PAPER IN FOREFRONT

Peptide collision cross sections of 22 post-translational modifications

Andreas Will[1] · Denys Oliinyk[1] · Christian Bleiholder[2] · Florian Meier[1]

Abstract

Recent advances have rekindled the interest in ion mobility as an additional dimension of separation in mass spectrometry (MS)-based proteomics. Ion mobility separates ions according to their size and shape in the gas phase. Here, we set out to investigate the effect of 22 different post-translational modifications (PTMs) on the collision cross section (CCS) of peptides. In total, we analyzed ~4300 pairs of matching modified and unmodified peptide ion species by trapped ion mobility spectrometry (TIMS). Linear alignment based on spike-in reference peptides resulted in highly reproducible CCS values with a median coefficient of variation of 0.26%. On a global level, we observed a redistribution in the *m/z* vs. ion mobility space for modified peptides upon changes in their charge state. Pairwise comparison between modified and unmodified peptides of the same charge state revealed median shifts in CCS between −1.4% (arginine citrullination) and +4.5% (O-GlcNAcylation). In general, increasing modified peptide masses were correlated with higher CCS values, in particular within homologous PTM series. However, investigating the ion populations in more detail, we found that the change in CCS can vary substantially for a given PTM and is partially correlated with the gas phase structure of its unmodified counterpart. In conclusion, our study shows PTM- and sequence-specific effects on the cross section of peptides, which could be further leveraged for proteome-wide PTM analysis.

Keywords Post-translational modifications · PTM · Ion mobility · TIMS · Collision cross section · CCS

Introduction

Post-translational modifications (PTMs) are key regulators of protein activity and function in health and disease. Not at least because most PTMs involve a specific shift in the molecular weight of modified amino acids, mass spectrometry (MS)-based proteomics has evolved as the method of choice for the investigation of PTMs on a proteome-wide scale [1–3]. MS-based proteomics analyzes complex mixtures of (modified) peptides derived from tryptic digests by liquid chromatography coupled to high-resolution MS. Since the modification mass shift also transfers to fragment ions, this allows identifying modified peptide sequences and also often localizing the PTM to a specific amino acid.

Recently, ion mobility spectrometry (IMS) has become a popular extension of the proteomics toolbox that adds one more dimension of separation [4–10]. IMS distinguishes ions in the gas phase by their size and shape, which can be inferred from time- and field-dispersive ion mobility measurements in the form of a collision cross section (CCS) or, more precisely, the momentum transfer collision integral [11–13]. In the CCS vs *m/z* dimension, electrosprayed tryptic peptides typically split into distinct populations according to their charge state [14]. A more fine-structured heterogeneity, most prominently for triply charged species, has been attributed to either more extended or more compact structures that are determined by the linear peptide sequence [15, 16]. As a result, even isobaric and isomeric peptides can have distinct cross sections [17, 18].

Trapped ion mobility spectrometry (TIMS) is a relatively new type of IMS that inverses the concept of classical drift tube IMS by holding ions in an electric field against an opposing gas flow [19–21]. Lowering the electric field strength releases ions sequentially from the TIMS device to

✉ Florian Meier
florian.meier@med.uni-jena.de

1 Functional Proteomics, Jena University Hospital, Am Klinikum 1, 07747 Jena, Germany

2 Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32304, USA

the downstream mass analyzer as a function of increasing ion mobility (or decreasing CCS). The PASEF acquisition mode synchronizes the separation with a quadrupole time-of-flight mass analyzer, which greatly enhances the speed and sensitivity of peptide sequencing [5, 22, 23]. We have recently shown that an intriguing feature of this setup is that it enables peptide CCS measurements at very large scale and with high precision, sufficient to train a deep learning model to predict peptide cross sections based solely given the linear amino acid sequence and charge state as an input [16].

The advantages of IMS in terms of speed, sensitivity, and specificity should equally apply to or even be enhanced in the analysis of PTMs, given the additional complexity arising from isobaric modifications and positional isomers [24–27]. This motivated researchers since the beginning of IMS to study the effect of specific modifications on the gas phase characteristics of model peptides. By far the most studied example is phosphorylation of serine and threonine residues, which, interestingly, often leads to more compact gas phase structures compared to unmodified peptides — despite the increase in mass [26, 28–30]. Another field of interest is N-glycosylated peptides, for which the glycan moiety results in a distinct separation from unmodified peptides in the ion mobility dimension, and even differentiation between isomeric localization variants has been demonstrated [27, 31, 32]. To model the effect of PTMs on peptide CCS values, Clemmer and co-workers generated a dataset of cysteine-palmitoylated as well as cysteine-carboxyamidomethylated peptides and derived modification-specific intrinsic size parameters [33]. Kaszycki and Shvartsburg extended this approach to predict the intrinsic size parameters of 100 different PTMs [34]. More recent large-scale studies further highlighted additional sequence-dependent effects on peptide cross sections resulting from intramolecular interactions [16, 26, 35].

The further exploration of the potential of IMS for PTM analysis and the development of accurate CCS prediction models for a wide range of PTMs, similar to those for unmodified peptides [16, 35, 36], is currently limited by the availability of comprehensive experimental data. Here, we investigate the effect of 22 different naturally occurring PTMs on the collision cross section of peptides, including 21 sets of modified peptides synthesized as part of the ProteomeTools project [37, 38] and an additional set of O-GlcNAcylated peptides.

Methods

Synthetic peptides

A library of ~4500 lyophilized synthetic peptides in 96-well format representing 21 naturally occurring post-translational modifications on K, R, P, and Y residues and their unmodified counterparts were obtained from the ProteomeTools project [38]. Additionally, we purchased synthetic O-GlcNAcylated peptides and corresponding unmodified peptides from JPT Peptide Technologies GmbH (SpikeMix PTM-Kit 57 and 55). The pre-pooled peptides were reconstituted in 2% acetonitrile/0.1% formic acid to a final concentration of ~100 fmol/μL. As a reference, we spiked each sample with a retention time standard (Biognosys iRT) in a ratio of 1:40 (vol/vol) [39].

Liquid chromatography and mass spectrometry

All solvents were HPLC grade and purchased from Sigma-Aldrich. Nanoflow reversed-phase liquid chromatography was performed on a nanoElute system (Bruker Daltonics). Peptides were separated with a 120-min gradient at a flow rate of 0.3 μL/min at 60 °C on a homemade 50 cm × 75 μm column with a pulled emitter tip, packed with 1.9-μm ReproSil-Pur C18 - AQ beads. Mobile phases A and B were water/0.1% formic acid and acetonitrile/0.1% formic acid. The LC system was connected online to a TIMS-quadrupole time-of-flight mass spectrometer (Bruker timsTOF Pro or timsTOF HT) via a CaptiveSpray nano-electrospray source [5]. TIMS analysis was performed in a range from $1/K_0 = 1.5$ to 0.6 Vs cm^{-2}, while accumulating and analyzing in parallel for 100 ms each. The ion mobility gas was nitrogen from ambient air without temperature control, and the pressure at the TIMS entrance was kept at ~2.4 mbar. Detailed experimental parameters are provided in Supplementary Tables 1 and 2. To calibrate the ion mobility dimension, we added low-concentration Agilent ESI LC/MS to the inlet filter of the CaptiveSpray ion source and fitted the TIMS elution voltages to $1/K_0$ values using a linear model for at least three ions (m/z, $1/K_0$: 622.0289, 0.9848 Vs cm^{-2}; 922.0097, 1.1895 Vs cm^{-2}; and 1221.9906, 1.3820 Vs cm^{-2}). All data were acquired in dda-PASEF mode and suitable precursor ions for fragmentation were selected by their relative position in the $1/K_0$ vs. m/z plane. The quadrupole isolation window was set to 2 Th for $m/z < 700$ and 3 Th for $m/z > 700$. To prevent repeated fragmentation of precursors for which the dda-PASEF target intensity value of 20,000 a.u. has been already achieved, we defined an exclusion time of 0.4 min.

