\Delta 80$  Da) normalised to the total number of potential PTM-containing product ions. Ions correlating to fragmentation at the same position in the peptide (e.g. a<sub>4</sub>/b<sub>4</sub>), y<sub>8</sub>/y<sub>8</sub><sup>2+</sup>/z<sub>8</sub>) were collapsed into a single entry. Modification site is depicted by '[Y]', and charge states are visualised separately. Green = all potential mass shift containing product ions detected; red = no potential mass shift containing product ions detected, white = no MS/MS spectra were confidently identified.

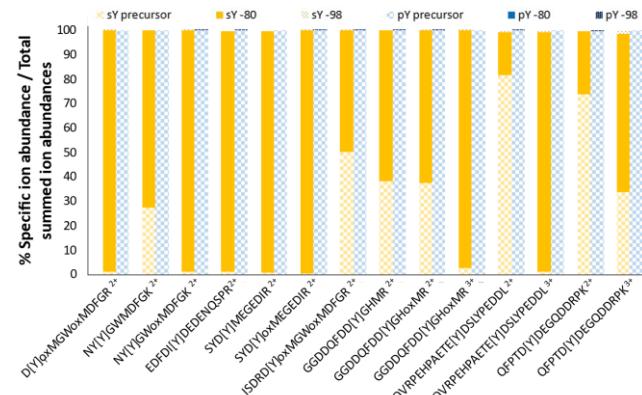
ions across all charge states. Interestingly, our observations with UVPD in terms of both automated peptide identification and generation of site localizing ions broadly mirrored the results with HCD for both sY- and pY-peptides. However, UVPD was highly inefficient for peptide fragmentation, requiring long irradiation times (>100 ms) and generating product ions of low intensity, irrespective of peptide sequence, charge or modifications status. sY-peptides were generally well identified with UVPD (8 out of 12 sY-peptides). However, sulfonate loss was extensive and it was not possible to localise the precise sites of sulfation on our sY-peptide panel.

### sY- and pY-peptides can be discriminated based on precursor neutral loss with HCD at low energy (10% NCE).

A number of strategies have been introduced to help distinguish sY from pY in peptides and proteins. These include a subtractive approach based on phosphatase treatment and acetylation of free (unmodified) tyrosine residues.<sup>67</sup> However, subtractive analytic approaches fundamentally rely on both complete phosphate removal and subsequent chemical modification of unmodified tyrosine residues, where inefficiencies in either step will result in mis-identification. The presence of additional labile tyrosine PTMs (such as nitration) also increases the potential of sY mis-identifications. Adduction of sY with trace metal ions (from LC solvents and ESI emitters) has also been reported to improve sY peptide identification and site localisation.<sup>64</sup> As well as increasing the average charge state (thus increasing ETD/EThcD efficiency), metal adducted sY-peptides have been reported to be more likely to retain the sulfate moiety (permitting localization by ETD mediated fragmentation regimes). However, while the relative proportion of Na<sup>+</sup>/K<sup>+</sup> adducts of the 2 sY-peptides reported in a previous study were relatively high, we failed to observe any sY-peptide metal-ion adduction following manual interrogation and open PTM searching of our data. Unfortunately, addition of Na<sup>+</sup>/K<sup>+</sup> salts into LC-MS systems compromises instrument performance and is thus not a practical solution.

Our observations (Fig. 4), and that of others<sup>28, 90</sup>, reveals a marked difference in the propensity of PTM neutral loss in both HCD and CID between sY- and pY-peptides. We thus hypothesised that we could exploit this feature to develop a low energy neutral loss EThcD triggering approach that could discriminate near-isobaric PTMs and enable sY-peptide identification from complex mixtures containing pY peptides. To test this, we quantified HCD precursor ion neutral loss at 10% NCE for sY- and pY-peptides, quantifying -80 Da (SO<sub>3</sub>/HPO<sub>3</sub>) and -98 Da (H<sub>2</sub>SO<sub>4</sub>/H<sub>3</sub>PO<sub>4</sub>) mass shifts (Fig. 5). For each peptide spectrum match (Fig. 4), the relative abundance of the precursor or neutral loss precursor ions were calculated as a percentage of total product ion current (Fig. 5). The predominant 10% NCE HCD product ion observed across our sY-peptide panel equated to loss of 80 Da (sulfonate) from the precursor, accounting for 20–100% of the MS/MS ion current (median = ~85%). Little to no peptide backbone cleavage was observed, in agreement with the finding that no peptides were identified with the search engine under this condition (Fig. 4). In contrast, pY-peptide exhibited minimal loss of either 80 or 98 Da (<1%) under the same conditions (Fig. 5) (Supp. Fig 1).

Interestingly, while the degree of neutral loss for sY peptide ions appeared to be dictated by the ratio between the number of



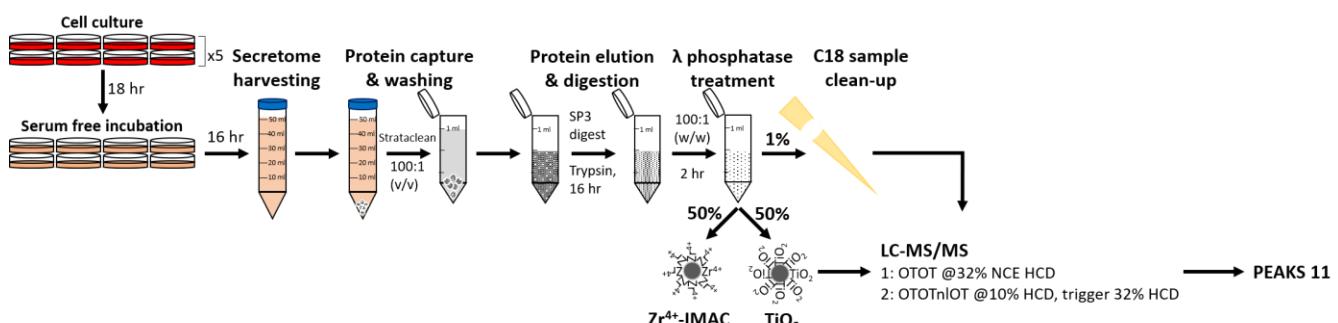
**Figure 5: Neutral loss propensity of sY- versus pY-peptides at 10% NCE HCD fragmentation.** The most intense PSM (determined from the mgf file) for each sY and pY peptide at different charges (including oxidised methionine variants) were selected for quantitation (14 peptide ions). The relative abundance (%) of the precursor ion (check), precursor ion -80 Da (block), or precursor ion -98 Da (dotted) was calculated as a proportion of the summed intensity of all ions in the MS2 spectrum. sY-containing peptides in yellow; pY peptides in blue.

basic residues and the charge state, the trend was the opposite to that observed for phosphopeptides.<sup>91</sup> Where the charge state was greater than the number of basic residues (H/K/R), contributing to a ‘mobile proton environment’ (MPE), we observed near complete neutral loss of 80 Da. However, the degree of precursor ion neutral loss was comparatively reduced when the charge state was smaller than the number of basic sites (-80 Da neutral loss: QFPTD[Y]DEGQDRPK: 2+ ~20%, 3+ ~65% and QVRPEHPAETE[Y]DSLYPEDDL: 2+ ~15%, 3+ ~95%). These data suggest that the charge-directed fragmentation mechanisms that appear to drive phosphopeptide neutral loss<sup>91, 92</sup> are not directly applicable to sulfopeptides, whose fragmentation (propensity for neutral loss) is also likely to be affected by the differences in electronegative and hydrogen bonding capabilities of the sulfonate moiety.

Having demonstrated the ability to distinguish sY- and pY-peptides based on HCD precursor ion neutral loss at 10% NCE, we next sought to implement this for sY-peptide identification using a neutral loss triggering approach in a mixture. Given optimal peptide identification across all charge states with HCD (32% NCE) (Fig. 4), we used the 80 Da precursor ion loss at 10% NCE HCD to trigger 32% NCE HCD on the same precursor to permit peptide identification, analysing a 1:1 ratio of our sY- and pY-peptide panel. Using this approach, we were able to positively identify all of the sY-peptides, and did not identify any of the pY-containing peptides, confirming the utility of this low energy HCD triggering strategy. To our knowledge, this is the first reported case of utilising low collision energy HCD fragmentation to efficiently distinguish sY- and pY-peptides by MS in a single run.

