

Profiling the proteome-wide selectivity of diverse electrophiles

Patrick R. A. Zanon¹, Fengchao Yu², Patricia Z. Musacchio³, Lisa Lewald¹, Michael Zollo¹, Kristina Krauskopf⁴, Dario Mrdović¹, Patrick Raunft¹, Thomas E. Maher¹, Marko Cigler⁴, Christopher J. Chang⁵, Kathrin Lang⁴, F. Dean Toste³, Alexey I. Nesvizhskii^{2,6}, and Stephan M. Hacker^{1,*}

¹ Department of Chemistry, Technical University of Munich, Garching, Germany

² Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA

³ Department of Chemistry, University of California, Berkeley, California, USA

⁴ Department of Chemistry, Group of Synthetic Biochemistry, Technical University of Munich, Garching, Germany

⁵ Departments of Chemistry and Molecular and Cell Biology, University of California, Berkeley, California, USA

⁶ Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA

* Correspondence: stephan.m.hacker@tum.de

Abstract: Targeted covalent inhibitors are powerful entities in drug discovery, but their application has so far mainly been limited to addressing cysteine residues. The development of cysteine-directed covalent inhibitors has largely profited from determining their proteome-wide selectivity using competitive residue-specific proteomics. Several probes have recently been described to monitor other amino acids using this technology and many more electrophiles exist to modify proteins. Nevertheless, a direct, proteome-wide comparison of the selectivity of diverse probes is still entirely missing. Here, we developed a completely unbiased workflow to analyse electrophile selectivity proteome-wide and applied it to directly compare 54 alkyne probes containing diverse reactive groups. In this way, we verified and newly identified probes to monitor a total of nine different amino acids as well as the *N*-terminus proteome-wide. This selection includes the first probes to globally monitor tryptophans, histidines and arginines as well as novel tailored probes for methionines, aspartates and glutamates.

Introduction

Targeted covalent inhibitors (TCIs) are powerful entities in drug discovery as they can have key advantages such as increased binding affinity to the target, the possibility to generate selectivity among closely related proteins and better pharmacodynamic properties.¹ Nevertheless, careful optimisation of the reactivity and selectivity of these inhibitors is key to avoiding toxicity as well as possible immunogenic reactions.¹

For TCIs that target cysteine residues, competitive residue-specific proteomics is an essential tool to tailor covalent inhibitors for high selectivity proteome-wide.^{2,3} The underlying methods are based on the ground-breaking isotopic tandem orthogonal proteolysis activity-based protein profiling (isoTOP-ABPP) platform.^{2,3} We have recently built on this platform by developing isotopically labelled desthiobiotin azide (isoDTB) tags to streamline the experimental procedure.⁴ In the underlying isoDTB-ABPP workflow⁴ (Fig. 1a), two samples of a proteome-of-interest are treated with a covalent ligand or the corresponding solvent as a control. Next, a broadly reactive alkyne probe is applied that labels many residues with alkynes. The residue(s) that are already engaged by the ligand are blocked from this reactivity and are not modified with alkynes. In the next step, isotopically differentiated isoDTB tags are attached using copper-catalysed azide-alkyne cycloaddition (CuAAC).⁵ As the proteins originating from the compound- and solvent-treated samples are now

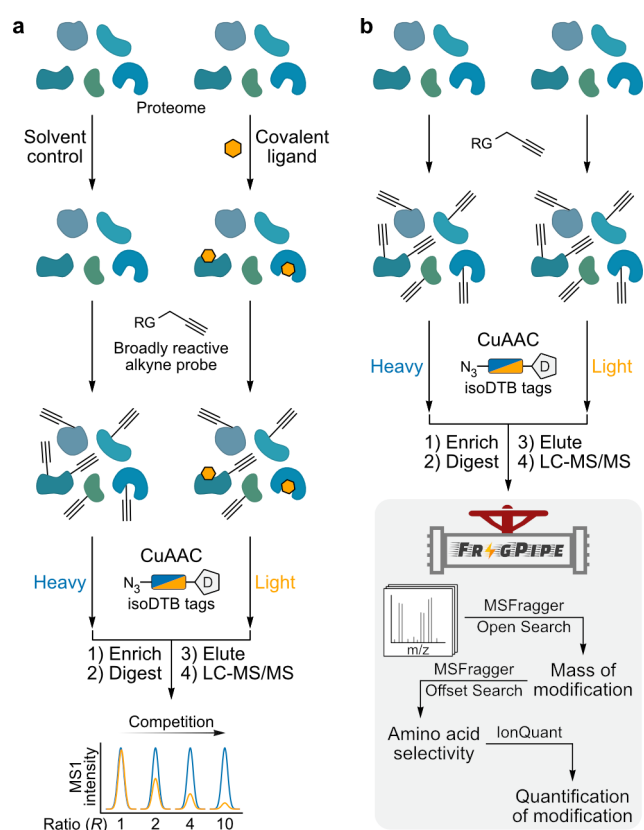


Fig. 1 | Workflows for competitive isoDTB-ABPP and the unbiased analysis of electrophile selectivity. **a**, Workflow for competitive, residue-specific chemoproteomic experiments using the isoDTB-ABPP workflow.⁴ RG = reactive group; D = desthiobiotin. **b**, Unbiased workflow to comprehensively investigate electrophile reactivity in the proteome using the MSFragger-based FragPipe computational platform.^{6,7}

differentiated, the two samples are mixed, enriched, proteolytically digested and the modified peptides eluted. Using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), the modified peptides are identified and quantified. Peptides containing residues that are engaged by the covalent ligand will show high ratios between the two samples ($R \gg 1$), while all unaffected peptides will show ratios close to 1. In this way, target engagement and selectivity of TCIs can be investigated proteome-wide in a quantitative fashion.⁴