Data processing

The MS raw data were processed with MaxQuant (version 2.1.4.0) using the default parameters for the analysis of timsTOF data [40, 41]. MS/MS spectra were searched against concatenated peptide sequences as provided by the ProteomeTools consortium and JPT Peptide Technologies,

both supplemented with the iRT peptide sequences. The digestion mode was set to "specific" according to the cleavage rules for Trypsin. We analyzed each PTM pool separately, defining methionine oxidation and each peptide pool`s respective modification as variable modifications (Supplementary Table 3) and cysteine carbamidomethylation as a fixed modification. The false discovery rate on the peptide spectrum match and protein level was controlled by a target-decoy approach <1%. For modified peptides, we required a minimum Andromeda score [42] of 40 and a minimum delta score of 6 [43].

Bioinformatic analysis

Data analysis and visualization were performed in R (v4.2.1) using the packages tidyverse (1.3.1), magrittr (2.0.3), data.table (1.14.2), ggplot2 (3.3.6), ggrepel (0.9.1), ggnewscale (0.4.8), heatmaply (1.4.0), orca (1.1.1), GGally (2.1.2), rstatix (0.7.2), and VennDiagram (1.7.3). Heatmaps were generated with the R package "heatmaply" and clustered using the Euclidean distance measure and the average linkage function.

To account for time-dependent drifts in measured ion mobility ($1/K_0$) and retention time (RT) values, we linearly aligned all experiments by a run-wise alignment value (see also Results). To determine appropriate reference values for this, we first measured three replicate injections of the 11-iRT peptide mixture, calibrating the TIMS dimension before each injection as described above, and calculated the mean $1/K_0$ (or RT) value for each iRT peptide (Supplementary Table 4). As we spiked iRT peptides in all our measurements, we could then extract the experimental $1/K_0$ for each identified iRT peptide in each experiment. From this, and separately for each run, we calculated the median deviation by subtracting each measured iRT peptide $1/K_0$ (and RT) value from its respective reference value. This resulted in one value per run, which we used to align all runs by subtracting it from all measured $1/K_0$ values, run by run. The aligned $1/K_0$ values were converted to CCS values using the Mason-Schamp equation. For further analysis, we kept only peptide sequences identified with a single modification site not localized on the C-terminus, excluding oxidized peptides. In cases, in which MaxQuant listed multiple "evidences" for a modified peptide sequence, we retained only the most abundant feature (highest intensity value) for charge state 2 and 3 for our analysis. CCS and RT values represent mean values calculated from one to three technical replicates. Differences in the collision cross section of modified peptides and their unmodified counterparts were calculated relative to the cross section of the unmodified peptide ($\Delta CCS = (CCS_{modified} - CCS_{unmodified})/CCS_{unmodified}$).

Modeling of peptide collision cross sections

The potential energy surface of the selected doubly protonated peptides was explored by molecular dynamics (MD) simulations using the OPLS all-atom force field [44, 45] in conjunction with the GROMACS suite of programs. For simulation of peptides with succinyl-modified lysine residues, we used the parameters tabulated for aliphatic molecular systems [46]. Extra protons were placed on the most basic sites of the peptides, i.e., the lysine residues that are known to sequester protons. We further used a deprotonated carboxy-terminus and a protonated amino-terminus. For the succinylated sequences, which lack one of the basic lysine residues, we neutralized the carboxy-terminus. Note that the semi-empirical calculations performed subsequent to the MD simulations would allow proton transfer processes to take place if they were to lead to energetically more stable structures. During the MD simulations, we used simulated annealing techniques to produce candidate structures for further refinement. These simulations were carried out for a duration of 100 ns at simulation temperatures of up to 600 K to effectively overcome barriers on the potential energy surface. Snapshots were saved every 100 ps during these simulations and analyzed by a conformer family search program which assigns structures into families within which the most important characteristic torsion angles are similar. Note that this approach was extensively used to discuss gas-phase fragmentation pathways of protonated peptides [46–48]. A full geometry optimization was subsequently carried out for the most stable structure of each conformer family at the PM6 level of theory [49] using the MOPAC suite of programs [50]. The optimized structure was then used as input for a cross section calculation using our projection superposition approximation (PSA) method [51] in nitrogen gas [52]. We restricted the putative assignment of gas phase structures to experimental mobility spectra to the respective lowest-energy conformers.

Data availability

The mass spectrometry proteomics data underlying this study have been deposited to the ProteomeXchange Consortium [53, 54] via the PRIDE partner repository with the dataset identifier PXD042416. A summary result file is provided as Supplementary File 2. Code to reproduce the data analysis and visualization in this study can be accessed via https://github.com/MeierLab/2022_22PTM.

Results and discussion

Constructing a peptide CCS dataset for 22 PTMs

To generate a high-quality CCS dataset of a wide range of naturally occurring peptide modifications, we analyzed

pooled libraries of synthetic peptides by nanoflow reversed-phase liquid chromatography and TIMS-quadrupole time-of-flight mass spectrometry (Fig. 1a). In dda-PASEF mode, the mass spectrometer selects suitable precursor ions from a survey TIMS-MS scan and targets them for fragmentation in the subsequent PASEF-MS/MS scans. As quadrupole and collision cell are positioned downstream of the TIMS analyzer, precursor and fragment spectra are linked through their position in the ion mobility spectrum. We processed this data in the MaxQuant software to assemble three-dimensional features in ion mobility, m/z and retention time dimension and match the associated MS/MS spectra to (modified) peptide sequences [41]. The inverse reduced ion mobility value ($1/K_0$) determined from the mobility spectrum of each feature can then be converted into ion-nitrogen $^{TIMS}CCS_{N_2}$ values using the Mason-Schamp equation [55].