### Application of our HTP sulfopeptide analytical pipeline to identify the secreted ‘sulfome’ of HEK-293 cells.

sY-proteins are generated in the Golgi compartment, and are predominantly destined for secretion or cell membrane localisation: of the 33 validated human sY-proteins in UniProt, 19 are secreted and 14 are membrane bound.<sup>7</sup> In order to test the capabilities of our optimised workflow for sTyr, we evaluated the



**Figure 6: Workflow for the identification of sulfated peptides from a HEK-293 cell secretome.** Following incubation in serum-free media, the secretome was harvested, and proteins captured with Strataclean beads. Eluted proteins were subject to SP3-based trypsin proteolysis and treatment with  $\lambda$  phosphatase. 1% was subjected to C18 sample clean-up prior to LC-MS/MS analysis. The remainder was subject to enrichment using either the optimised  $Zr^{4+}$ -IMAC or  $TiO_2$  protocols. All samples were analysed using i) high resolution (Orbitrap, OT) data-dependent acquisition (DDA) with 32% NCE HCD, or ii) the neutral loss triggering strategy where loss of 80 Da at 10% NCE HCD invoked precursor fragmentation with 32% NE HCD. Data were analysed using PEAKS 11.

secretome of the adherent HEK-293 model cell system (Figure 6). After an 18 h incubation in serum-free medium, the HEK-293 cell secretome was collected, purified and prepared using Strataclean resin and an SP3-based trypsin digestion protocol (adapted from<sup>93</sup>) and treated with protein phosphatase.

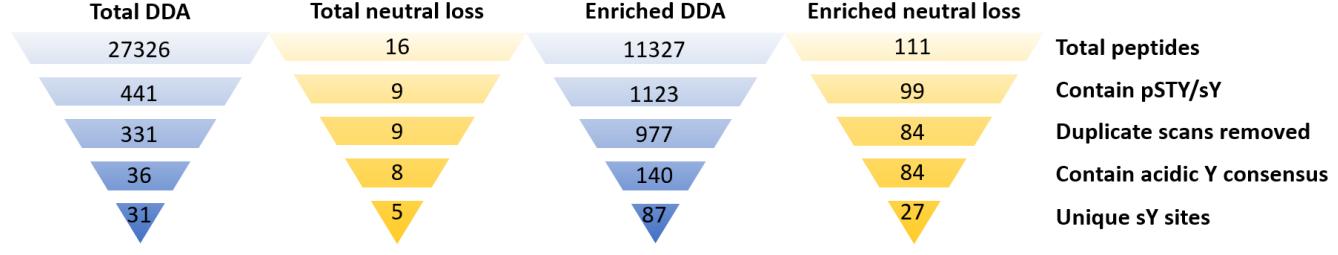
Given differences in pY and sY peptide enrichment between  $Zr^{4+}$ -IMAC and  $TiO_2$  protocols, we elected to compare both optimised enrichment protocols to investigate the cellular secretome. However, initial analysis of enriched material from agarose coated (Purecube)  $Zr^{4+}$ -IMAC resin using a standard data-dependent acquisition (DDA) pipeline revealed extensive binding of non-modified peptides. We thus employed MagReSyn®  $Zr$ -IMAC HP (ReSyn Biosciences), which exhibits substantially reduced non-specific peptide binding.<sup>94</sup>

To allow us to evaluate (i) the efficiency of sY- versus pY-peptide enrichment using the  $TiO_2$  and MagReSyn®  $Zr^{4+}$ -IMAC HP resins, and (ii) the neutral loss triggered-strategy for sulfopeptide identification, we acquired MS/MS data using a DDA pipeline alongside our neutral-loss triggered approach (Figures 6, 7).

To confirm effective purification of the secretome, 1% of the sample was subject to DDA analysis using HCD (NCE 32%). Performing an initial open PTM search (PEAKs PTM) we identified a high degree of Met oxidation (~5,900 peptide spectral matches (PSMs)) and Asn/Asp deamidation (~5,500 PSMs). Met oxidation and Asp/Asn deamidation, as well as Ser/Thr/Tyr phosphorylation and Tyr sulfation were thus included as variable modifications in subsequent searches. From the total

secretome DDA analysis, we identified 27,326 peptides from 2,695 proteins at a 1% FDR (Figure 7). Gene Ontology (GO) analysis with DAVID (Supp. Table 2) revealed that 36% of the identified proteins were defined as being localised to extracellular exosomes (Benjamini-Hochberg corrected p-value <1.6e<sup>-163</sup>), 29% as membrane-bound (Benjamini-Hochberg corrected p-value = 6.49e<sup>-56</sup>) and 44% as cytosolic (Benjamini-Hochberg corrected p-value = 7.72e<sup>-101</sup>). Of these ~27k peptides, 441 (<2%) were annotated as containing either pS/T/Y or sY (330 and 141 sites respectively) (Figure 7). A substantive proportion (~25%) of these annotations were from duplicate scan numbers where the same scan generated PSMs annotated as containing either pSTY or sY, or differences in deamidation status of Asn/Asp. Manually filtering this list, retaining the highest scoring scan number unique PSMs yielded 331 identifications. Enzymatic deposition by TPST1/2 is known to occur on Tyr residues within an acidic consensus (with Asp or Glu localized at either the +1 or -1 position), therefore we further filtered these identifications based on an acidic consensus, revealing 36 peptides, 31 of which were unique sites of Tyr sulfation (Figure 7).

A total of 11,327 peptides were identified following DDA analysis of both enriched samples. No substantive differences in terms of isoelectric point or *m/z* distribution were observed for those PSMs identified from the total or the enriched secretome samples (Supp. Fig. 2). Ten percent of the enriched samples (1,123) were annotated as phosphorylated and/or sulfated (898 and 304 respectively). While markedly lower than the level of enrichment typically seen from a cell extract using standard phosphopeptide enrichment strategies<sup>58, 95-97</sup> (usually



**Figure 7: Evaluation of sulfopeptide enrichment and neutral loss triggering data acquisition for sulfopeptide identification.** Secretome samples were analysed using either a data-dependent acquisition (DDA) strategy (32% NCE HCD) or the 10% HCD neutral loss triggered strategy for HCD acquisition (32% NCE), with or without enrichment using  $TiO_2$  /  $Zr^{4+}$ -IMAC enrichment (Figure 6). Listed are the total number of peptides identified, those containing either pSTY or sY before or after removal of duplicate scan PSMs. Peptide lists were subsequently filtered for those where the site of modifications was in an acidic consensus (with D/E at the +1 or -1 position relative to Y), and then for unique sY sites.

~85-90% using  $\text{TiO}_2$  in our hands), this equates to over a 6-fold increase in the identification of modified peptides from the secretome sample, and a ~2.2-fold increase in the peptides annotated as being sulfated. Removal of duplicate scans (977 identifications) and filtering for an acidic consensus left some 140 peptides with 87 sites (Figure 7).

Performing our neutral loss triggered MS acquisition method with the same sample significantly reduced the number of peptides identified in both the total (0.06%) and enriched (1.0%) samples. The numbers of peptides identified in the enriched DDA experiment compared to the enriched neutral loss analysis indicates that the enrichment strategy alone is insufficient for enhanced sulfopeptide identification (in agreement with our synthetic peptide analysis, Figure 3), but that enrichment does serve to improve the sensitivity of the MS acquisition.

Of the two different enrichment strategies, the sample prepared using  $\text{Zr}^{4+}$ -IMAC HP resin triggered the greatest number of MS/MS spectra (demonstrating higher likely enrichment of sY-peptides) following 10% NCE HCD, yielding 977 MS2 spectra; the  $\text{TiO}_2$ -enriched sample resulted in 754 triggering events. Despite >1700 total triggering events, only 327 spectra were matched to a peptide sequence (~19%) suggesting issues associated with fragmentation, and/or ion intensity that compromised identification. Surprisingly, of these 327 PSMs, 319 (~98%) were from the  $\text{TiO}_2$  sample, with only 8 (~2%) spectra from  $\text{Zr}^{4+}$ -IMAC HP resin. We are currently at a loss to explain this marked difference in the ratio of triggering events to peptide identifications, noting that the  $\text{TiO}_2$ -enriched samples generally yielded ions of lower average intensity with a lower charge state distribution (Supp. Fig. 3). In terms of total peptide identifications – this equates to 111 for  $\text{TiO}_2$  of which 62 were annotated as sY-containing (54% enrichment efficiency), and 7 for  $\text{Zr}^{4+}$ -IMAC of which 6 were annotated as sulfated (86% enrichment efficiency). Applying the acidic-Tyr filter and concatenating for the highest scoring PSM per scan for our enriched neutral loss triggered dataset (Figure 7), we identified 84 modified peptides, all of which contained a Tyr residue within an acidic motif. The resultant 27 unique sites of modification mapped to 23 proteins (Table 1). To note, only 5 of these sites (from 4 proteins) were found in the unenriched DDA experiment.