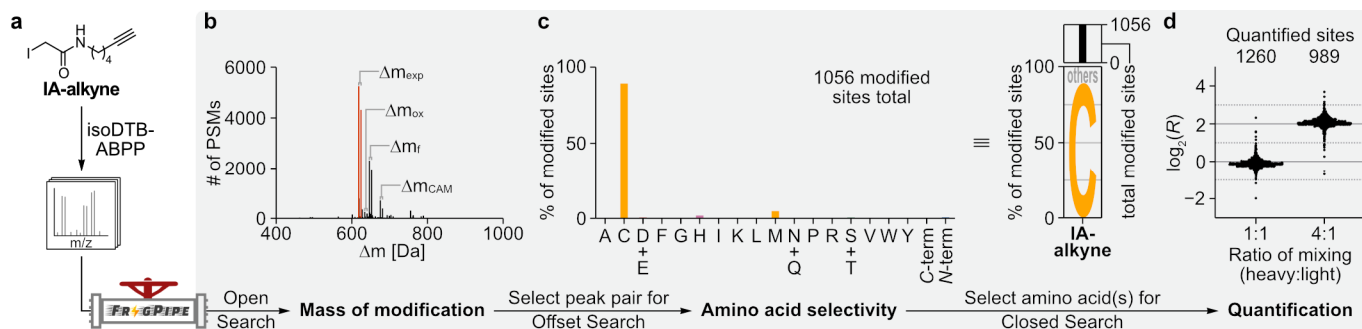


Fig. 2 | An unbiased workflow to study electrophile selectivity using the MSFragger-based FragPipe computational platform. **a**, Labelling of the proteome of *S. aureus* SH1000 with 1 mM IA-alkyne and analysis using the isoDTB-ABPP workflow (Fig 1b)⁴ results in MS data for the analysis of proteome-wide reactivity and selectivity. **b**, Through analysis with an Open Search in MSFragger,^{6,7} the masses of modification that occur proteome-wide are assigned. The peaks highlighted in red are the expected modifications (Δm_{exp}) resulting from alkylation and modification with the light and heavy isoDTB tags, respectively. Further modification of the alkylated peptides by oxidation (Δm_{ox}), formylation (Δm_{f}) or carbamidomethylation on a second cysteine (Δm_{CAM}) is also detected. PSMs = peptide spectrum matches. **c**, One peak pair (Δm_{exp}) is selected for an MSFragger Offset Search that localises this modification to the modified amino acid(s). In this way, selectivity is assessed across all proteinogenic amino acids. The bar graph represents the fraction of all modified sites that is modified at the indicated amino acid. The same data is also represented in a letter plot, in which the size of the letter is scaled by the fraction of all modified sites that is modified at the indicated amino acid. All amino acids that are modified in less than 5% of the cases are summarised as “others”. The total number of modified sites is given as a bar graph on top of the letter plot. **d**, One amino acid (cysteine) is selected for quantification at the selected masses (Δm_{exp}) using an MSFragger Closed Search and the IonQuant quantification module.⁸ Here, two datasets were analysed, in which the heavy and light samples were mixed at a ratio of 1:1 and 4:1, respectively. The grey, solid lines indicate the expected values of $\log_2(R) = 0$ and $\log_2(R) = 2$. The grey, dashed lines indicate the respective preferred window of quantification ($-1 < \log_2(R) < 1$ for the 1:1 ratio, $1 < \log_2(R) < 3$ for the 4:1 ratio). The total number of quantified sites is given on top of the plot. The mass spectrometric experiments used for this analysis were performed in an earlier study.⁴ All data is based on technical duplicates.

Despite the promise of TCIs to address diverse reactive amino acids, they so far almost exclusively address cysteine residues.⁹ However, as cysteine is a very rare amino acid, many interesting binding pockets contain no suitable cysteine for covalent engagement.¹⁰ Recently, many reactive groups have been developed that selectively target other residues in proteins and that have promise for TCI development against a much broader set of target proteins.¹¹ For these compounds, one key challenge is to globally investigate their target and amino acid selectivity. Alkyne-, azide- or (desthio)biotin-containing probes have been successfully developed in various cases to directly investigate global selectivity.¹² Nevertheless, besides often being synthetically challenging, this approach requires that the modification is stable at all modified residues, which excludes this strategy e.g. for acylating reagents that form unstable thioesters at cysteine, which evade direct detection.¹³ Similarly, this approach is not suitable for reversibly covalent inhibitors,^{14,15} whose modifications would also be lost before the analysis. In contrast, if one broadly reactive alkyne probe is developed that leads to a stable modification of an amino acid of interest, the targets for all of these compounds can be identified competitively. Therefore, being able to competitively profile target engagement at various amino acids using residue-specific proteomics would enable a global profile of the selectivity of TCIs regardless of their amino acid selectivity or the stability of their adducts.

To date, broadly reactive alkyne probes that can be used in this approach have been reported for cysteines,^{2,3,16,17} lysines,^{18,19} aspartates and glutamates,^{10,20} methionines^{21,22} as well as tyrosines.^{23,24} In the different underlying studies various strategies for the affinity tags, the isotopic labelling, the mass spectrometric instrumentation and the data analysis have been used. In this way, it is impossible to directly compare the reactivity and selectivity of these probes and to choose the best probe for a certain application. Furthermore, for other amino acids, broadly reactive alkyne probes are still entirely missing. To address this challenge, we here directly compared a large variety of electrophiles and established their proteome-wide amino acid selectivity using isoDTB-ABPP.⁴ This method is ideally suited to perform this analysis as it allows specific enrichment of the modified peptides to increase sensitivity of detection. Furthermore, it fosters confident assignment of probe-

labelled peptides due to the isotopic pattern introduced by the isoDTB tags.

As we are most interested in the application of TCIs for antibacterial applications,⁴ we performed the main part of our analysis in the lysate of *Staphylococcus aureus* SH1000.²⁵ We enhanced and developed new features for the MSFragger-based^{6,7} FragPipe computational platform (<https://fragpipe.nesvilab.org>), which now allows studying proteome-wide electrophile reactivity in a completely unbiased fashion (Fig. 1b). In this way, we verified or newly identified probes to competitively study a total of nine different amino acids as well as the N-terminus proteome-wide. For the chosen amino acid-selective probes, we verify similar selectivity in the human proteome indicating that amino acid selectivity is not dependent on the specific proteome. This set of probes will enable competitive profiling of TCIs not only against cysteine but against a variety of reactive amino acids and thereby guide the development of novel TCIs.

Results and Discussion

Tailoring the MSFragger-based FragPipe computational platform for unbiased analysis of proteome-wide electrophile selectivity. One important challenge for the widespread application of competitive residue-specific proteomics is that tailored software is not readily available to the broader community.²⁶ Furthermore, it was so far not possible to study electrophile reactivity and selectivity in a completely unbiased fashion. In order to address these challenges, we set out to tailor the MSFragger software and other components of the FragPipe computational platform for this application.^{6,7} We chose to use FragPipe as it is freely available to the academic community and as the ultrafast fragment-ion indexing method is especially powerful for the complex data analyses needed to identify and localise modifications on peptides in an unbiased fashion. To validate the new software features, we utilised a published dataset,⁴ in which 1 mM iodoacetamide alkyne (IA-alkyne) was used in a non-competitive isoDTB-ABPP workflow (Fig. 1b).