Figure 1b shows an overview of all PTMs in our dataset. The ProteomeTools library contributed matching modified and unmodified tryptic peptides derived from human protein sequences, which were selected for their favorable LC-MS properties and synthesizability as described in more detail in the original publications [37, 38]. The library contains four different sets of base peptide sequences that carry N-terminal or internal modifications on one of four amino acids (lysine, arginine, tyrosine, and proline). All peptide sequences with the same modification are combined, resulting in 21 pools of modified peptides and four matching pools of unmodified peptides, which we measured in randomized order and in triplicate. In addition, we analyzed a pool of O-GlcNAcylated peptides (serine/threonine) and their unmodified counterparts in the same way. Overall, lysine modifications represent the largest group in our dataset, including three homologous series for acylation with aliphatic residues (formylation to butyrylation), carboxylic acid residues (malonylation to glutarylation) and methylation. Further, lysine biotinylation is included as well as the GlyGly remnant of ubiquitination and hydroxylated proline. Tyrosine modifications in our dataset are phosphorylation and nitration. The
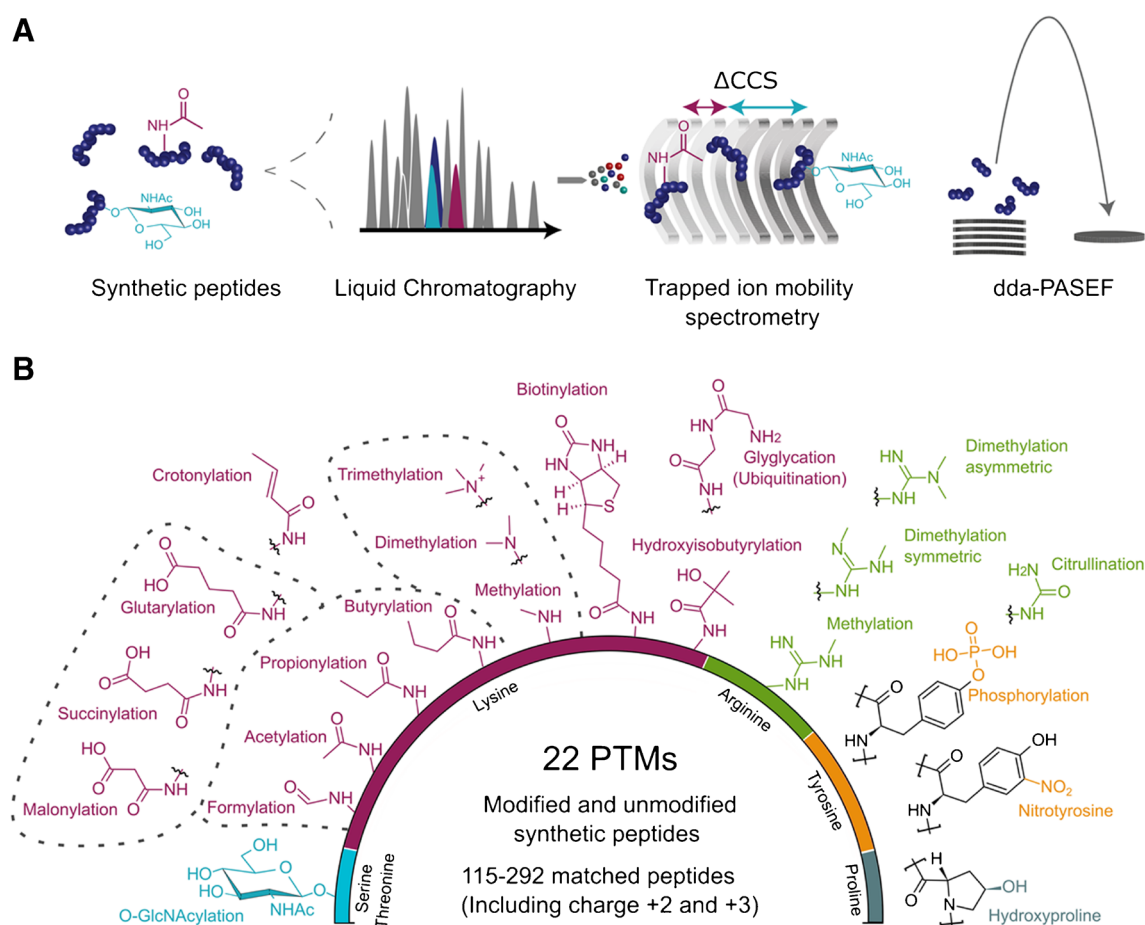
Fig. 1 Overview of 22 post-translational modifications (PTMs) analyzed by trapped ion mobility spectrometry. **A** Schematic workflow for the analysis of synthetic peptide pools by liquid chromatography and trapped ion mobility-quadrupole time-of-flight mass spectrometry with data-dependent acquisition (dda)-PASEF. **B** Dataset summary and chemical structures of the peptide modifications investigated in this study color-coded by their respective modified amino acid. Homologous series are highlighted by dashed lines

arginine pool adds further subtleties through symmetric and asymmetric di-methylations.

In total, we compiled 84 raw files, which resulted in ~58,000 peptide spectrum matches mapping to ~5000 unique combinations of peptide sequence, charge state, and modification. Of these, 74% and 26% were detected as doubly and triply charged species, and only 1% in charge state 4. The median Andromeda score was 105 with a very high localization probability close to 1 for all modifications except for O-GlcNAc, which is labile in collision induced dissociation experiments (Supplementary Fig. 1). Plotting the m/z vs. CCS distribution of modified and unmodified peptides shows the expected clustering by charge state distributed over an m/z range of about 300–1200 and a CCS range of about 300–700 $Å^2$ (Supplementary Fig. 2). For further analysis, we extracted the CCS value of the most abundant feature for each modified peptide sequence, while keeping doubly and triply charged ions separate because of their distinct ion mobility. This yielded approximately 4300 matched pairs of modified and unmodified peptides and about 115 to 292 pairs per modification.

Precision of TIMS CCS measurements

Experimental ion mobility values depend on the analyte itself as well as the electric field and experimental parameters such as temperature and the nature of the mobility gas [13, 56]. To make them comparable between experiments, TIMS is usually calibrated by a linear regression of known $1/K_0$ values to the elution voltage, resulting in good agreement with conventional drift tube experiments [21]. In our previous study, we demonstrated that $^{TIMS}CCS_{N2}$ values from multiple experiments can additionally be linearly aligned based on overlapping peptide identifications, leading to a remarkable reproducibility over long periods of time

and across instruments [16]. To allow a similar alignment in our dataset of non-overlapping peptide pools and avoid external calibration before each injection, we here spiked a standard of eleven synthetic iRT peptides into each sample. The peptides eluted evenly distributed over the 2-h chromatographic gradient and were detected as doubly protonated species with CCS values ranging from 320 to 430 $Å^2$. Our analysis revealed a time-dependent shift of their measured CCS values over the course of the experiment (84 LC-MS injections, acquired on two independent instruments), which could be attributed to changes in experimental conditions such as ambient temperature and pressure that are typically not controlled in TIMS experiments [11] (Fig. 2a). A linear alignment to the average iRT $1/K_0$ values from three reference experiments successfully corrected these drifts (Methods, Fig. 2a lower panel), while using the median deviation as a correction factor makes it robust against outliers (e.g., "DGLDAA…" in Fig. 2a). Across all experiments, the median coefficient of variation of the iRT peptide CCS values was 1.41% before and 0.30% after the alignment.

Next, we determined the precision of the CCS measurements for all other peptides in our dataset (Fig. 2b). Because of the high sequencing rate of dda-PASEF and the relatively low complexity of the synthetic peptide pools, 76% of all peptides were identified in all three replicates. For these, the coefficient of variation was significantly improved ($p < 0.001$, Kolmogorov-Smirnov test) from 1.23% before to only 0.26% after linear alignment. This indicates an excellent reproducibility of our TIMS measurements and is in line with previous reports on this instrument platform [5, 16].

Global view on CCS values of modified peptides

The fact that charge is a major determinant of ion mobility prompted us to investigate the occurrence of different charge

Fig. 2 Reproducibility and cross-run alignment of peptide collision cross sections. **A** CCS values of eleven reference peptides spiked into 84 consecutive LC-MS experiments. Data are shown in chronological order before (top) and after (below) performing a linear alignment to correct for drifts in the ion mobility measurement (Methods). **B** Coefficients of variation of CCS values of (modified) peptides measured in three replicate injections in the full dataset (*top*: raw data, *bottom*: aligned data). $N = 4213$, bin width = 0.05, 1% outliers out of bounds not shown
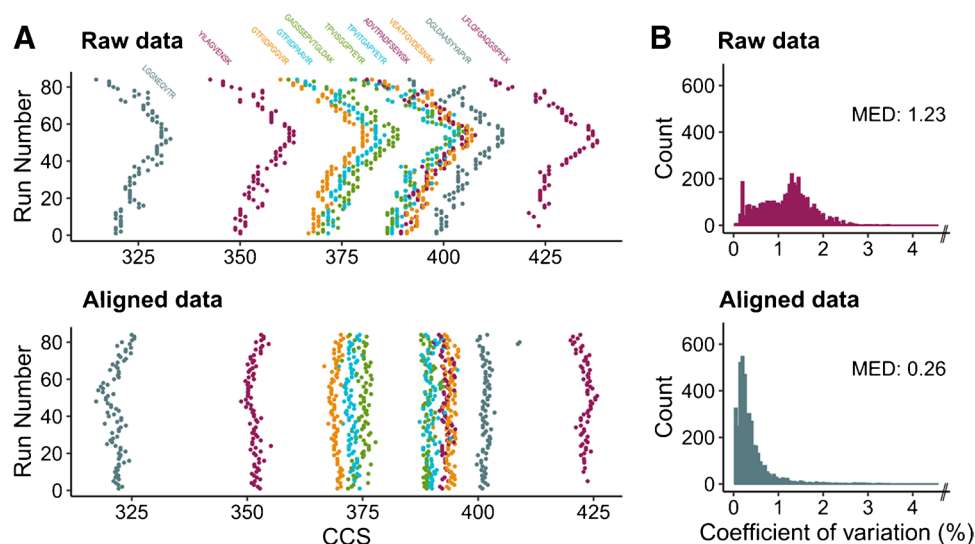
states in our dataset. Figure 3a provides an overview of predominant charge states as well as the relative abundance fraction of charge states for all peptides. Tryptic peptides generally take up two to four protons in the electrospray process, depending on the length of the amino acid sequence and the number of basic residues. PTMs can alter the pK_a and gas phase basicity of the modified amino acid and hence the charge state distribution. This effect was most striking for lysine modifications, as doubly and triply charged species appeared in roughly equal abundance for the unmodified peptides, whereas the various acylations shifted the charge distribution almost completely to charge 2. By contrast, lysine methylations as well as the GlyGly residue retain basic properties at the lysine site and thus showed only little effect on the relative abundance distribution, but rather tipped the predominant charge state to 3. We observed similar trends for arginine methylation and, as expected, citrullination reduced the charge state. Hydroxylation of proline, O-GlcNAcylation of serine and threonine as well as nitration and phosphorylation of tyrosine did not alter the charge state distribution as compared with their unmodified counterparts. Overall, these results are in agreement with the preceding analysis of the 21 PTM library on a different instrument platform [38].