Considering that our neutral loss triggering MS acquisition method can efficiently discriminate sY- from pY peptides, we attempted to utilise confidence scores ( $-\text{LogP}$ ) and  $\Delta\text{ppm}$  to distinguish between pSTY/sY of these duplicated scans (Supp. Fig 4). The majority of PSMs were annotated as being sulfated, with or without N/Q deamidation. Although  $\Delta\text{ppm}$  was generally lower for those PSMs with a higher  $-\text{Log10P}$  value, this was not consistent. PSMs annotated as deamidated generally had both a higher  $\Delta\text{ppm}$  and were lower scoring. We did not observe any score or ppm-related features that could confidently differentiate assigned PSMs as being either phosphorylated or sulfated. We thus concatenated the data, removing the lower scoring duplicate PSM to retain a single identification per scan – this yielded a final dataset of 23 different peptides and 30 pSTY/sY annotations from the HEK-293 secretome (Table 1).

We next compared our dataset of 27 modified peptides with the human Tyr ‘Sulfome’ in UniProt. Interestingly, only Y217 on Amyloid beta precursor protein (A4, P05067) and Y1604 on collagen alpha-1 (CO5A1, P20908) have been previously

identified as sites of sulfation. While other sites of sY have been reported in both of these proteins, these were not identified in our data. A number of these, including Y262 on A4 is situated in extremely acidic regions which likely do not lend themselves to trypsin-based identification.

Six of the 27 peptides contain multiple sites of (+80 Da) modification, being annotated as either doubly sulfated, doubly phosphorylated, or a mixture of the two. Focussing in the first instance on a doubly ‘phosphorylated’ peptide from PCSK5 (Proprotein convertase subtilisin/kexin type 5), two of the five Tyr residues have an Asp at -1, and are localised within a highly acidic region (the others being clustered at the peptide C-terminus). It is also worth noting that the tandem MS data that generated these identifications was triggered based on neutral loss of 80 Da at 10% NCE HCD, which we have shown does not induce phosphate loss (Fig. 5, Supp. Fig. 1). Like the mixed PTM containing peptides (sY and pS/T/Y) identified from APLP2 (amyloid  $\beta$  precursor like protein 2), CO5A1 (collagen  $\alpha$ -1) and STC2 (stanniocalcin-2), it is therefore likely that (at least) one of these sites is in fact sulfated; we assume that neutral loss from the sulfated residue under lower energy conditions triggered MS/MS data acquisition and identification of a peptide that is **also** phosphorylated. Likewise, the singly modified site on STC2 (which is embedded in an EYxD motif) is likely also sulfated. To investigate this further, we interrogated the 10% HCD scans for these precursors, determining that the MS2 spectra from both APLP2 and STC2 contained a neutral loss peak equivalent to a single sulfated residue (-80 Da), while spectra for the secreted metalloproteinase BMP1, MCFD2, CO5A1, PCSK5 contain losses equivalent to (80 and) 160 Da, suggestive of two sites of tyrosine sulfation. This agrees with information in UniProt that identifies Y1601 on CO5A1 as an additional site of sulfation, strongly indicative of mis-identification of sulfation sites as phosphosites by the search engine.

Given our synthetic peptide panel data (Fig. 4), an unexpected finding from our final secretome-derived (non-unique) dataset was the large number of ions (61 out of the 84 peptides i.e. >60%) that were observed with a charge state  $\geq 3$ . The sulfation-induced reduction in charge state observed for the panel meant that we had expected a substantive proportion of those peptides observed using our sulfoproteome pipeline (Figure 6) to be doubly protonated. In fact, while this was true across all peptides in the DDA dataset from the enriched secretome, this was not the case for the NL triggered dataset, with the majority of peptide ions appearing as 3+ species (Supp. Fig. 5A). We believe that this is because the peptides identified with the NL triggering approach are much larger (and thus of higher mass) compared to those from the enriched DDA set (Supp. Fig. 5B, C) or our peptide panel; this can be explained by virtue of the fact that the sY-containing (NL-triggering) peptides from the secretome contained a substantially higher proportion of acidic residues (9.0 D/E residues on average) compared with those peptides in the total protein DDA set (average of 1.7 D/E residues) (Supp. Fig. 5D), and consequently a lower relative proportion of K/R residues.

To validate the secretome data obtained using our pipeline for identification of Tyr sulfation, we expressed and affinity precipitated the novel sulfated substrates H6ST1 and H6ST2 (Golgi-localized Heparan-sulfate 6-O-sulfotransferase 1/2) from HEK-293T cells, and subjected both proteins to *in vitro* PAPS-dependent sulfation with recombinant TPST1/2. We also immunoprecipitated the related isoform H6ST3, which contains

Accession	Gene name	Description	Peptide	-10LogP	ppm	m/z	z	Search engine defined PTM	Putative site of sulfation	Prior knowledge
P05067	A4	Amyloid beta precursor protein	GVEFVCCPLAEESDNVDSDAEEEDSDVWWGG ADTDYADGSEDK	200	-2.5	1227.464	4	Cm;Cm;Sulfation	Y217	UniProt- sY (sequence analysis)
Q06481	APLP2	Amyloid beta precursor like protein 2	IIGSVKEEEEDDEEEEEEDEEDDYDVYK IIGSVKEEEEDDEEEEEEDEEDDYDVYK	84.08	8.8	1244.822	3	Sulfation	Y233	PSP - 1 HTP (pY233)
P13497	BMP1	Bone morphogenetic protein 1	LDLADTYDLAEEEDSPELNLYKDPCK LDLADTYDLAEEEDSPELNLYKDPCK	103.32	1.3	1084.763	3	Sulfation;Sulfation;Cm	Y28; Y30	Y30: PSP - 1 HTP (pY30)
O43852	CALU	Calumenin	VHNDAQSFDYDHAFLGAEAK	100.52	4.6	853.6876	3	Sulfation	Y47	UniProt- pY47 (computational/experimental evidence); PSP - 14 HTP (pY47)
P20908	CO5A1	Collagen alpha-1	NIDASQLLDDGNGENYDVYADGMEEIFGSLNSLK	85.5	5.4	1934.298	2	Deamidation (NQ); Phosphorylation (STY);Sulfation	Y1604	UniProt- sY manual assertion (sequence analysis)
O75718	CRTAP	Cartilage-associated protein	ENIMDDDEGEVVEYVDDLLEETS	35.28	9	1490.617	2	Phosphorylation (STY)	Y390	
Q9UBP4	DKK3	Dickkopf-related protein 3	EVPDEYEVGSFMEVR	82.22	3.8	997.9047	2	Sulfation	Y308	
P02751	FINC	Fibronectin	EDVDYHLYPH	40.11	4	684.2651	2	Sulfation	Y2187	
O00461	GOLI4	Golgi integral membrane protein 4	GREENHYYYYEEEEEEDGAAVAEK	99.2	3.8	882.0118	3	Sulfation	Y673	UniProt- pY (sequence similarity); PSP - 145 HTP (pY)
Q8NBJ4	GOLM1	Golgi membrane protein 1	ELEHNAEETYGENDENTDDKNNDEEQEVR	84.95	6.7	1201.472	3	Sulfation	Y641	PSP - 4 HTP (pY)
O60243	H6ST1	Heparan-sulfate 6-O-sulfotransferase 1	LRGEDDYYNMDENEASETDKQALAGNDR	200	-6.6	1112.781	3	Sulfation	Y351	PSP - 2 HTP (pY)
Q96MM7	H6ST2	Heparan-sulfate 6-O-sulfotransferase 2	EDADEPGRVPTEDYMSHIEK	48.59	-2.6	843.0226	3	Sulfation;Oxidation (M)	Y403	
Q14766	LTBP1	Latent-transforming growth factor beta-binding protein 1	EQNDNTSNTGNDYIGSVEK	41.31	9.3	1083.926	2	Deamidation (NQ); Sulfation;Deamidation (NQ)	Y597	
Q8NI22	MCFD2	Multiple coagulation factor deficiency protein 2	DGDKNNNDGYIDYAEFAK DGDKNNNDGYIDYAEFAK	109.84	-9.2	1076.873	2	Sulfation;Sulfation	Y135; Y138	PSP - 3 HTP (pY135); 1 HTP (pY138)
Q14165	MLEC	Malectin	KEEEEEEYDEGSNLKK	58.56	6.6	1147.962	2	Phosphorylation (STY); Deamidation (NQ)	Y239	PSP - 5 HTP (pY239)
Q9UHG2	PCS1N	ProSAAS	PRPPVYDDGPGDAAEAGDETPDVPDELLR	85.01	-3.8	1123.826	3	Sulfation	Y171	
Q92824	PCSK5	Proprotein convertase subtilisin/kinin type 5	VEDPTDDYGTEDYAGPCDPECSEVGCDGPGPDH CNDCLHYYYK	27.06	-3.6	1297.955	4	Phosphorylation (STY); Phosphorylation (STY);Cm;Cm;Cm;	Y627; Y631	
O00264	PGRC1	Membrane-associated progesterone receptor component 1	EALKDEYDDSLDLTAAQQETLSDWESQFTFK	89.69	-0.1	1235.209	3	Sulfation	Y139	PSP - 4 HTP (pY139)
Q8WZA1	PMGT1	Protein O-linked-mannose-beta-1,2-N-acetylglucosaminyltransferase 1	AISEANEDPEPEQDYDEALGR	83.65	5.9	1214.494	2	Sulfation	Y78	UniProt- pY78 (computational/experimental evidence)
O76061	STC2	Stanniocalcin-2	VGGGLGAQGPSSSEWEDEQSEYSDIR VGGGLGAQGPSSSEWEDEQSEYSDIR	84.9	4.4	1410.588	2	Phosphorylation (STY)	Y297	
O60279	SUSD5	Sushi domain-containing protein 5	DEAAHIDYEDNFPDDR	50.05	-6.3	967.3721	3	Phosphorylation (STY);Sulfation	Y208	
Q5JRA6	TGO1	Transport and Golgi organization protein 1 homolog	QGKPQSATDYSDPDNVDDGLFIVDIPK	95.14	2.8	1065.898	2	Sulfation	Y434	