In the first step, we optimised FragPipe’s Open Search workflow to identify modified peptides and assign the masses of their respective modifications. For this purpose, we performed an Open Search using MSFragger^{6,7,27,28} followed by removal of mass shift artifacts with

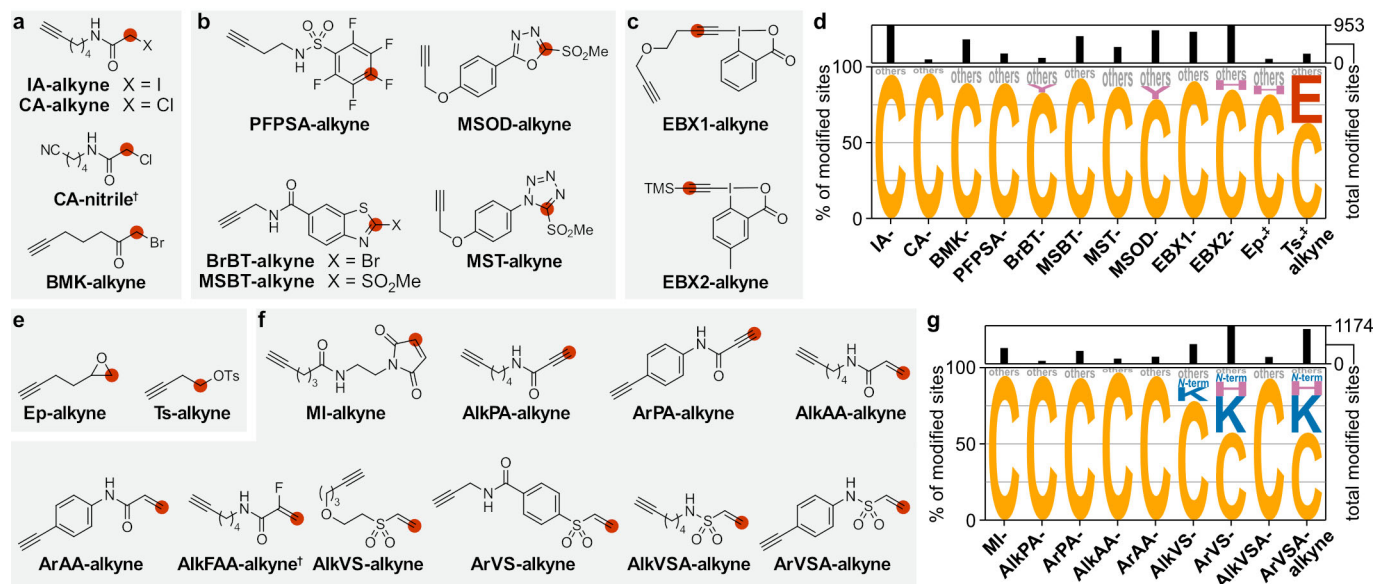


Fig. 3 | Amino acid selectivity of electrophiles targeting cysteines. a,b,c,e,f, Structures of alkyne-containing electrophilic probes that were investigated for their proteome-wide amino acid selectivity. Red circles indicate the initial site of electrophilic reactivity. d,g, Amino acid selectivity of the probes upon treatment of the proteome of *S. aureus* SH1000 at 100 μ M probe concentration. The data is represented as letter plots. The size of the letter is scaled by the fraction of all modified sites that is modified at the indicated amino acid. All amino acids that are modified in less than 5% of the cases are summarised as “others”. The total number of modified sites is given as a bar graph on top of the letter plot. †: No clear mass of modification was detected and therefore no analysis of the amino acid selectivity was possible. ‡: Data for the indicated probe at 1 mM is shown. All data is based on technical duplicates.

Crystal-C,²⁷ processing of peptide and protein identifications using PeptideProphet²⁹ and ProteinProphet,³⁰ false discovery-rate (FDR) filtering *via* Philosopher³¹ and summarisation of the mass shifts observed on peptides using PTM-Shepherd.²⁸ Using this procedure, we identified the expected alkylation with **IA-alkyne** as the main modification (Δm_{exp}) in the form of two peaks corresponding to the isotopically differentiated isoDTB tags (Fig. 2a,b). Furthermore, we detected a significant amount of additional formylation (Δm_f) of the modified peptides, which is known to occur when eluting, re-dissolving and storing peptides in solutions containing formic acid.³² Additionally, we identified smaller peaks that correspond to oxidation of the formed thioether (Δm_{ox}) or additional carbamidomethylation at a second cysteine residue (Δm_{CAM}). The detection of these expected minor modifications verifies that MSFragger, with its ability to perform accurate deisotoping³³ and mass calibration,⁷ can identify different modifications occurring after treatment with an electrophilic probe proteome-wide in an unbiased fashion with high mass accuracy. As the highest deviation of all of these masses of modification from the expected value was 0.0044 Da (6.9 ppm), the molecular formula can be directly deduced from the MS data for unknown modifications.

After having determined the mass of the modification, we performed an MSFragger Offset Search, which allows searching at the mass with an indicated offset from the mass of the unmodified peptide.^{6,7} Here, we chose an offset that accounts for the detected alkylation with **IA-alkyne** (Δm_{exp}). The modification is computationally localised to an amino acid residue or a stretch of amino acids without having to previously specify, which amino acids are potentially modified.⁷ Thus, the FragPipe computational platform allows to globally study selectivity towards all amino acid residues and the protein termini at once. After thorough data filtering, we verified that **IA-alkyne** shows high selectivity of 89% for cysteines with alkylation of methionines (5%) being the most prominent additional reactivity (Fig. 2c).