To illustrate the global effect of charge state alterations on the ion mobility of modified peptides, we plotted them in their predominant charge state in the m/z vs. CCS space (Fig. 3b). The unmodified peptides of the lysine pool (left panel) are not strictly tryptic peptides because of their internal lysine. Nevertheless, they followed the well-characterized distribution of ion mobility and charge state occupancy. Malonylation, as an example for the group of acylations, thinned the population of charge 3 species and led to a more dense population of charge 2 peptides. Conversely, GlyGlycation caused a distinct shift in the opposite direction and predominantly populated the area of triply charged species. Thus, simple charge state alterations can already contribute to the separation of peptides and modifications in the ion mobility dimension.

Pairwise comparison of modified and unmodified peptides

A distinct feature of our dataset is the large number of matching pairs of modified and unmodified peptide sequences. Thus, having determined the position of modified peptide populations in the CCS vs. m/z space, we next performed a pairwise analysis of the respective counterparts. Intuitively, one could expect increasing CCS values throughout, as all modifications in our dataset introduce an additional functional group to the peptide (with the only exception of citrullination). However, we observed relative differences ($\Delta CCS = (CCS_{modified} - CCS_{unmodified})/CCS_{unmodified}$) ranging all the way from about -10 to $+10\%$, depending on the type of modification, the modified amino acid, and the charge state (Fig. 4a). As a proxy for the experimental precision in the dataset, we indicate the $+/-$ three-fold coefficient of variation (0.78%) interval in the boxplot. The median ΔCCS shifts of specific modifications ranged from almost no difference (0% for lysine methylation) to slightly negative values

Fig. 3 Analysis of charge state and collision cross section (CCS) distribution for different PTMs. **A** Number of identified peptides per modification in their predominant charge state (top). Relative abundance of different charge states for all identified peptides (bottom). For R and K, only unmodified peptide sequences with at least one internal R or K are shown. **B** Peptide m/z vs. CCS distribution of selected modifications from our dataset of 22 PTMs. The scatter plots show peptides in their predominant charge state
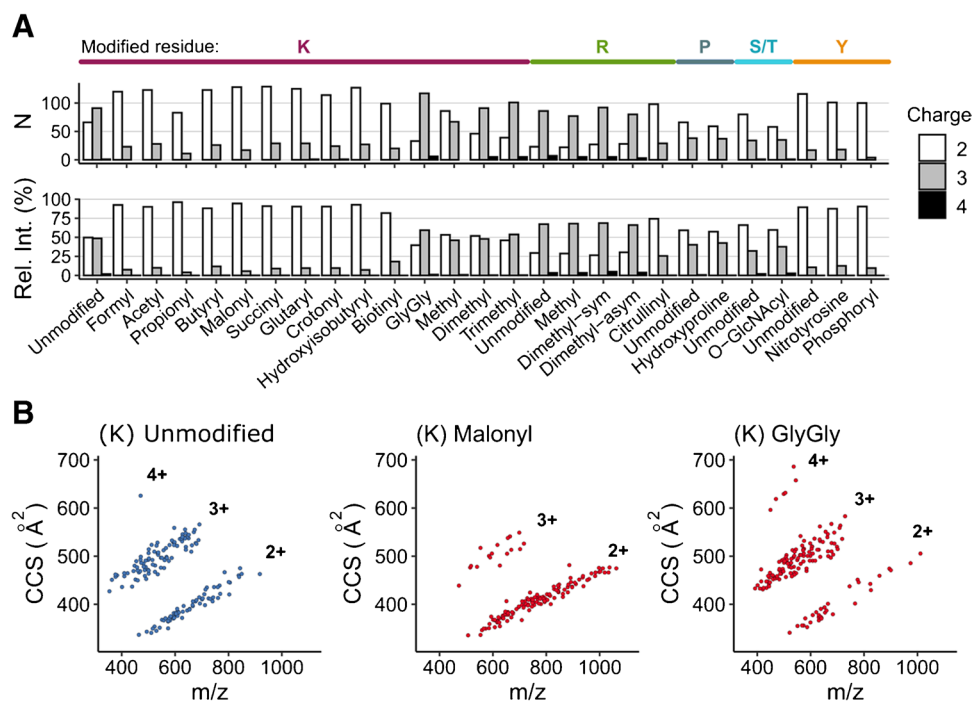
(−1.2% for lysine formylation) and clear positive shifts for large modifications such as biotinylation (4.3%) and O-Glc-NAcylation (4.5%).

To delve deeper into factors that determine CCS values of modified peptides, it is insightful to first consider the two different homologous series in the subset of lysine modifications: acylations with mono- (formyl to butyryl) and di-carboxylic acids (malonyl to glutaryl). Both series show a nearly linear increase in CCS with longer acyl chains, consistent for both charges 2 and 3 (Fig. 4a). This is in line with the intuition of increasing size with increasing chain length. To our surprise, and despite the fact that both series differ by one carboxylic acid, the median shifts in CCS of, for example, acetyl and malonyl were almost identical (−0.2%, −0.5% respectively). Moreover, they were similar to those of lysine methylation (0%), an even smaller moiety. A possible explanation for this observation are the different chemical properties of the modifications. While methylation tends to increase the basicity at the modified lysine, the acetyl group is electron withdrawing and malonylation adds an acidic

functionality. Peptide conformations in the gas phase can be partially explained by Coulomb interactions and intra-molecular charge solvation [14]. Analyzing a larger dataset of peptide CCS values, Chang et al. concluded that internal acidic residues facilitate more compact conformations and, conversely, internal basic residues result in more extended conformations with larger cross sections [35]. Our data is in line with this model and suggests that the carboxyl residue of malonyl compensates for its increased size.