**Table 1: Confidently identified sY-peptides from a HEK293 cellular secretome.** Detailed are the accession number, gene name, protein description, peptide sequence, -10LogP score, Δppm from theoretical *m/z*, the observed *m/z*, charge state (z) and identified PTMs as determined by the search engine. Listed is a concatenated version of all modified peptide identifications, maintaining only the highest -10LogP score peptide for a unique sY-site. Predicted (acidic) sY-site is **highlighted**. Cm – carbamidomethylation. All predicted sY sites were searched for prior annotation as being either phosphorylated or sulfated in UniProt and the PhosphositePlus (PSP) database [accessed July 2023]; HTP – details the number of times phosphorylation was reported in PSP in a high throughput study.

an analogous Tyr residue within the acid consensus sequence in its C-terminal region. After enzymatic reaction, proteins were digested with trypsin and analysed by LC-MS/MS using both DDA and our neutral loss triggered method. As well as confirming TPST1/2-dependent sulfation of H6ST1 and H6ST2 at the same sites on tryptic peptides observed from our HTP study (sTyr403 and sTyr597 respectively, Supp. Tables 3, 4), we also show that the related protein, H6ST3 can also be Tyr sulfated at sTyr285 and sTyr464, Supp. Tables 3, 4). Of particular interest, sTyr403, sTyr597 and sTyr 464 lie in a conserved acidic motif in the C-terminal region of H6ST1-3 outside of the catalytic domain (Supp. Fig. 6). Not only does this analysis validate the exploratory potential of our discovery pipeline for sTyr detection, but it also demonstrates how this information can be extrapolated to predict additional sites of modification in closely-related proteins such as H6ST3. Interestingly, only one of the two TPST1/2-dependent sulfation sites on H6ST3 contained an acidic residue in close proximity (sTyr464 – [EDYX]), suggesting that an acidic motif around the site of sulfation may not be an absolute requirement for TPST1/2 substrates.

As well as validating our HTP sulfomics pipeline, our enzymatic assays also revealed additional novel TPST1/2-dependent sulfation sites on proteins that co-immunoprecipitated with

H6ST2 and/or H6ST3: Gem-associated protein 5 (GEMIN5; sTyr992), tubulin α-1A/B/C chain (TUBA1A/TUBA1B/TUBA1C; sTyr161, sTyr432), Tubulin β chain (TUBB; sTyr[50 or 51], sTyr340), insulin receptor substrate 4 (IRS4; sTyr921) and heterogeneous nuclear ribonucleoprotein H (HNRNPH1; sTyr266). Gemin5 is thought to reside in the nucleoplasm and in specific nuclear bodies (Gemini of Cajal Bodies), as well as the cytoplasm, and has not previously been shown to be sulfated (or modified on Tyr992). Likewise, none of the other TSPT1/2 substrates have previously been reported to be sulfated. However, the *in vitro* sulfated Tyr sites identified on these proteins have previously been reported as being Tyr phosphorylated in PhosphoSitePlus in HTP proteomics screens, with IRS4 Tyr921 described as a putative substrate for the Tyr protein kinases Fer and IGFR1. Finally, we also identified what we believe to be the first auto-Tyr sulfation site on TPST1 (likely at Tyr326), which might itself be relevant to cellular TPST1/2 regulation or enzyme:substrate interactions, similar to the signaling paradigm in which tyrosine kinases are themselves controlled by tyrosine phosphorylation.

Together, these data provide strong evidence that our sulfomics pipeline is capable of defining novel human sulfopeptides in an HTP manner, and suggest Tyr sulfation cross-talk as a

potential regulatory modification between Golgi co-localized human heparan sulfate sulfotransferases and TPSTs.

## Conclusions

In this study, we investigated several strategies for the analytical discrimination of sulfopeptides from phosphopeptides by employing, to our knowledge, the largest panel of synthetic peptides based on tryptic human peptides derived from known sites of sY-modifications from cellular proteins. In developing a workflow specifically for HTP ‘sulfomics’ we characterised a number of sY-peptide discriminatory features that include RT shifts, susceptibility (or lack thereof) to treatment with protein phosphatase, conditions for sY-peptide enrichment and peptide fragmentation. We developed, and implemented the first HTP based ‘sulfomics’ pipeline and demonstrated its utility using a HEK-293 cellular secretome, where we identified a total of 21 novel (experimentally-determined) sY-containing proteins and 28 sY-sites, which expands the known human ‘sulfome’ by ~70%. In the process of validating H6ST1 and H6ST2 (and HS6ST3) as *in vitro* targets of the sulfotransferases TSPT1/2, we also identified additional *in vitro* substrates – GEMIN5, TUBA1, TUBB, IRS4 and HNRNPH1, which are themselves part of the known H6ST2/3 interactome. Interestingly, while the site on GEMIN5 has not previously been identified as being modified, there is extensive HTP evidence of phosphorylation of the other TSPT1/2 substrates. It thus remains to be seen whether i) these sites can be **both** sulfated and phosphorylated *in vivo*, ii) our identification of these specific sulfation sites *in vitro* is biased by the assay conditions, or iii) the phosphorylation site identifications reported in PhosphoSite Plus are actually mis-representations of a Tyr sulfation event. Since Tyr sulfation is believed to be irreversible, our findings raise the possibility that false identification of sTyr as pTyr, may be meaningful for several proteins, and urgently needs to be clarified for multiple proteins. In addition, since sTyr could act as a Golgi-based signal that competes with, or prevents, Tyr phosphorylation further along the secretory pathway, we are in the process of evaluating this combinatorial phenomenon further. Regardless, the identification of both glycan and protein sulfotransferases as sY-containing proteins potentially opens up a new research area for understanding the regulation and substrate targeting of these enzymes, similar in many ways to the phosphoregulatory paradigms established for protein kinases.<sup>98</sup>

While Fe<sup>3+</sup>/Ga<sup>3+</sup> IMAC has previously been used for sY peptide enrichment<sup>26,27</sup>, we were unable to validate these findings using our peptide library. However, we demonstrate the utility of Zr<sup>4+</sup>-IMAC and TiO<sub>2</sub> in acetic acid-based solutions for the semi-specific enrichment of sY-peptides in a complex (phospho) peptide mixture. Noting the differential enrichment of the relatively acidic non-modified peptides from our library (average pI ~4.5), compared with unmodified peptides from BSA/casein (average pI ~5.7), our evidence suggests that the acidic consensus (thought to be required for TPST1/2 directed sulfation) also contributes to the efficiency of sY-peptide enrichment. This is also reflected in the complex secretome analysis where the neutral loss triggered enriched peptides (sY-containing) identified were of much greater acidity than the total secretome DDA identifications (pI ~4.6 versus ~6.2).