Finally, to use probes for competitive residue-specific proteomics, their relative intensity between the light and the heavy channel needs to be quantified. For this purpose, complex in-house software^{3,34} or

workarounds in existing software⁴ had to be used, so far. Therefore, we extended the quantification tool IonQuant⁸ – another FragPipe component – to allow quantification of modified peptides in MSFragger Offset Searches as well as in conventional Closed Searches, in which the modified residue is specified. To do so, we implemented a new functionality in IonQuant to perform relative quantification of isotopically labelled peaks to support stable isotope labelling-based quantification. In a Closed Search, we quantified 1260 cysteines with >99% in the preferred quantification window of $-1 < \log_2(R) < 1$ (Fig. 2d). For the Offset Search, we quantified 1896 modified peptides (99% in the preferred quantification window, Supplementary Fig. 1). It should be noted that the number of total modified sites, which is used for selectivity analysis based on an Offset Search (1056 total modified sites in this case, Fig. 2c), and the number of quantified residues in the Closed Search (1260 quantified cysteines in this case, Fig. 2d) or the Offset Search (1896 quantified peptides in this case, Supplementary Fig. 1) will be different in most cases, because of different data analysis workflows and downstream data filtering. We also analysed an additional published dataset,⁴ in which the heavy and light sample were mixed at a ratio of 4:1. Here, we quantified 989 cysteines with 98% in the preferred quantification window of $1 < \log_2(R) < 3$ (Fig. 2d). The high quality of this quantification data was comparable to data evaluation with MaxQuant³⁵ using our previously described workaround⁴ or pFind 3³⁶ using a custom script for downstream analysis³⁴ (Supplementary Fig. 2). Importantly, our automated FragPipe workflow simplifies the data analysis and allows analysis for probes that are not selective for a certain amino acid type.

Overall, the optimised FragPipe computational platform enables the completely unbiased analysis of residue-specific proteomic data obtained with various probes including identification of the mass of the modification, its amino acid selectivity and its use for quantitative applications. Having this unbiased analysis workflow at hand, we applied a standardised sample preparation protocol for all probes throughout this project to directly compare them regarding their proteome-wide reactivity and selectivity. Here, two identical samples of *S. aureus* lysate were treated with 100 μ M of the respective probe,

modified with 100 μ M of the light or heavy isoDTB tag, respectively, mixed at a ratio of 1:1 and analysed using the isoDTB-ABPP workflow.

Diverse chemistries allow monitoring of cysteines. Using the described standard conditions for **IA-alkyne**³ (Fig. 3a), we detected 95% selectivity for cysteines and quantified 1197 cysteines (Fig. 3d, Supplementary Fig. 3 and 4). An increase of the concentration to 1 mM resulted in 86% selectivity for cysteines. Chloroacetamide **CA-alkyne**³ and α -bromomethyl ketone **BMK-alkyne**³⁷ (Fig. 3a) also demonstrated high selectivity for cysteines (96% and 89%) and allowed quantification of 230 and 976 cysteines, respectively (Fig. 3d, Supplementary Fig. 3 and 4). Importantly, a chloroacetamide negative control (**CA-nitrile**, Fig. 3a), with a nitrile replacing the alkyne, did not yield a clear mass of modification (Supplementary Fig. 3).

Nucleophilic aromatic substitution is an alternative chemistry to label cysteine residues in proteins (**PFPSA**,³⁸ **BrBT**-, **MSBT**-, **MST**- and **MSOD-alkyne**,¹⁶ Fig. 3b). All of the studied probes showed high cysteine selectivity (70-92%) and allowed quantification of 362-1061 cysteines (Fig. 3d, Supplementary Fig. 5 and 6). We also investigated hypervalent iodine reagents that have been described for labelling of cysteines in the proteome.¹⁷ Using the originally reported chemistry (**EBX1-alkyne**, Fig. 3c, Supplementary Fig. 7),³⁹ we detected several different modifications (Supplementary Fig. 8).⁴⁰ While all of these showed high cysteine selectivity (91-97%, Fig. 3d, Supplementary Fig. 8 and 9), this complex modification behaviour complicates downstream data analysis. Therefore, we turned to the recently published probe **EBX2-alkyne**⁴¹ (Fig. 3c), which selectively led to the minimal modification with an ethynyl group (Supplementary Fig. 8, 9 and 10). This modification showed high selectivity for cysteines (85%) and allowed quantification of 1251 cysteines in the proteome (Fig. 3d, Supplementary Fig. 8 and 9). Finally, we also turned to nucleophilic substitution reactions at unactivated sp³-carbon centres (**Ep**- and **Ts-alkyne**, Fig. 3e). Both probes showed a clear preference for cysteines, but for **Ts-alkyne** a significant number of modifications was also found at glutamates (Fig. 3d, Supplementary Fig. 11 and 12).

When we combined the quantified cysteines of all these probes, we in aggregate quantified 1941 cysteines in *S. aureus* covering 37% of the 5268 cysteines encoded in the *S. aureus* genome. **IA-alkyne** can still be considered the gold standard for monitoring cysteine residues with residue-specific proteomics *in vitro*. Nevertheless, we verified that many other complementary probes exist that further increase the coverage with **EBX2**- and **MSBT-alkyne** being especially powerful (Supplementary Fig. 13).

Michael acceptors preferentially react with cysteines. *Michael* acceptors are currently mainstays for the design of TCIs (Fig. 3f). For a maleimide probe (**MI-alkyne**, after hydrolysis of the resulting succinimide during the workflow) and for propiolamide probes (**AlkPA**- and **ArPA-alkyne**), we detected high cysteine selectivity (93-95%) and quantified a total of 283-752 cysteines (Fig. 3g, Supplementary Fig. 14 and 15). We also observed the expected modification and strong proteomic labelling (160-1174 localised sites, Fig. 3f,g, Supplementary Fig. 16-19) for all acceptor-substituted terminal alkenes except for **AlkFAA-alkyne**. Strikingly, we observed different degrees of selectivity. While **AlkAA-alkyne** labelled cysteine residues with 97% selectivity, less than 60% of labelled sites were cysteines for **ArVSA**- and **ArVS-alkyne** (>20% lysines, ~9% histidines, ~5% *N*-termini). Interestingly, the labelling of *N*-termini showed a preference for the secondary amine of proline when the initial methionine was removed (Supplementary Fig. 19).⁴² This data demonstrates that, even though *Michael* acceptors can be

designed to also target other amino acid residues, cysteines are the major modification sites in all studied cases. In this way, while lysine-selective TCIs have been developed based on *Michael* acceptors,⁴³ monitoring their cysteine off-targets is an important prerequisite for their further development.