Next, we plotted the median ΔCCS values as a function of the modification's molecular weight (Fig. 4b, Suppl. Fig. 3a). This analysis reproduced the large effect size for O-GlcNAcylation and biotinylation and showed an overall correlation of $r^2 = 0.74$ between ΔCCS and ΔM. However, when excluding the latter from the analysis, r^2 dropped to only 0.38, indicating that mass alone is a poor predictor of ΔCCS values for this group of chemically diverse modifications. In line with the results above, the homologous lysine modification series appear on parallel lines in this plot (dashed gray lines). In addition, we observed modifications

Fig. 4 Pairwise comparison of modified and unmodified peptide collision cross section (CCS). **A** Relative difference of CCS values of matching pairs, separated by charge state (ΔCCS). The number of pairs is plotted on the top axis. The median relative deviation (MED) indicates the combined value for charges 2 and 3, which was tested for significance using a one-sample Wilcoxon test with Benjamini-Hochberg correction (*$p<0.05$; **$p<0.01$; ***$p<0.001$; ****$p<$1e-04). Dashed lines indicate the 3-fold median coefficient of variation in the full dataset as a proxy for experimental precision. Boxplot elements: Interquartile range within boxes; median indicated by horizontal line; whiskers spanning 1.5-fold the interquartile range. **B** Median ΔCCS values of all investigated modifications as a function of the molecular weight change (ΔM). Dashed lines visualize homologous series. See also Supplementary Fig. 3 for absolute ΔCCS and relative ΔM values. **C** Same as **B**, but as a function of the observed shift in retention time (ΔRT)
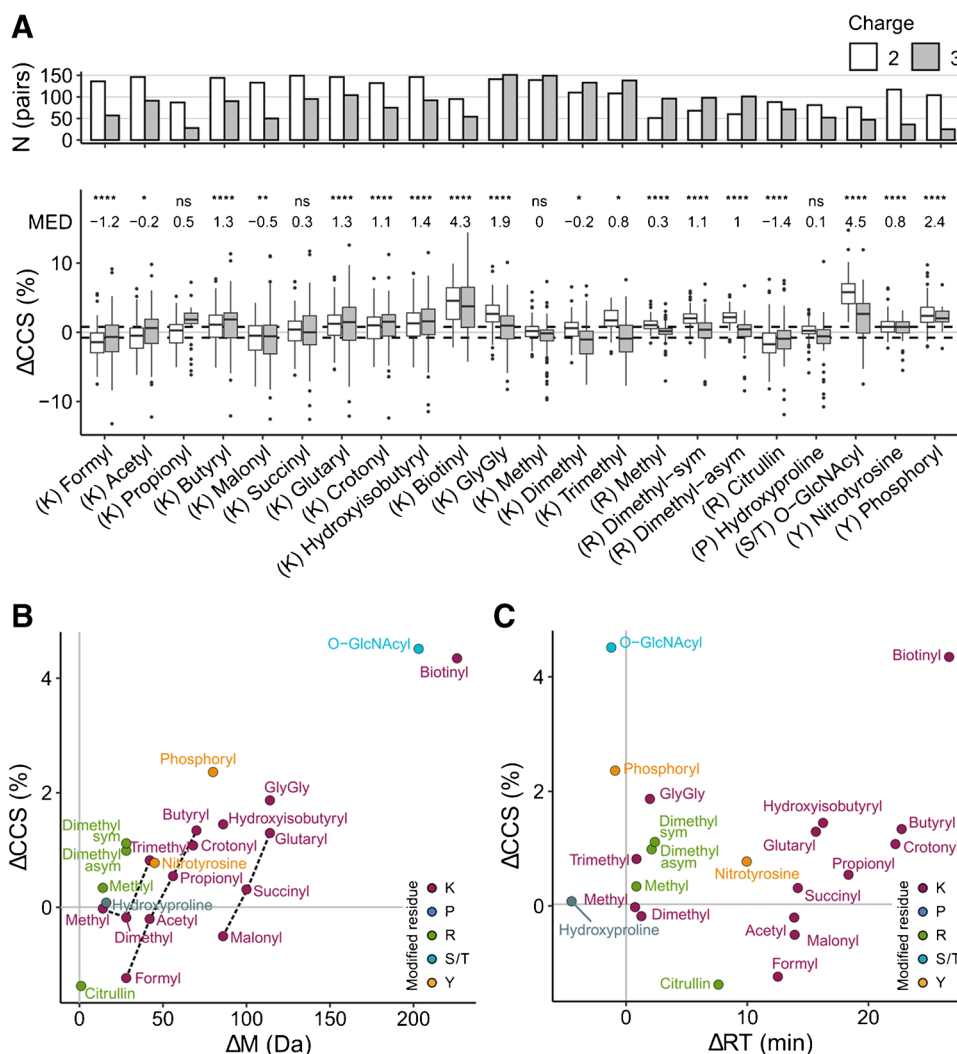
with different molecular weight but similar ΔCCS such as the methyl-acetyl-malonyl example above (7, 21, and 43 Da). Conversely, other modifications such as malonyl and hydroxyisobutyryl have a similar molecular weight but different ΔCCS (−0.5% vs. +1.4%). As the example of lysine- and arginine-dimethylation shows, this effect is not limited to the chemistry of the modification itself, but can also depend on the modified amino acid. Performing a similar correlation analysis of ΔCCS and changes in the chromatographic retention time (Fig. 4c, Suppl. Fig. 3b) revealed some degree of orthogonality between LC and ion mobility. O-GlcNAcylation, phosphorylation, and GlyGlycation, for example, only slightly altered the retention time, while resulting in relatively large CCS shifts.

Although our analysis highlighted conceivable trends for each modification in our data, we also noted a relatively high variance within the modification groups (Fig. 4a). In most cases, the interquartile range of the pairwise analysis was several-fold larger than the precision of our experiments. Furthermore, some modifications showed less variance compared to others and the variance in charge 3 species was generally higher. This hints towards sequence-dependent effects within the peptide pools that modulate the effect of modifications on peptide cross sections.

## Resolving sequence-dependent CCS determinants

To dissect the high variance within the modification groups, we next resolved our data by peptide sequences. This is possible because all peptides from one amino acid pool have the same unmodified base sequence. Figure 5a shows absolute ΔCCS values of all doubly charged peptides in the subset of lysine modifications color-coded by their respective base peptide sequence. Strikingly, the rank order of ΔCCS values remained largely unchanged throughout the homologous series of acyl modifications. In other words, while a particular modification could alter a peptide's cross section from +10 Å² to −25 Å² (formylation) depending on its amino acid sequence, elongating the modification, e.g., from formyl to butyryl, resulted in consistent increments across all peptide sequences. This observation also applied to triply charged peptide ions (Suppl. Fig. 4) and only few peptides deviated from this trend. Even for the chemically rather distant biotinylation, we observed a similar behavior, although more peptides swapped positions, in particular those with larger ΔCCS values. In contrast, when extending the line plot to lysine methylations and GlyGly, the trend was interrupted. However, the latter modifications are also distinct from the former acyl-type modifications in terms of charge state distribution (Fig. 3). This suggests that the observed grouping is driven by intramolecular charge localization and solvation. Similarly, we found that for individual sequences, the absolute ΔCCS as well as the shift relative to other sequences can

be discordant for charge states (Fig. 5b). These results raised the question of whether there are commonalities in peptide sequences that undergo similar changes in their gas phase structure upon modification. To this end, we plotted the m/z vs. CCS distribution of the unmodified peptides and overlaid the corresponding ΔCCS values for different modifications. As a visual aid, we divided them into extended (larger CCS values) and compact structures (smaller CCS values) by fitting a linear model to each charge state. Figure 5c shows lysine succinylation and trimethylation as examples. For succinylation, close inspection of the doubly and triply charged ion populations revealed a tendency towards negative ΔCCS values for extended structures and vice versa. We also observed a compaction of extended structures for triply charged trimethylated peptides, indicating that the internally localized charge destabilizes extended conformations. This sequence dependency of trimethylation was less pronounced for doubly charged species, in line with their narrower ΔCCS distribution in Fig. 4a.