In undertaking comprehensive MS analysis of tryptic sulfopeptides using CID, HCD, EThcD, ETcID and UVPD fragmentation regimes, we were able to quantify the generation of site-localising product ions and compare them with equivalent synthetic phosphopeptides. Taken together, our data indicates

that specific sulfosite localisation is poor, irrespective of fragmentation regime or conditions used. Our data with the commercially available UVPD configuration did not prove as promising as anticipated, based on previous work.<sup>43,44,39</sup> A number of factors likely contribute to this: difference in UVPD wavelength, energy and/or laser frequency, the use of positive rather than negative ion mode, and the fact that many of the studies reported to date use a handful of exemplar peptides. EThcD appeared to be the best overall for site localisation, but was inconsistent in its ability to identify sulfopeptides, in part because of the reduction in charge state. During the course of our studies we determined that low energy (10% NCE) HCD could serve to differentiate sulfopeptides from phosphopeptides, based on the lability of the sulfate group. We thus made use of the single feature that caused the greatest analytical challenge in terms of localisation to allow sY discrimination ‘on the fly’; a low energy (10% NCE) HCD neutral loss-triggering approach to define sulfopeptide-containing scans for subsequent identification. Our initial thinking based on the reduced ionisation efficiency of sulfopeptides was that application of EThcD to global ‘sulfome’ analysis would be compromised. However, our secretome-derived dataset suggests that the average increase in length and charge state of sulfopeptides may make a neutral-loss triggered EThcD strategy feasible for HTP sulfosite identification. However, this strategy will need to be evaluated for throughput given the additional time requirements of EThcD over HCD, and the necessity given that i) most peptides only contain a single Tyr residue and ii) the strong requirements for an acidic consensus for TSPT1/2 activity.

Thus, while the pipeline described herein facilitates the identification of sY-peptides from complex mixtures, unambiguously pinpointing the precise site Tyr of modification remains challenging. Specifically, the electronegativity of sY and issues associated with site localisation in positive ion mode MS (given current fragmentation strategies) continue to compromise definitive site determination, unless peptides only contain a single Tyr residue.

Recent reports have commented on the inherent metal ion-binding affinity of sY that might be exploited for localisation<sup>(99, 28 & 90)</sup>, with one report identifying metal adducted species as dominant when using standard positive ion mode LC-MS/MS methods (specifically Na<sup>+</sup>/K<sup>+</sup>, <sup>64</sup>). As such, we performed open PTM searches on our datasets in an attempt to identify some of the missing triggered peptides (86% of triggers). However, we failed to observe these species in our investigations. No additional sY-containing proteins were identified with the open PTM search, although differentially modified species of already identified sY peptides were observed. We presume this is a direct result of how PEAKs PTM performs its searches, requiring a protein to be identified by a first round database search with pre-defined search parameters, prior to performing a 2<sup>nd</sup> round mass shift search on a concatenated database containing only these identified proteins.<sup>100</sup> Since our neutral loss triggering method drastically reduces the number of peptides and proteins identified, this compromises the ability for PEAKs PTM to identify additional sulfated peptides in the absence of defining specific additional modifications. In summary, we believe that the analytical developments reported in this paper will serve as the basis of a new pipeline for HTP sulfome analysis, providing a resource to the community to better define the extent and roles of protein sulfation from biological samples.

## ASSOCIATED CONTENT

### Supporting Information

Supplementary Figures and Supplementary Table 3 – PDF  
Supplementary Tables 1, 2 and 4 - .xlsx

All mass spectrometry data has been deposited at the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifiers: Project Accession - PXD043713 and PXD043723.

## AUTHOR INFORMATION

### Corresponding Author

Claire E. Eyers - Centre for Proteome Research, Institute of Systems, Molecular & Integrative Biology, University of Liverpool, Crown Street, Liverpool, L69 7ZB, U.K. <https://orcid.org/0000-0002-3223-5926>; Email:ceyers@liverpool.ac.uk

### Author Contributions

LAD: performed all sample preparation, enrichment and mass spectrometry analysis. Contributed to design of experiments and manuscript writing. DPB: performed biochemical analyses – phosphatase experiments and sulfation assays. SP & EM: contributed to MS/MS fragmentation data analysis. PJB: supported mass spectrometry methods development. ARJ: contributed to MS/MS fragmentation data analysis. PAE: contributed to design of experiments and manuscript writing. CEE: contributed to design of experiments, analysis of data and manuscript writing.

The manuscript was written by LD and CEE with contributions from PAE, DPB and ARJ. All authors have given approval to the final version of the manuscript.

## ACKNOWLEDGMENT

This work is supported by from funding from the Biotechnology and Biosciences Research Council (BBSRC; BB/S018514/1, BB/M012557/1, BB/S017054/1 and BB/R000182/1). We thank Prof David Fernig (University of Liverpool) for useful discussions.