Studying lysine residues proteome-wide. The activated ester probe **STP-alkyne** (Fig. 4a) allows selective monitoring of many lysine residues in the proteome.¹⁸ Using our workflow, we verified its high selectivity for lysines (78%, Fig. 4b, Supplementary Fig. 20 and 21), which led to the quantification of 3277 lysines. The remaining peptides were mainly labelled at serines (9%), threonines (2%) or *N*-termini (5%). This selectivity was fully retained even at 1 mM probe concentration. It is noteworthy that peptides labelled at threonine preferentially have a histidine at position -2 and those labelled at serine show a preference for cysteine at position +2, arginine at position -1 and histidine at position -2 indicating that these sequence contexts might increase reactivity towards **STP-alkyne** (Supplementary Fig. 21).⁴² We also studied four additional acylation reagents (**TFP**-,¹⁸ **NHS**-,¹⁹ **ATT**-⁴⁴ and **NASA-alkyne**,⁴⁵ Fig. 4a) and detected an overall similar selectivity, which allowed quantification of 2145-4404 lysines (Fig. 4b, Supplementary Fig. 20 and 22). All of these probes are therefore suitable to study lysines proteome-wide and hold potential especially for applications that are hampered by the negative charge of **STP-alkyne**. All of the probes showed similar preferences in the sequence around labelled serines and threonines as seen for **STP-alkyne** (Supplementary Fig. 22). As there are no tools to globally study serines and threonines in a residue-specific fashion and as tools to study the *N*-terminus lead to limited coverage (*vide infra*), it is furthermore noteworthy that **STP-alkyne** also quantified 428 serines, 165 threonines and 152 *N*-termini. This points to the fact that, while **STP-alkyne** mainly labels lysines, it could also moonlight to detect effects on these other amino acids.

Besides acylation reagents, squaric acid derivatives (**AlkSq**- and **ArSq-alkyne**, Fig. 4c) also react with amines under physiological conditions.⁴⁶ Both probes demonstrated very high selectivity for lysines (93% and 90%, Fig. 4b, Supplementary Fig. 23 and 24). Therefore, **ArSq-alkyne** is a promising broadly reactive alkyne probe for lysines (2990 quantified lysines), while structures like **AlkSq-alkyne** might be interesting for TCI design due to tempered reactivity (1339 quantified lysines). Another reactivity of lysines is the formation of imines with aldehydes, but irreversible stabilisation of the product is needed for the application as broadly reactive alkyne probe. Here, we were enticed by 2-ethynyl-benzaldehyde-based probes like **EBA-alkyne** (Fig. 4d) that were previously shown to react with amines in proteins to form an imine that cyclises to yield an irreversible isoquinolinium salt (Supplementary Fig. 25).⁴⁷ Exhibiting high lysine selectivity (81%) and good proteomic coverage (3796 quantified lysines, Fig. 4b, Supplementary Fig. 26 and 27), **EBA-alkyne** is an additional probe promising for lysine monitoring. Importantly, all of these probes also quantified a significant number of *N*-termini (243 for **AlkSq-alkyne**, 252 for **ArSq-alkyne** and 253 for **EBA-alkyne**), which demonstrates monitoring of *N*-termini as another possible application for these probes.

Light-activated probes hold the potential to monitor covalent target engagement in living cells as the probe itself is unreactive before activation and therefore has the potential to be non-toxic.¹⁰ Proteome-wide, residue-specific monitoring of lysines in a light-dependent fashion would therefore be highly promising to study effects on lysines with temporal and spatial resolution in living cells. Nevertheless, no probe had been previously shown to allow this application. For this purpose, we set out to study **oNBA-alkyne** that had been shown to react with lysines in live cells after formation of a nitrosobenzaldehyde upon irradiation (Fig. 4e, Supplementary

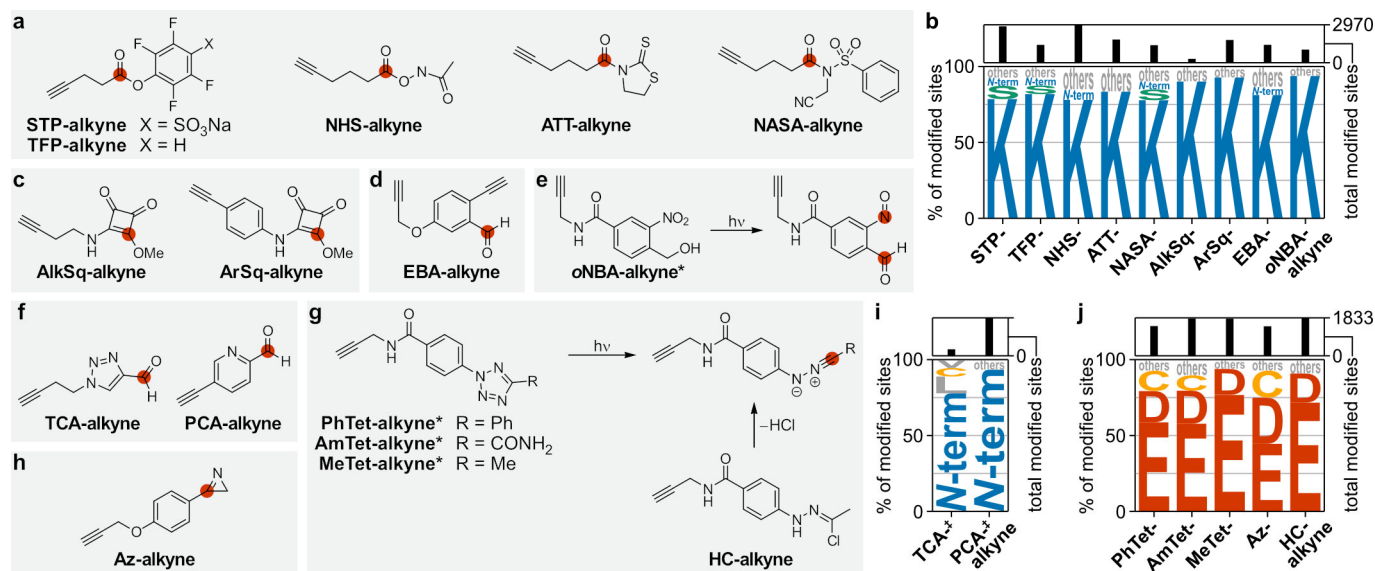


Fig. 4 | Amino acid selectivity of electrophiles targeting lysines, N-termini, aspartates and glutamates. a,c,d,e,f,g,h, Structures of alkyne-containing electrophilic probes that were investigated for their proteome-wide amino acid selectivity. Red circles indicate the initial site of electrophilic reactivity. For probes **oNBA-alkyne**, **PhTet-alkyne***, **AmTet-alkyne***, **MeTet-alkyne*** and **HC-alkyne**, the reactions leading to the reactive intermediate are also shown. b,i,j, Amino acid selectivity of the probes upon treatment of the proteome of *S. aureus* SH1000 at 100 μ M probe concentration. The data is represented as letter plots. The size of the letter is scaled by the fraction of all modified sites that is modified at the indicated amino acid. All amino acids that are modified in less than 5% of the cases are summarised as “others”. The total number of modified sites is given as a bar graph on top of the letter plot. *: Labelling was performed using UV-activation at 280-315 nm (**PhTet-alkyne***, **AmTet-alkyne*** and **MeTet-alkyne***) or 365 nm (**oNBA-alkyne**) for 10 min. †: Data for the indicated probe at 1 mM is shown. All data is based on technical duplicates.