To investigate this further, we selected two peptide sequences from our lysine succinylation data and modeled the gas phase structures of their doubly charged ions using a computational approach based on molecular dynamics (Methods, Suppl. Fig. 5). Our modeling recapitulated the expected charge reduction at the modified lysine residue, which for the succinylated peptides resulted in proton localization at the peptide N-terminus and the C-terminal lysine residue. By contrast, one acidic proton was assumed to be sequestered at the internal lysine residues for the unmodified peptides. For "VGID…K(succinyl)LK," the proton localization to the terminal residues resulted in re-folding to a more extended structure as compared to its unmodified counterpart (+32 Å², +6.7%), while the lysine succinylation prevented interactions between the terminal residues. This was in good agreement with our experiment (+32 Å²). Conversely, for the modeled gas phase structure of the doubly charged "GTI…K(succinyl)…AK," proton relocalization and succinylation of the internal lysine did not prevent the adoption of folded peptide conformations by interaction of the terminal residues. Consequently, only minor changes of the gas phase collisional cross sections were computed (+8 Å², +1.9%), while we even found a negative ΔCCS value (−25 Å²) experimentally.

These results led us to hypothesize that the gas phase conformation of the unmodified peptide ion is indicative of the relative effect of specific modifications. To test this on all our data, we plotted the ΔCCS value for each modification as a function of the corresponding residuals of the linear fit (indicating whether the unmodified peptide adapts a more compact or extended structure) (Suppl. Figs. 6 and 7). The slope of the resulting linear trend lines can be interpreted as "ΔCCS gradients" within the CCS vs. m/z ion populations (Fig. 5d). Indeed, all lysine acylations and citrullination followed the trend described above for succinylation, while

Fig. 5 Sequence-dependent fine structure of PTM-induced CCS shifts. **A** ΔCCS values of individual peptide sequences for lysine modifications at charge 2. Only peptides which were detected for all shown modifications are included ($N = 63$). Peptides with the same unmodified sequence are connected and color-coded according to their rank order for formylation, from highest (pink) to lowest ΔCCS (orange). **B** Heat maps of ΔCCS values comparing charges 2 and 3 for both succinylation ($N = 89$) and trimethylation ($N = 102$). **C** Peptide m/z vs. CCS distribution of unmodified reference peptides for succinylation ($N = 244$) and trimethylation ($N = 246$), overlaid with each peptides PTM-induced CCS shift. Linear trend lines are fitted to both charge states. **D** ΔCCS gradient within ion clouds for each investigated PTM, separated by charge. Each peptide's residual to a linear regression, as depicted in panel **C**, was plotted against its ΔCCS value. The slope of a subsequent linear regression represents a ΔCCS gradient within the respective ion clouds (panel **C**)
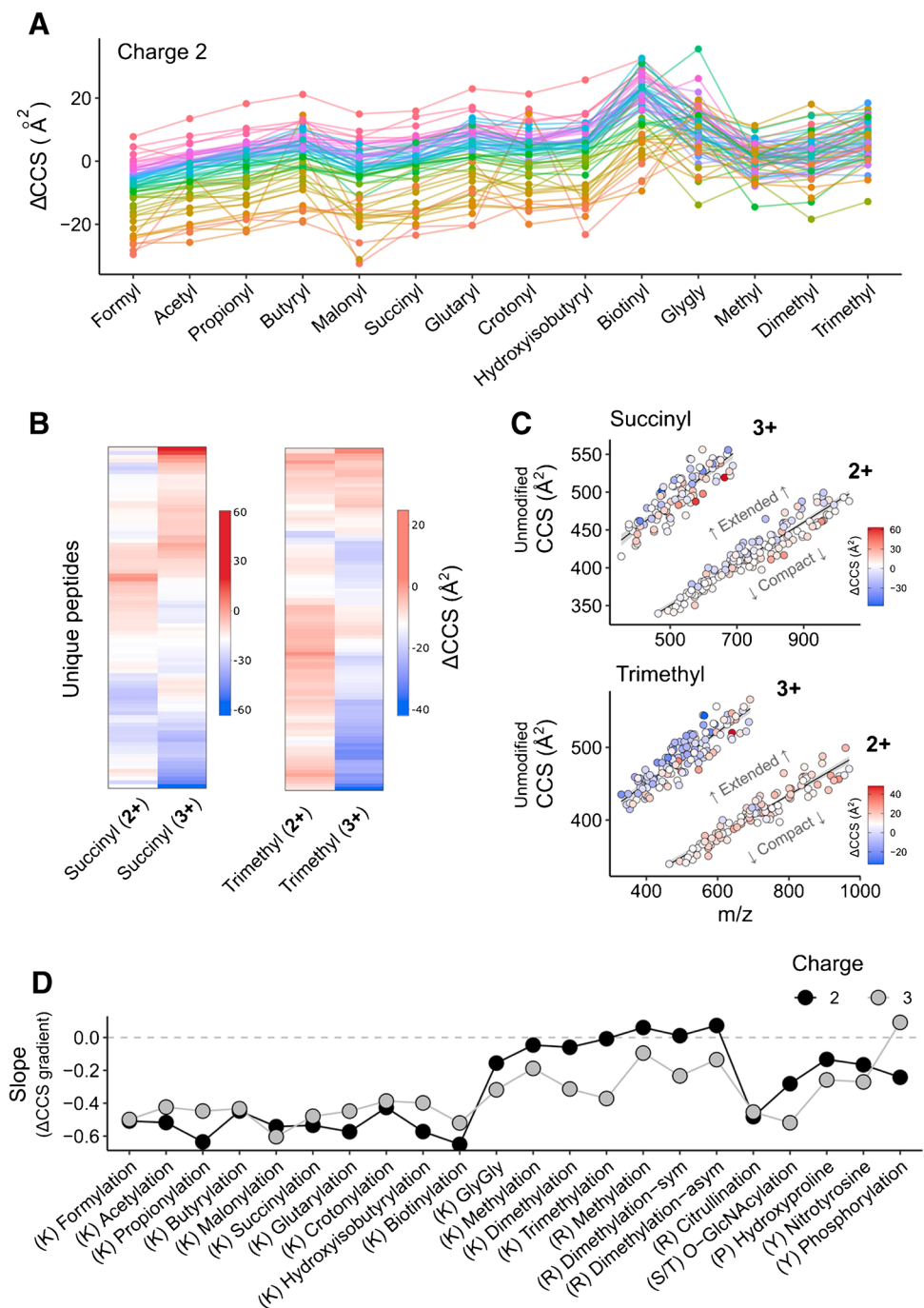
GlyGly and methylated peptides resembled trimethylation. In particular for doubly charged peptides with lysine acylations, the ion populations in CCS vs. m/z appeared narrower with respect to their trendlines (Supplementary Fig. 8). Taken together, our data suggests that the ΔCCS value associated with a modification is indeed correlated with the "starting conformation" of the unmodified peptide and hence at least partially dependent on the amino acid sequence, rather than a fixed increment determined by its chemical composition.

Conclusions

Recent advances in the application of ion mobility spectrometry to MS-based proteomics also promise new opportunities for the proteome-wide characterization of post-translational modifications. In particular, the combination of TIMS and PASEF has enabled the precise measurement of CCS values on the scale of hundreds of thousands to more than a million data points and contributed to a

better understanding of sequence- and position-dependent determinants of peptide cross sections [16, 35, 57]. To extend this work beyond unmodified peptides, here, we used synthetic peptide libraries with known ground truth to characterize the effect of 22 different PTMs on peptide CCS values.

Our study provides data on 115 to 292 matched modified and unmodified peptides per modification with a precision <1% after linear alignment, which is on par with previous studies on the same instrument platform [5, 16]. Limitations of this approach include its reliance on the accuracy of the software-based feature detection as well as the fact that the accuracy of ^{TIMS}CCS measurements, particularly for complex samples, can be affected by exceeding the local charge capacity in the TIMS cartridge. However, the precision of the dataset enabled us to investigate different layers of modification-specific effects on peptide CCS values. On a global level, we observed major shifts in the m/z vs. ion mobility distribution for modified peptides, which we attributed to changes in their predominant charge state. In proteomics practice, such effects can be important, for example, to optimize the precursor selection scheme in dia-PASEF experiments [24, 25] or to bias data-dependent acquisition towards modified peptides [27, 58].