## REFERENCES

1. Eyers, C. E.; Simpson, D. M.; Wong, S. C.; Beynon, R. J.; Gaskell, S. J., QCAL--a novel standard for assessing instrument conditions for proteome analysis. *J Am Soc Mass Spectrom* **2008**, *19* (9), 1275-80.
2. Hardman, G.; Perkins, S.; Brownridge, P. J.; Clarke, C. J.; Byrne, D. P.; Campbell, A. E.; Kalyuzhnay, A.; Myall, A.; Eyers, P. A.; Jones, A. R.; Eyers, C. E., Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation. *The EMBO Journal* **2019**, *38* (21), e100847.
3. Attwood, P. V.; Piggott, M. J.; Zu, X. L.; Besant, P. G., Focus on phosphohistidine. *Amino Acids* **2007**, *32* (1), 145-56.
4. Besant, G. P.; Attwood, V. P.; Piggott, J. M., Focus on Phosphoarginine and Phospholysine. *Current Protein & Peptide Science* **2009**, *10* (6), 536-550.
5. Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Lo Muzio, L., The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med* **2017**, *40* (2), 271-280.
6. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanan, S., The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298* (5600), 1912-1934.
7. Consortium, T. U., UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **2022**, *51* (D1), D523-D531.
8. Baeuerle, P. A.; Huttner, W. B., Tyrosine sulfation of yolk proteins 1, 2, and 3 in *Drosophila melanogaster*. *J Biol Chem* **1985**, *260* (10), 6434-9.
9. Moore, K. L., The biology and enzymology of protein tyrosine O-sulfation. *J Biol Chem* **2003**, *278* (27), 24243-6.
10. Gregory, H.; Hardy, P. M.; Jones, D. S.; Kenner, G. W.; Sheppard, R. C., THE ANTRAL HORMONE GASTRIN. STRUCTURE OF GASTRIN. *Nature* **1964**, *204*, 931-3.
11. Niehrs, C.; Kraft, M.; Lee, R. W.; Huttner, W. B., Analysis of the substrate specificity of tyrosylprotein sulfotransferase using synthetic peptides. *J Biol Chem* **1990**, *265* (15), 8525-32.
12. Lee, R. W.; Huttner, W. B., (Glu62, Ala30, Tyr8)n serves as high-affinity substrate for tyrosylprotein sulfotransferase: a Golgi enzyme. *Proc Natl Acad Sci U S A* **1985**, *82* (18), 6143-7.
13. Braun, S.; Raymond, W. E.; Racker, E., Synthetic tyrosine polymers as substrates and inhibitors of tyrosine-specific protein kinases. *J Biol Chem* **1984**, *259* (4), 2051-4.
14. Ippel, J. H.; de Haas, C. J.; Bunschoten, A.; van Strijp, J. A.; Kruijzer, J. A.; Liskamp, R. M.; Kemmink, J., Structure of the tyrosine-sulfated C5a receptor N terminus in complex with chemotaxis inhibitory protein of *Staphylococcus aureus*. *J Biol Chem* **2009**, *284* (18), 12363-72.
15. Seibert, C.; Veldkamp, C. T.; Peterson, F. C.; Chait, B. T.; Volkman, B. F.; Sakmar, T. P., Sequential tyrosine sulfation of CXCR4 by tyrosylprotein sulfotransferases. *Biochemistry* **2008**, *47* (43), 11251-62.
16. Costagliola, S.; Panneels, V.; Bonomi, M.; Koch, J.; Many, M. C.; Smits, G.; Vassart, G., Tyrosine sulfation is required for agonist recognition by glycoprotein hormone receptors. *Embo J* **2002**, *21* (4), 504-13.
17. Hortic, G. L., Sulfation of tyrosine residues in coagulation factor V. *Blood* **1990**, *76* (5), 946-52.
18. Leyte, A.; van Schijndel, H. B.; Niehrs, C.; Huttner, W. B.; Verbeet, M. P.; Mertens, K.; van Mourik, J. A., Sulfation of Tyr1680 of human blood coagulation factor VIII is essential for the interaction of factor VIII with von Willebrand factor. *J Biol Chem* **1991**, *266* (2), 740-6.
19. Bundgaard, J. R.; Vuust, J.; Rehfeld, J. F., Tyrosine O-sulfation promotes proteolytic processing of progastrin. *Embo J* **1995**, *14* (13), 3073-9.
20. Farzan, M.; Mirzabekov, T.; Kolchinsky, P.; Wyatt, R.; Cayabyab, M.; Gerard, N. P.; Gerard, C.; Sodroski, J.; Choe, H., Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell* **1999**, *96* (5), 667-76.
21. Cormier, E. G.; Persuh, M.; Thompson, D. A.; Lin, S. W.; Sakmar, T. P.; Olson, W. C.; Dragic, T., Specific interaction of CCR5 amino-terminal domain peptides containing sulfotyrosines with HIV-1 envelope glycoprotein gp120. *Proc Natl Acad Sci U S A* **2000**, *97* (11), 5762-7.
22. Rodgers, S. D.; Camphausen, R. T.; Hammer, D. A., Tyrosine sulfation enhances but is not required for PSL-1 rolling adhesion on P-selectin. *Biophys J* **2001**, *81* (4), 2001-9.
23. Westmuckett, A. D.; Hoffhines, A. J.; Borghei, A.; Moore, K. L., Early postnatal pulmonary failure and primary hypothyroidism in mice with combined TPST-1 and TPST-2 deficiency. *Gen Comp Endocrinol* **2008**, *156* (1), 145-53.
24. Xu, P.; Xi, Y.; Wang, P.; Luka, Z.; Xu, M.; Tung, H.-C.; Wang, J.; Ren, S.; Feng, D.; Gao, B.; Singh, A. D.; Monga, S. P.; York, J. D.; Ma, X.; Huang, Z.; Xie, W., Inhibition of p53 Sulfoconjugation Prevents Oxidative Hepatotoxicity and Acute Liver Failure. *Gastroenterology* **2022**, *162* (4), 1226-1241.
25. Amano, Y.; Shinohara, H.; Sakagami, Y.; Matsubayashi, Y., Ion-selective enrichment of tyrosine-sulfated peptides from complex protein digests. *Analytical Biochemistry* **2005**, *346* (1), 124-131.
26. Balderrama, G. D.; Meneses, E. P.; Orihuela, L. H.; Hernández, O. V.; Franco, R. C.; Robles, V. P.; Batista, C. V. F., Analysis of sulfated peptides from the skin secretion of the Pachymedusa

dacnicolor frog using IMAC-Ga enrichment and high-resolution mass spectrometry. *Rapid Communications in Mass Spectrometry* **2011**, *25* (8), 1017-1027.

27. Capriotti, A. L.; Cerrato, A.; Laganà, A.; Montone, C. M.; Piovesana, S.; Zenezini Chiozzi, R.; Cavaliere, C., Development of a Sample-Preparation Workflow for Sulfopeptide Enrichment: From Target Analysis to Challenges in Shotgun Sulfoproteomics. *Analytical Chemistry* **2020**, *92* (11), 7964-7971.

28. Chen, G.; Zhang, Y.; Trinidad, J. C.; Dann, C., III, Distinguishing Sulfotyrosine Containing Peptides from their Phosphotyrosine Counterparts Using Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2018**, *29* (3), 455-462.

29. Monigatti, F.; Hekking, B.; Steen, H., Protein sulfation analysis--A primer. *Biochim Biophys Acta* **2006**, *1764* (12), 1904-13.

30. Kehoe, J. W.; Velappan, N.; Walbolt, M.; Rasmussen, J.; King, D.; Lou, J.; Knopp, K.; Pavlik, P.; Marks, J. D.; Bertozi, C. R.; Bradbury, A. R., Using phage display to select antibodies recognizing post-translational modifications independently of sequence context. *Mol Cell Proteomics* **2006**, *5* (12), 2350-63.

31. Byrne, D. P.; Li, Y.; Ngamlert, P.; Ramakrishnan, K.; Eyers, C. E.; Wells, C.; Drewry, D. H.; Zuercher, W. J.; Berry, N. G.; Fernig, D. G.; Eyers, P. A., New tools for evaluating protein tyrosine sulfation: tyrosylprotein sulfotransferases (TPSTs) are novel targets for RAF protein kinase inhibitors. *Biochem J* **2018**, *475* (15), 2435-2455.

32. Yun, H. Y.; Keutmann, H. T.; Eipper, B. A., Alternative splicing governs sulfation of tyrosine or oligosaccharide on peptidyl-glycine alpha-amidating monooxygenase. *J Biol Chem* **1994**, *269* (14), 10946-55.

33. Preobrazhensky, A. A.; Dragan, S.; Kawano, T.; Gavrilin, M. A.; Gulina, I. V.; Chakravarty, L.; Kolattukudy, P. E., Monocyte chemotactic protein-1 receptor CCR2B is a glycoprotein that has tyrosine sulfation in a conserved extracellular N-terminal region. *J Immunol* **2000**, *165* (9), 5295-303.

34. Farzan, M.; Schnitzler, C. E.; Vasilieva, N.; Leung, D.; Kuhn, J.; Gerard, C.; Gerard, N. P.; Choe, H., Sulfated tyrosines contribute to the formation of the C5a docking site of the human C5a anaphylatoxin receptor. *J Exp Med* **2001**, *193* (9), 1059-66.

35. Forbes, E. G.; Cronshaw, A. D.; MacBeath, J. R.; Hulmes, D. J., Tyrosine-rich acidic matrix protein (TRAMP) is a tyrosine-sulfated and widely distributed protein of the extracellular matrix. *FEBS Lett* **1994**, *351* (3), 433-6.

36. Gao, J.; Choe, H.; Bota, D.; Wright, P. L.; Gerard, C.; Gerard, N. P., Sulfation of tyrosine 174 in the human C3a receptor is essential for binding of C3a anaphylatoxin. *J Biol Chem* **2003**, *278* (39), 37902-8.

37. Benedum, U. M.; Lamouroux, A.; Konecki, D. S.; Rosa, P.; Hille, A.; Baeuerle, P. A.; Frank, R.; Lottspeich, F.; Mallet, J.; Huttner, W. B., The primary structure of human secretogranin I (chromogranin B): comparison with chromogranin A reveals homologous terminal domains and a large intervening variable region. *Embo J* **1987**, *6* (5), 1203-11.

38. Hortic, G.; Fok, K. F.; Toren, P. C.; Strauss, A. W., Sulfation of a tyrosine residue in the plasmin-binding domain of alpha 2-antiplasmin. *Journal of Biological Chemistry* **1987**, *262* (7), 3082-3085.