Fig. 28).⁴⁸ Indeed, we detected the expected modification (Supplementary Fig. 29 and 30) and a high degree of lysine selectivity (94%) for this probe, which was retained even at 1 mM probe concentration (93%, Fig. 4b, Supplementary Fig. 29 and 30). Therefore, this is the first demonstration that **oNBA-alkyne** can act as a photoprobe to globally study lysines using residue-specific proteomics (1456 quantified lysines).

Taken the data of all lysine-directed probes together, we quantified 9129 lysines, which covers 15% of the 62,166 lysines encoded in the genome of *S. aureus*. Although **STP-alkyne** remains the reagent of choice to study lysines for standard applications, **ArSq-**, **EBA-** and **oNBA-alkyne** also display high selectivity using complementary chemistries. These probes will be important not only for increasing overall coverage (Supplementary Fig. 31), but also to allow specific applications, especially within living cells.

Global monitoring of N-termini of proteins. Although several lysine-directed probes like **STP-**, **ArSq-** and **EBA-alkyne** also allow monitoring of N-termini, selective chemistry is highly desirable. Here, carboxaldehydes of electron-poor heteroaromatics have been previously applied to modify proteins (**TCA-**⁴⁹ and **PCA-alkyne**,⁵⁰ Fig. 4f). While we were not able to detect more than a few sites for **TCA-alkyne** in the proteome even at 1 mM, **PCA-alkyne** resulted in the expected modification by initial formation of an imine with the N-terminus followed by cyclisation involving the first amide nitrogen of the backbone (Supplementary Fig. 32 and 33). This modification showed high selectivity for the N-terminus (93%) and allowed quantification of 167 N-termini at 1 mM probe concentration (Fig. 4i, Supplementary Fig. 33 and 34). Even though the coverage certainly still needs to be improved, **PCA-alkyne** is a suitable starting point to selectively monitor the N-terminus proteome-wide. Taking the data of **PCA-** (1 mM), **STP-**, **ArSq-** and **EBA-alkyne** together, we were able to quantify 464 N-termini in 412 proteins (some proteins were detected with and without clipping of the N-terminal methionine) covering 14% of the 2959 proteins encoded in the *S. aureus* genome.

Monitoring aspartates and glutamates in the proteome.

Recently, we have developed 2,5-disubstituted tetrazoles (Fig. 4g) into light-activatable tools to globally study aspartates and glutamates in living bacteria.¹⁰ Using our workflow, we verified the expected modification by formation of a nitrilimine upon light irradiation and subsequent reactivity eventually leading to an diacylated hydrazine (**PhTet-**, **AmTet-** and **MeTet-alkyne**, Fig. 4g, Supplementary Fig. 35 and 36). This modification showed a strong preference for aspartates and glutamates (79-94%, Fig. 4j, Supplementary Fig. 36-38). The best probe, **MeTet-alkyne**, quantified 2192 aspartates and glutamates and its high selectivity was retained at 1 mM probe concentration (93%, Supplementary Fig. 36 and 38). While peptides labelled at aspartate were consistently enriched for aliphatic amino acids at position +1 (leucine, valine, isoleucine), those labelled at glutamate showed a preference for valine and alanine at positions -1 and -2 (Supplementary Fig. 37 and 38).⁴² When prolonging the photoreaction with **MeTet-alkyne**, we observed a decrease in the expected modification and an increasing additional modification that could be assigned as arylation, which localised to all aromatic amino acids and to cysteines (Supplementary Fig. 36 and 38). While this chemistry might therefore also be interesting to monitor aromatic amino acids, care must be taken to not overextend irradiation times when monitoring aspartates and glutamates with **MeTet-alkyne**.

2H-Azirines have been described as constitutively active electrophiles to target aspartates and glutamates by initial nucleophilic attack at the azirine eventually leading to an acylated α -aminoketone (**Az-alkyne**,²⁰ Fig. 4h, Supplementary Fig. 39). Using our workflow, we could show that besides the expected reactivity (Δm_{exp}), these probes also showed a mass of modification with a mass of $\Delta m_{exp}+1$ Da, which formally corresponds to hydrolysis leading to a loss of ammonia (Supplementary Fig. 39 and 40). As computationally differentiating these modifications was not fully possible, we further analysed their combined selectivity (Fig. 4j, Supplementary Fig. 40 and 41) and identified 75% of all modifications at aspartates and glutamates and 18% at cysteines. **Az-alkyne** is therefore a valuable probe to study aspartates and