Turning to pairwise comparisons of modified peptides and their unmodified counterparts, we observed median ΔCCS values in a range of −1.4 to 4.8%. Surprisingly, despite the correlation between ion mass and mobility, the modification mass alone proved to be a poor predictor of ΔCCS values for most modifications in our dataset. In parts, we could rationalize these observations by the counteracting effects of increased modification size on the one hand and intramolecular charge solvation on the other hand. In addition, our data revealed substantial sequence-dependent effects on the cross section of modified peptides. This is in line with another recent study focused on phosphorylated peptides [26]. All in all, our study adds to the increasing body of work indicating that peptide cross sections are determined by the amino acid composition [59, 60] as well as their linear sequence [16, 35, 61].

Accurate predictions of peptide properties such as retention time, MS/MS spectra, and CCS values are increasingly used in MS-based proteomics [62, 63]. In this context, synthetic peptides can provide important training data as they have a known ground truth [37]. This applies in particular to modified peptides, which are not always readily accessible from biological sources via efficient and affordable enrichment protocols. We envision that our high-quality dataset fills this gap and helps to extend CCS prediction algorithms to various post-translational modifications, for example, via transfer learning [36].

Declarations

**Conflict of interest** The authors declare no competing interests.

References

1. Olsen JV, Mann M. Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol Cell Proteom. 2013;12(12):3444–52.

2. Doll S, Burlingame AL. Mass spectrometry-based detection and assignment of protein posttranslational modifications. ACS Chem Biol. 2015;10(1):63–71.

3. Leutert M, Entwisle SW, Villén J. Decoding post-translational modification crosstalk with proteomics. Mol Cell Proteom. 2021;20:100129.

4. Bekker-Jensen DB, Martínez-Val A, Steigerwald S, Rüther P, Fort KL, Arrey TN, et al. A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients. Mol Cell Proteom. 2020;19(4):716–29.

5. Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. Mol Cell Proteom. 2018;17(12):2534–45.

6. Hebert AS, Prasad S, Belford MW, Bailey DJ, McAlister GC, Abbatiello SE, et al. Comprehensive single-shot proteomics with FAIMS on a hybrid orbitrap mass spectrometer. Anal Chem. 2018;90(15):9529–37.

7. Pfammatter S, Bonneil E, McManus FP, Prasad S, Bailey DJ, Belford M, et al. A novel differential ion mobility device expands the depth of proteome coverage and the sensitivity of multiplex proteomic measurements. Mol Cell Proteom. 2018;17(10):2051–67.

8. Helm D, Vissers JPC, Hughes CJ, Hahne H, Ruprecht B, Pachl F, et al. Ion Mobility tandem mass spectrometry enhances performance of bottom-up proteomics. Mol Cell Proteom. 2014;13(12):3709–15.

9. Ibrahim YM, Baker ES, Danielson WF, Norheim RV, Prior DC, Anderson GA, et al. Development of a new ion mobility time-of-flight mass spectrometer. Int J Mass Spectrom. 2015;377:655–62.

10. Distler U, Kuharev J, Navarro P, Tenzer S. Label-free quantification in ion mobility–enhanced data-independent acquisition proteomics. Nat Protoc. 2016;11(4):795–812.

11. Gabelica V, Marklund E. Fundamentals of ion mobility spectrometry. Curr Opin Chem Biol. 2018;42:51–9.

12. Dodds JN, Baker ES. Ion mobility spectrometry: fundamental concepts, instrumentation, applications, and the road ahead. J Am Soc Mass Spectrom. 2019;30(11):2185–95.

13. Gabelica V, Shvartsburg AA, Afonso C, Barran P, Benesch JLP, Bleiholder C, et al. Recommendations for reporting ion mobility Mass Spectrometry measurements. Mass Spectrom Rev. 2019;38(3):291–320.

14. Counterman AE, Clemmer DE. Large Anhydrous Polyalanine Ions: Evidence for Extended Helices and Onset of a More Compact State. J Am Chem Soc. 2001;123(7):1490–8.

15. Lietz CB, Yu Q, Li L. Large-scale collision cross-section profiling on a traveling wave ion mobility mass spectrometer. J Am Soc Mass Spectrom. 2014;25(12):2009–19.

16. Meier F, Köhler ND, Brunner A-D, Wanka J-MH, Voytik E, Strauss MT, et al. Deep learning the collisional cross sections of the peptide universe from a million experimental values. Nat Commun. 2021;12(1):1185.

17. Wu C, Siems WF, Klasmeier J, Hill HH. Separation of Isomeric Peptides Using Electrospray Ionization/High-Resolution Ion Mobility Spectrometry. Anal Chem. 2000;72(2):391–5.

18. Srebalus Barnes CA, Hilderbrand AE, Valentine SJ, Clemmer DE. Resolving Isomeric Peptide Mixtures: A Combined HPLC/Ion Mobility-TOFMS Analysis of a 4000-Component Combinatorial Library. Anal Chem. 2002;74(1):26–36.

19. Fernandez-Lima FA, Kaplan DA, Park MA. Note: integration of trapped ion mobility spectrometry with mass spectrometry. Rev Sci Instrum. 2011;82(12):126106.

20. Fernandez-Lima F, Kaplan DA, Suetering J, Park MA. Gas-phase separation using a trapped ion mobility spectrometer. Int J Ion Mobil Spectrom. 2011;14(2):93–98.

21. Ridgeway ME, Lubeck M, Jordens J, Mann M, Park MA. Trapped ion mobility spectrometry: a short review. Int J Mass Spectrom. 2018;425:22–35.

22. Meier F, Beck S, Grassl N, Lubeck M, Park MA, Raether O, et al. Parallel Accumulation-Serial Fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. J Proteome Res. 2015;14(12):5378–87.

23. Meier F, Park MA, Mann M. Trapped ion mobility spectrometry and parallel accumulation-serial fragmentation in proteomics. Mol Cell Proteom. 2021;20:100138.

24. Oliinyk D, Meier F. Ion mobility-resolved phosphoproteomics with dia-PASEF and short gradients. Proteomics. 2023;23(7-8):e2200032.

25. Skowronek P, Thielert M, Voytik E, Tanzer MC, Hansen FM, Willems S, et al. Rapid and in-depth coverage of the (phospho-) proteome with deep libraries and optimal window design for dia-PASEF. Mol Cell Proteom. 2022;21(9):100279.

26. Ogata K, Chang CH, Ishihama Y. Effect of Phosphorylation on the Collision Cross Sections of Peptide Ions in Ion Mobility Spectrometry. Mass Spectrom (Tokyo). 2021;10:A0093.

27. Mukherjee S, Jankevics A, Busch F, Lubeck M, Zou Y, Kruppa G, et al. Oxonium Ion-Guided Ion Mobility-Assisted Glycoproteomics on the timsTOF Pro. bioRxiv. 2022:2022.07.04.498688.

28. Ruotolo BT, Verbeck GFt, Thomson LM, Woods AS, Gillig KJ, Russell DH. Distinguishing between phosphorylated and non-phosphorylated peptides with ion mobility-mass spectrometry. J Proteome Res. 2002;1(4):303–6.

29. Thalassinos K, Grabenauer M, Slade SE, Hilton GR, Bowers MT, Scrivens JH. Characterization of phosphorylated peptides using traveling wave-based and drift cell ion mobility mass spectrometry. Anal Chem. 2009;81(1):248–54.