39. Robinson, M. R.; Moore, K. L.; Brodbelt, J. S., Direct identification of tyrosine sulfation by using ultraviolet photodissociation mass spectrometry. *J Am Soc Mass Spectrom* **2014**, *25* (8), 1461-71.

40. Zubarev, R. A., Reactions of polypeptide ions with electrons in the gas phase. *Mass Spectrom Rev* **2003**, *22* (1), 57-77.

41. Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E.; Coon, J. J., Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* **2007**, *79* (2), 477-85.

42. Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S., Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Analytical Chemistry* **2012**, *84* (22), 9668-9673.

43. Halim, M. A.; MacAleese, L.; Lemoine, J.; Antoine, R.; Dugourd, P.; Girod, M., Ultraviolet, Infrared, and High-Low Energy Photodissociation of Post-Translationally Modified Peptides. *J Am Soc Mass Spectrom* **2018**, *29* (2), 270-283.

44. Hersberger, K. E.; Håkansson, K., Characterization of O-sulfopeptides by negative ion mode tandem mass spectrometry: superior performance of negative ion electron capture dissociation. *Anal Chem* **2012**, *84* (15), 6370-7.

45. Riley, N. M.; Rush, M. J.; Rose, C. M.; Richards, A. L.; Kwiecien, N. W.; Bailey, D. J.; Hebert, A. S.; Westphall, M. S.; Coon, J. J., The Negative Mode Proteome with Activated Ion Negative Electron Transfer Dissociation (AI-NETD). *Mol Cell Proteomics* **2015**, *14* (10), 2644-60.

46. McAlister, G. C.; Russell, J. D.; Rumachik, N. G.; Hebert, A. S.; Syka, J. E.; Geer, L. Y.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J., Analysis of the acidic proteome with negative electron-transfer dissociation mass spectrometry. *Anal Chem* **2012**, *84* (6), 2875-82.

47. Wang, T.; Nha Tran, T. T.; Andreazza, H. J.; Bilusich, D.; Brinkworth, C. S.; Bowie, J. H., Negative ion cleavages of (M-H)(-) anions of peptides. Part 3. Post-translational modifications. *Mass Spectrom Rev* **2018**, *37* (1), 3-21.

48. Zuo, M.-Q.; Sun, R.-X.; Fang, R.-Q.; He, S.-M.; Dong, M.-Q., Characterization of collision-induced dissociation of deprotonated peptides of 4-16 amino acids using high-resolution mass spectrometry. *International Journal of Mass Spectrometry* **2019**, *445*, 116186.

49. Bowie, J.; Brinkworth, C.; Dua, S., Collision-Induced Fragmentations of the (M-H)- Parent Anions of Underivatized Peptides: An Aid to Structure Determination and Some Unusual Negative Ion Cleavages. *Mass spectrometry reviews* **2002**, *21*, 87-107.

50. Steinborner, S. T.; Bowie, J. H., The negative ion mass spectra of [M-H]<sup>-</sup> ions derived from caeridin and dynastin peptides. Internal backbone cleavages directed through Asp and Asn residues. *Rapid communications in mass spectrometry* **1997**, *11* (3), 253-258.

51. Boontheung, P.; Brinkworth, C. S.; Bowie, J. H.; Baudinette, R. V., Comparison of the positive and negative ion electrospray mass spectra of some small peptides containing pyroglutamate. *Rapid Communications in Mass Spectrometry* **2002**, *16* (4), 287-292.

52. Sugasawa, N.; Kawase, T.; Oshikata, M.; Iimuro, R.; Motoyama, A.; Takayama, M., Formation of c- and z-ions due to preferential cleavage at the NC bond of Xxx-Asp/Asn residues in negative-ion CID of peptides. *International Journal of Mass Spectrometry* **2015**, *383-384*, 38-43.

53. Byrne, D. P.; Shrestha, S.; Galler, M.; Cao, M.; Daly, L. A.; Campbell, A. E.; Eyers, C. E.; Veal, E. A.; Kannan, N.; Eyers, P. A., Aurora A regulation by reversible cysteine oxidation reveals evolutionarily conserved redox control of Ser/Thr protein kinase activity. *Science Signaling* **2020**, *13* (639), eaax2713.

54. Mohanty, S.; Oruganty, K.; Kwon, A.; Byrne, D. P.; Ferries, S.; Ruan, Z.; Hanold, L. E.; Katiyar, S.; Kennedy, E. J.; Eyers, P. A.; Kannan, N., Hydrophobic Core Variations Provide a Structural Framework for Tyrosine Kinase Evolution and Functional Specialization. *PLOS Genetics* **2016**, *12* (2), e1005885.

55. Daly, L. A.; Brownridge, P. J.; Batie, M.; Rocha, S.; Sée, V.; Eyers, C. E., Oxygen-dependent changes in binding partners and post-translational modifications regulate the abundance and activity of HIF-1 $\alpha$ /2 $\alpha$ . *Sci Signal* **2021**, *14* (692).

56. Byrne, D. P.; Clarke, C. J.; Brownridge, P. J.; Kalyuzhnyy, A.; Perkins, S.; Campbell, A.; Mason, D.; Jones, A. R.; Eyers, P. A.; Eyers, C. E., Use of the Polo-like kinase 4 (PLK4) inhibitor centrinone to investigate intracellular signalling networks using SILAC-based phosphoproteomics. *Biochem J* **2020**, *477* (13), 2451-2475.

57. Byrne, D. P.; Li, Y.; Ramakrishnan, K.; Barsukov, I. L.; Yates, E. A.; Eyers, C. E.; Papy-Garcia, D.; Chantepie, S.; Pagadala, V.; Liu, J.; Wells, C.; Drewry, D. H.; Zuercher, W. J.; Berry, N. G.; Fernig, D. G.; Eyers, P. A., New tools for carbohydrate sulfation analysis: heparan sulfate 2-O-sulfotransferase (HS2ST) is a target for small-molecule protein kinase inhibitors. *Biochem J* **2018**, *475* (15), 2417-2433.

58. Ferries, S.; Perkins, S.; Brownridge, P. J.; Campbell, A.; Eyers, P. A.; Jones, A. R.; Eyers, C. E., Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer. *J Proteome Res* **2017**, *16* (9), 3448-3459.

59. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551-67.

60. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P., A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **2012**, *30* (10), 918-920.

61. Huang da, W.; Sherman, B. T.; Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **2009**, *4* (1), 44-57.

62. Rapp, C.; Klerman, H.; Levine, E.; McClendon, C. L., Hydrogen bond strengths in phosphorylated and sulfated amino acid residues. *PLoS One* **2013**, *8* (3), e57804.

63. Wishart, D. S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, Brian L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, Vicki W.; Varshavi, D.; Varshavi, D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.; Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schiöth, Helgi B.; Greiner, R.; Gautam, V., HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Research* **2021**, *50* (D1), D622-D631.

64. Piovesana, S.; Capriotti, A. L.; Cavaliere, C.; Cerrato, A.; Montone, C. M.; Zenezini Chiozzi, R.; Laganà, A., The Key Role of Metal Adducts in the Differentiation of Phosphopeptide from Sulfopeptide Sequences by High-Resolution Mass Spectrometry. *Analytical Chemistry* **2022**, *94* (26), 9234-9241.

65. Ohguro, H.; Palczewski, K., Separation of phospho- and non-phosphopeptides using reverse phase column chromatography. *FEBS Lett* **1995**, *368* (3), 452-4.

66. Steen, H.; Jebanathirajah, J. A.; Rush, J.; Morrice, N.; Kirschner, M. W., Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. *Mol Cell Proteomics* **2006**, *5* (1), 172-81.

67. Yu, Y.; Hoffhines, A. J.; Moore, K. L.; Leary, J. A., Determination of the sites of tyrosine O-sulfation in peptides and proteins. *Nature Methods* **2007**, *4* (7), 583-588.

68. Önnerfjord, P.; Heathfield, T. F.; Heinegård, D., Identification of Tyrosine Sulfation in Extracellular Leucine-rich Repeat Proteins Using Mass Spectrometry\*. *Journal of Biological Chemistry* **2004**, *279* (1), 26-33.

69. Byrne, D. P.; London, J. A.; Eyers, P. A.; Yates, E. A.; Cartmell, A., Mobility shift-based electrophoresis coupled with fluorescent detection enables real-time enzyme analysis of carbohydrate sulfatase activity. *Biochemical Journal* **2021**, *478* (4), 735-748.