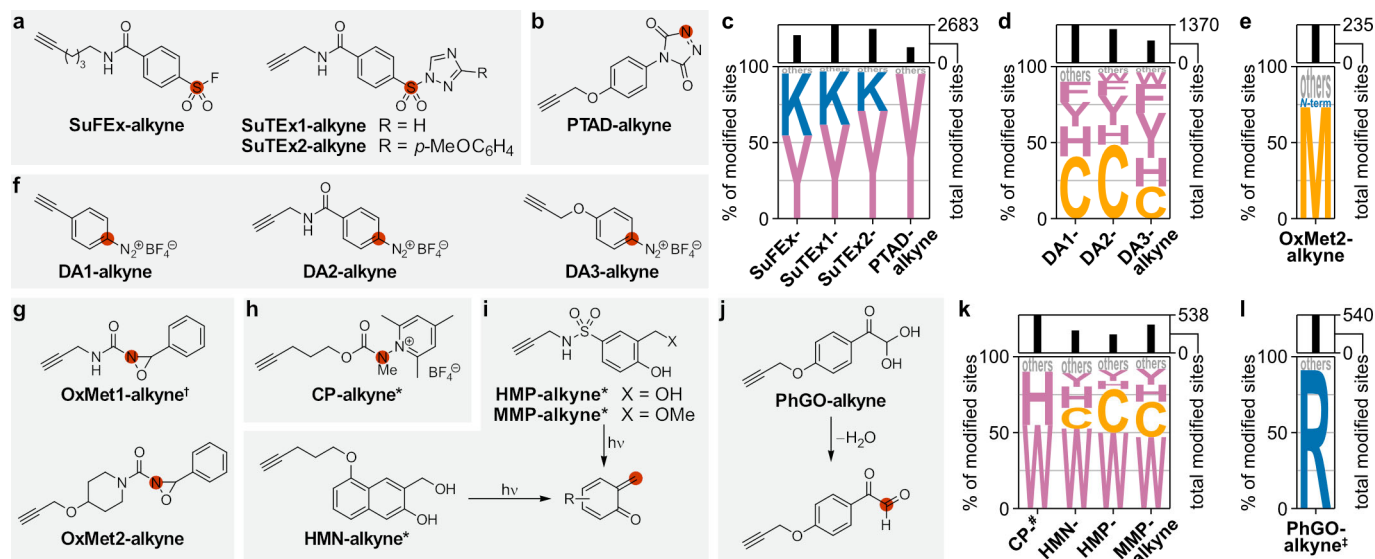


Fig. 5 | Amino acid selectivity of electrophiles targeting tyrosines, aromatic amino acids, methionines, tryptophans, histidines and arginines. a,b,f,g,h,i,j. Structures of alkyne-containing electrophilic probes that were investigated for their proteome-wide amino acid selectivity. Red circles indicate the initial site of electrophilic reactivity. For probes **HMN-**, **HMP-**, **MMP-** and **PhGO-alkyne**, the reactions leading to the reactive intermediate are also shown. c,d,e,k,l. Amino acid selectivity of the probes upon treatment of the proteome of *S. aureus* SH1000 at 100 μ M probe concentration. The data is represented as letter plots. The size of the letter is scaled by the fraction of all modified sites that is modified at the indicated amino acid. All amino acids that are modified in less than 5% of the cases are summarised as “others”. The total number of modified sites is given as a bar graph on top of the letter plot. †: No clear mass of modification was detected and therefore no analysis of the amino acid selectivity was possible. *: Labelling was performed using UV-activation at 280-315 nm (**CP-**, **HMP-** and **MMP-alkyne**) or 365 nm (**HMN-alkyne**) for 10 min. ‡: Data for the indicated probe at 1 mM is shown. #: Labelling was performed in degassed lysate under argon. All data is based on technical duplicates.

glutamates, but care must be taken to account for cysteine off-targets.

Enticed by the high selectivity of **MeTet-alkyne**, we also explored similar chemistry for monitoring aspartates and glutamates with constitutively active electrophiles. Hydrazonoyl chlorides such as **HC-alkyne** (Fig. 4g) have been shown to liberate the same reactive nitrilimine as the light-activated tetrazoles (Supplementary Fig. 35).⁵¹ We detected the expected modification and high selectivity for aspartates and glutamates (91%), which was also retained at 1 mM probe concentration (92%, Fig. 4j, Supplementary Fig. 42 and 43). The sequence around the labelled amino acids showed the same preferences as also seen for **MeTet-alkyne** (Supplementary Fig. 43).⁴² **HC-alkyne** allowed the quantification of 2450 aspartates and glutamates in the proteome at 100 μ M and is therefore a novel promising probe to study these amino acids without the need for light-activation.

Interestingly, **MeTet-** and **HC-alkyne** showed a strong preference to label glutamates (77% and 71% over aspartates (17% and 19%), while **Az-alkyne** demonstrated a smaller difference (44% and 30%, respectively) indicating that it is better able to also react with the sterically more hindered aspartate. Taking all carboxylic acid-directed probes together, 7811 aspartates and glutamates were quantified corresponding to 7.8% of the 100,780 aspartates and glutamates encoded in the *S. aureus* genome. Specifically, **MeTet-**, **HC-** and **Az-alkyne** constitute a set of complementary probes that allow a deep profiling of these amino acids in the bacterial proteome (Supplementary Fig. 44). Considering that no probes exist to selectively monitor C-termini proteome-wide, it is furthermore noteworthy that all carboxylic acid-directed probes together quantified 179 C-termini with **MeTet-** and **HC-alkyne** at 1 mM being most promising (101 and 109 quantified C-termini).

Residue-specific proteomics at tyrosines. Tyrosines offer a unique opportunity for various selective chemistries through reactions with the hydroxyl group as well as with the electron-rich aromatic system. Recently, sulfonylation of the hydroxyl group using

sulfur-fluoride (**SuFEx-alkyne**)⁵² and sulfur-triazole exchange chemistry (**SuTex1-** and **SuTex2-alkyne**)^{23,24} has been established for proteome-wide approaches (Fig. 5a). We verified the tyrosine reactivity of these probes (55-71%, Fig. 5c, Supplementary Fig. 45 and 46) with lysine residues being the most prominent off-targets (26-41%). In agreement with a previous study in human proteomes, **SuTex2-alkyne** showed the highest tyrosine selectivity and allowed quantifying 2653 tyrosines in bacterial lysates.

For labelling of the aromatic system of tyrosines in proteins, reagents like **PTAD-alkyne** have been established (Fig. 5b, Supplementary Fig. 47).⁵³ We identified the expected adduct that shows high selectivity for tyrosines (95%, Fig. 5c, Supplementary Fig. 48 and 49). Furthermore, we also detected a modification that corresponds to fragmentation of **PTAD-alkyne** to the isocyanate and subsequent reactivity, which showed some selectivity for lysines and N-termini and could be strongly reduced through addition of excess primary amine (Supplementary Fig. 48 and 49).⁵⁴ Due to the exquisite selectivity of the expected modification, it would be interesting to optimise the stability of these reagents to further reduce this side reactivity.

Through combination of the data for all tyrosine probes, we quantified 3968 tyrosines covering 12% of the 32,172 tyrosines encoded in the *S. aureus* genome. While **SuTex2-alkyne** is currently the probe of choice for tyrosines, reagents like **PTAD-alkyne** should allow for the development of optimised complementary probes in the future (Supplementary Fig. 50).