30. Glover MS, Dilger JM, Acton MD, Arnold RJ, Radivojac P, Clemmer DE. Examining the Influence of Phosphorylation on Peptide Ion Structure by Ion Mobility Spectrometry-Mass Spectrometry. J Am Soc Mass Spectrom. 2016;27(5):786–94.

31. Hinneburg H, Hofmann J, Struwe WB, Thader A, Altmann F, Varón Silva D, et al. Distinguishing N-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. Chem Commun (Camb). 2016;52(23):4381–4.

32. Creese AJ, Cooper HJ. Separation and identification of isomeric glycopeptides by high field asymmetric waveform ion mobility spectrometry. Anal Chem. 2012;84(5):2597–601.

33. Li Z, Dilger JM, Pejaver V, Smiley D, Arnold RJ, Mooney SD, et al. Intrinsic size parameters for palmitoylated and carboxyamidomethylated peptides. Int J Mass Spectrom. 2014;368:6–14.

34. Kaszycki JL, Shvartsburg AA. A priori intrinsic PTM size parameters for predicting the ion mobilities of modified peptides. J Am Soc Mass Spectrom. 2017;28(2):294–302.

35. Chang C-H, Yeung D, Spicer V, Ogata K, Krokhin O, Ishihama Y. Sequence-specific model for predicting peptide collision cross section values in proteomic ion mobility spectrometry. J Proteome Res. 2021;20(7):3600–10.

36. Zeng W-F, Zhou X-X, Willems S, Ammar C, Wahle M, Bludau I, et al. AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. Nat Commun. 2022;13(1):7238.

37. Zolg DP, Wilhelm M, Schnatbaum K, Zerweck J, Knaute T, Delanghe B, et al. Building proteometools based on a complete synthetic human proteome. Nat Methods. 2017;14(3):259–62.

38. Zolg DP, Wilhelm M, Schmidt T, Médard G, Zerweck J, Knaute T, et al. Proteometools: systematic characterization of 21 post-translational protein modifications by liquid chromatography tandem mass spectrometry (LC-MS/MS) using synthetic peptides. Mol Cell Proteom. 2018;17(9):1850–63.

39. Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, et al. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics. 2012;12(8):1111–21.

40. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367–72.

41. Prianichnikov N, Koch H, Koch S, Lubeck M, Heilig R, Brehmer S, et al. MaxQuant software for ion mobility enhanced shotgun proteomics. Mol Cell Proteom. 2020;19(6):1058–69.

42. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011;10(4):1794–805.

43. Sharma K, D'Souza Rochelle CJ, Tyanova S, Schaab C, Wiśniewski Jacek R, Cox J, et al. Ultradeep human phosphoproteome reveals a distinct regulatory nature of tyr and Ser/Thr-based signaling. Cell Rep. 2014;8(5):1583–94.

44. Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc. 1988;110(6):1657–66.

45. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B. 2001;105(28):6474–87.

46. Jorgensen WL. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J Am Chem Soc. 1981;103(2):335–40.

47. Bleiholder C, Osburn S, Williams TD, Suhai S, Van Stipdonk M, Harrison AG, et al. Sequence-scrambling fragmentation pathways of protonated peptides. J Am Chem Soc. 2008;130(52):17774–89.

48. Harrison AG, Young AB, Bleiholder C, Suhai S, Paizs B. Scrambling of sequence information in collision-induced dissociation of peptides. J Am Chem Soc. 2006;128(32):10364–5.

49. Bleiholder C, Suhai S, Harrison AG, Paizs B. Towards understanding the tandem mass spectra of protonated oligopeptides. 2: the proline effect in collision-induced dissociation of protonated Ala-Ala-Xxx-Pro-Ala (Xxx = Ala, Ser, Leu, Val, Phe, and Trp). J Am Soc Mass Spectrom. 2011;22(6):1032–9.

50. Stewart JJP. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. J Mol Model. 2007;13(12):1173–213.

51. Stewart JJP. MOPAC2016. Steward Computational Chemistry. Colorado Springs, CO, USA; 2016.

52. Bleiholder C, Wyttenbach T, Bowers MT. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method. Int J Mass Spectrom. 2011;308(1):1–10.

53. Deutsch EW, Bandeira N, Sharma V, Perez-Riverol Y, Carver JJ, Kundu DJ, et al. The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. Nucleic Acids Res. 2020;48(D1):D1145–52.

54. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022;50(D1):D543–52.

55. Revercomb HE, Mason EA. Theory of plasma chromatography/gaseous electrophoresis. Review. Anal Chem. 1975;47(7):970–83.

56. McDaniel EW, Viehland LA. The transport of slow ions in gases: Experiment, theory, and applications. Phys Rep. 1984;110(5):333–67.

57. Dickinson Q, Meyer JG. Positional SHAP (PoSHAP) for Interpretation of machine learning models trained from biological sequences. PLoS Comput Biol. 2022;18(1):e1009736.

58. Steigenberger B, van den Toorn HWP, Bijl E, Greisch JF, Räther O, Lubeck M, et al. Benefits of collisional cross section assisted precursor selection (caps-PASEF) for cross-linking mass spectrometry. Mol Cell Proteom. 2020;19(10):1677–87.

59. Henderson SC, Li J, Counterman AE, Clemmer DE. Intrinsic size parameters for Val, Ile, Leu, Gln, Thr, Phe, and Trp residues from ion mobility measurements of polyamino acid ions. J Phys Chem B. 1999;103(41):8780–5.

60. Valentine SJ, Ewing MA, Dilger JM, Glover MS, Geromanos S, Hughes C, et al. Using ion mobility data to improve peptide identification: intrinsic amino acid size parameters. J Proteome Res. 2011;10(5):2318–29.

61. Liu FC, Kirk SR, Caldwell KA, Pedrete T, Meier F, Bleiholder C. Tandem Trapped Ion Mobility Spectrometry/Mass Spectrometry (tTIMS/MS) reveals sequence-specific determinants of top-down protein fragment ion cross sections. Anal Chem. 2022;94(23):8146–55.

62. Meyer JG. Deep learning neural network tools for proteomics. Cell Rep Methods. 2021;1(2):100003.

63. Mann M, Kumar C, Zeng W-F, Strauss MT. Artificial intelligence for proteomics and biomarker discovery. Cell Syst. 2021;12(8):759–70.
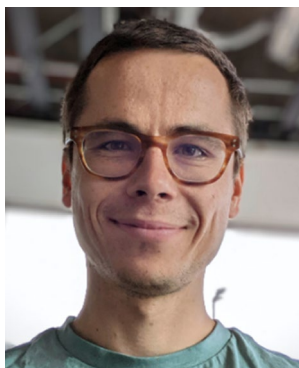
Andreas Will is currently a doctoral researcher at Jena University Hospital jointly supervised by Prof. Florian Meier and Prof. Regine Heller. His research focuses on the application of trapped ion mobility-mass spectrometry to detect post-translational modifications, with a particular interest in the role of O-GlcNAcylation during aging and cellular senescence.

Christian Bleiholder is Associate Professor of Chemistry and Biochemistry at Florida State University with a D.Sc. in chemistry (University of Heidelberg & German Cancer Research Center). His current research focuses on the development and application of ion mobility spectrometry/mass spectrometry methods. His awards include a fellowship from the Alexander-von-Humboldt Foundation and a CAREER award from the National Science Foundation.

Denys Oliinyk is currently a PhD student at Jena University Hospital in Prof. Florian Meier's group. His expertise is in the development of mass spectrometry-based methods to analyze protein phosphorylation networks in human health and disease at the systems scale.

Florian Meier is Junior Professor and independent research group leader at Jena University Hospital with a PhD in biochemistry from Ludwig Maximilian University Munich. His interest is in the development of (ion mobility) mass spectrometry-based omics technologies and their applications in biomedical research. His awards include the 2019 Wolfgang Paul Study Award of the German Mass Spectrometry Society.