70. Bagwan, N.; El Ali, H. H.; Lundby, A., Proteome-wide profiling and mapping of post translational modifications in human hearts. *Scientific Reports* **2021**, *11* (1), 2184.

71. Zhao, Y.; Jensen, O. N., Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* **2009**, *9* (20), 4632-41.

72. Larsen, M. R.; Trelle, M. B.; Thingholm, T. E.; Jensen, O. N., Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *BioTechniques* **2006**, *40* (6), 790-798.

73. Low, T. Y.; Mohtar, M. A.; Lee, P. Y.; Omar, N.; Zhou, H.; Ye, M., WIDENING THE BOTTLENECK OF PHOSPHOPROTEOMICS: EVOLVING STRATEGIES FOR PHOSPHOPEPTIDE ENRICHMENT. *Mass Spectrometry Reviews* **2021**, *40* (4), 309-333.

74. Riley, N. M.; Coon, J. J., Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Analytical Chemistry* **2016**, *88* (1), 74-94.

75. Urban, J., A review on recent trends in the phosphoproteomics workflow. From sample preparation to data analysis. *Analytica Chimica Acta* **2022**, *1199*, 338857.

76. Grimsley, G. R.; Scholtz, J. M.; Pace, C. N., A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci* **2009**, *18* (1), 247-51.

77. Balsved, D.; Bundgaard, J. R.; Sen, J. W., Stability of tyrosine sulfate in acidic solutions. *Anal Biochem* **2007**, *363* (1), 70-6.

78. Yagami, T.; Kitagawa, K.; Aida, C.; Fujiwara, H.; Futaki, S., Stabilization of a tyrosine O-sulfate residue by a cationic functional group: formation of a conjugate acid-base pair. *J Pept Res* **2000**, *56* (4), 239-49.

79. Connor, P. A.; McQuillan, A. J., Phosphate Adsorption onto TiO<sub>2</sub> from Aqueous Solutions: An in Situ Internal Reflection Infrared Spectroscopic Study. *Langmuir* **1999**, *15* (8), 2916-2921.

80. Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jørgensen, T. J. D., Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns\*. *Molecular & Cellular Proteomics* **2005**, *4* (7), 873-886.

81. Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127* (3), 635-48.

82. Thingholm, T. E.; Jørgensen, T. J. D.; Jensen, O. N.; Larsen, M. R., Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nature Protocols* **2006**, *1* (4), 1929-1935.

83. Jensen, S. S.; Larsen, M. R., Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. *Rapid Communications in Mass Spectrometry* **2007**, *21* (22), 3635-3645.

84. Thingholm, T. E.; Jensen, O. N., Enrichment and characterization of phosphopeptides by immobilized metal affinity chromatography (IMAC) and mass spectrometry. *Methods Mol Biol* **2009**, *527*, 47-56, xi.

85. Thingholm, T. E.; Larsen, M. R., Phosphopeptide Enrichment by Immobilized Metal Affinity Chromatography. *Methods Mol Biol* **2016**, *1355*, 123-33.

86. Arribas Diez, I.; Govender, I.; Naicker, P.; Stoychev, S.; Jordaan, J.; Jensen, O. N., Zirconium(IV)-IMAC Revisited: Improved Performance and Phosphoproteome Coverage by Magnetic Microparticles for Phosphopeptide Affinity Enrichment. *Journal of Proteome Research* **2021**, *20* (1), 453-462.

87. Kweon, H. K.; Tang, B. L.; McClory, P. J.; Andrews, P. C.; Hakansson, K., Developing an analytical framework for sulfoproteomics: Sulfopeptide enrichment and positive mode mass spectrometric analysis. *ASMS 2017 Indianapolis* **2017**.

88. Salek, M.; Costagliola, S.; Lehmann, W. D., Protein Tyrosine-O-Sulfation Analysis by Exhaustive Product Ion Scanning with Minimum Collision Offset in a NanoESI Q-TOF Tandem Mass Spectrometer. *Analytical Chemistry* **2004**, *76* (17), 5136-5142.

89. Medzihradzsky, K. F.; Guan, S.; Maltby, D. A.; Burlingame, A. L., Sulfopeptide Fragmentation in Electron-Capture and Electron-Transfer Dissociation. *Journal of the American Society for Mass Spectrometry* **2007**, *18* (9), 1617-1624.

90. Edelson-Averbukh, M.; Shevchenko, A.; Pipkorn, R.; Lehmann, W. D., Discrimination between peptide O-sulfo- and O-phosphotyrosine residues by negative ion mode electrospray tandem mass spectrometry. *J Am Soc Mass Spectrom* **2011**, *22* (12), 2256-68.

91. Lanucara, F.; Lee, D. C.; Eyers, C. E., Unblocking the sink: improved CID-based analysis of phosphorylated peptides by enzymatic removal of the basic C-terminal residue. *J Am Soc Mass Spectrom* **2014**, *25* (2), 214-25.

92. Boersema, P. J.; Mohammed, S.; Heck, A. J., Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* **2009**, *44* (6), 861-78.

93. Hammond, D. E.; Kumar, J. D.; Raymond, L.; Simpson, D. M.; Beynon, R. J.; Dockray, G. J.; Varro, A., Stable Isotope Dynamic Labeling of Secretomes (SIDLS) Identifies Authentic Secretory Proteins Released by Cancer and Stromal Cells. *Mol Cell Proteomics* **2018**, *17* (9), 1837-1849.

94. Arribas Diez, I.; Govender, I.; Naicker, P.; Stoychev, S.; Jordaan, J.; Jensen, O. N., Zirconium(IV)-IMAC Revisited: Improved Performance and Phosphoproteome Coverage by Magnetic Microparticles for Phosphopeptide Affinity Enrichment. *J Proteome Res* **2021**, *20* (1), 453-462.

95. Hunter, J. E.; Campbell, A. E.; Kerridge, S.; Fraser, C.; Hannaway, N. L.; Luli, S.; Ivanova, I.; Brownridge, P. J.; Coxhead, J.; Taylor, L.; Leary, P.; Hasoon, M. S. R.; Eyers, C. E.; Perkins, N. D., Up-regulation of the PI3K/AKT and RHO/RAC/PAK signalling pathways in CHK1 inhibitor resistant Emicro-Myc lymphoma cells. *Biochem J* **2022**, *479* (19), 2131-2151.

96. Hunter, J. E.; Campbell, A. E.; Hannaway, N. L.; Kerridge, S.; Luli, S.; Butterworth, J. A.; Sellier, H.; Mukherjee, R.; Dhillon, N.; Sudhindar, P. D.; Shukla, R.; Brownridge, P. J.; Bell, H. L.; Coxhead, J.; Taylor, L.; Leary, P.; Hasoon, M. S. R.; Collins, I.; Garrett, M. D.; Eyers, C. E.; Perkins, N. D., Regulation of CHK1 inhibitor resistance by a c-Rel and USP1 dependent pathway. *Biochem J* **2022**, *479* (19), 2063-2086.

97. Hunter, J. E.; Campbell, A. E.; Butterworth, J. A.; Sellier, H.; Hannaway, N. L.; Luli, S.; Floudas, A.; Kenneth, N. S.; Moore, A. J.; Brownridge, P. J.; Thomas, H. D.; Coxhead, J.; Taylor, L.; Leary, P.; Hasoon, M. S. R.; Knight, A. M.; Garrett, M. D.; Collins, I.; Eyers, C. E.; Perkins, N. D., Mutation of the RelA(p65) Thr505 phosphosite disrupts the DNA replication stress response leading to CHK1 inhibitor resistance. *Biochem J* **2022**, *479* (19), 2087-2113.

98. Cohen, P., The origins of protein phosphorylation. *Nat Cell Biol* **2002**, *4* (5), E127-30.

99. Liu, H.; Håkansson, K., Electron Capture Dissociation of Tyrosine O-Sulfated Peptides Complexed with Divalent Metal Cations. *Analytical Chemistry* **2006**, *78* (21), 7570-7576.

100. Han, X.; He, L.; Xin, L.; Shan, B.; Ma, B., PeaksPTM: Mass Spectrometry-Based Identification of Peptides with Unspecified Modifications. *Journal of Proteome Research* **2011**, *10* (7), 2930-2936.