Diazonium salts perform arylation chemistry in the bacterial proteome. During the investigation of tyrosine-directed chemistries, we also considered aryl diazonium salts that have been shown to lead to azo coupling on tyrosines in isolated proteins (Fig. 5f, Supplementary Fig. 51).⁵⁵ Proteome-wide, we only detected minor azo coupling and almost exclusively arylation, corresponding to a formal loss of molecular nitrogen (Supplementary Fig. 51-53).⁵⁶ Strikingly, next to modifications on cysteines,⁵⁷ this also led to up to 75% of all modifications being localised to aromatic amino acids for

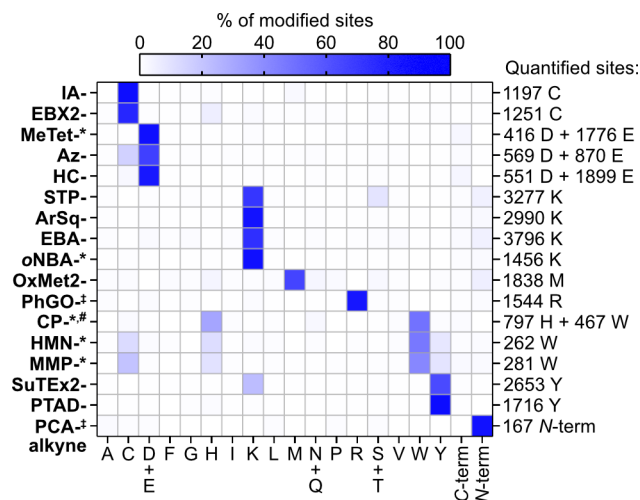


Fig. 6 | A set of 17 electrophilic probes that enables studying nine different amino acid residues and the N-terminus. For a selection of the probes that allow studying diverse residues in the proteome of *S. aureus*, their selectivity is plotted in a heatmap. The colour is scaled by the fraction of all modified sites that is modified at the indicated amino acid. *: Labelling was performed using UV-activation at 280-315 nm (MeTet-, CP- and MMP-alkyne) or 365 nm (oNBA- and HMN-alkyne) for 10 min. ‡: Data for the indicated probe at 1 mM is shown. #: Labelling was performed in degassed lysate under argon. All data is based on technical duplicates.

DA3-alkyne (19% phenylalanines, 19% histidines, 8% tryptophans, 30% tyrosines, Fig. 5d, Supplementary Fig. 52 and 53). In this way, **DA3-alkyne** enabled the quantification of 1218 aromatic amino acids. While aryl diazonium salts are therefore not suitable for azo couplings in bacterial proteomes, it is noteworthy that they can be used to monitor various aromatic amino acids and that preferred reactivity is possible at many residues in the proteome including even the unactivated aromatic system of phenylalanines.

An optimised tool for global monitoring of methionine residues.

Hypervalent iodine reagents have been described to monitor methionines but require an additional reaction step to give a stable modification.²² Therefore, we focused on oxaziridines for labelling of methionines, which proceeds through initial attack of the thioether at the nitrogen of the oxaziridine eventually giving a sulfimide (Supplementary Fig. 54).²¹ The initially described reagent **OxMet1-alkyne** (Fig. 5g) did not result in detection of the expected modification using our standard workflow (Supplementary Fig. 55). Implementing recently described design principles for more stable methionine modification,⁵⁸ we synthesised **OxMet2-alkyne** (Fig. 5g), which led to the detection of a high number of modified peptides with a preference for methionine modification (73%, 1838 quantified methionines, 8.5% of the 21,677 methionines encoded in the *S. aureus* genome, Fig. 5e, Supplementary Fig. 55 and 56). In this way, we developed **OxMet2-alkyne** as a tailored reagent for proteome-wide monitoring of methionines.

The first technologies to monitor tryptophans and histidines proteome-wide.

Recently, *N*-carbamoylpyridinium salts like **CP-alkyne** (Fig. 5h) have been introduced to photochemically label tryptophans in proteins through photoinduced electron transfer leading to the attachment of the carbamoyl group to tryptophans (Supplementary Fig. 57).⁵⁹ However, proteome-wide studies were still missing. Upon irradiation under protective gas, **CP-alkyne** showed the expected mass of modification in the proteome with almost complete selectivity for tryptophans (55%) and histidines (35%, Fig. 5k, Supplementary Fig. 58 and 59) allowing the quantification of 467 tryptophans and 797 histidines. The selectivity was retained, when the reaction was run open to air or at 1 mM

CP-alkyne (Supplementary Fig. 58 and 59). While we could not detect any preferences in the sequence around labelled histidine, peptides modified at tryptophan consistently showed an enrichment of glutamate in the -1-position (Supplementary Fig. 59).⁴² The reactivity with histidines is especially noteworthy, as we were unsuccessful to detect the expected modification with reported histidine-selective thiophosphorodichloridates (**TPAC-alkyne**, Supplementary Fig. 60), most likely due to instability of the conjugate during our workflow.⁶⁰ The fraction of >35% of histidine labelling for **CP-alkyne** is the highest we have detected for any probe, making it the probe of choice to study histidines proteome-wide. Taking all conditions for **CP-alkyne** together, we quantified 1697 histidines covering 9.1% of the 18,746 histidines in the *S. aureus* genome. **CP-alkyne** is, therefore, the first probe to globally monitor both tryptophans and histidines using residue-specific proteomics.

Simultaneously, we also investigated UV-activatable *o*-quinone methide precursors for tryptophan labelling through a formal [4+2]-cycloaddition (**HMN-**, **HMP-** and **MMP-alkyne**, Fig. 5i, Supplementary Fig. 61). Also in this case, we observed a preference for tryptophans (47-53%, Fig. 5k, Supplementary Fig. 62 and 63) with cysteines (14-29%), histidines (5-13%) and tyrosines (7-10%) being the main off-targets. In this way, **HMN-** and **MMP-alkyne** are complementary probes to study tryptophans in a proteome-wide sense. Through the combination of the data for all probes, we quantified a total of 701 tryptophans covering 11% of the 6183 tryptophans encoded in the *S. aureus* genome (Supplementary Fig. 64).

A first-in-class method to monitor arginines in the whole proteome.

Although arginine has been targeted with glyoxal-based reagents for bioconjugation⁶¹ and crosslinking before,⁶² no method is currently available to globally monitor arginines with residue-specific proteomics. We therefore synthesised **PhGO-alkyne** based on the known reactivity of phenylglyoxals with arginines that eventually produces a stable imidazole derivative (Fig. 5j, Supplementary Fig. 65). Using our workflow, we detected only minor modification of the proteome with the expected modification accompanied by an oxidation product (Supplementary Fig. 66).⁶³ At increased concentration of 1 mM **PhGO-alkyne**, however, we detected many modifications of the proteome (1544 arginines quantified, 5.4% of the 28,550 arginines encoded in the *S. aureus* genome) with the expected mass and high arginine selectivity (91%) alongside some oxidation (Fig. 5l, Supplementary Fig. 66 and 67). While the presence of the oxidation product still necessitates some optimisation, **PhGO-alkyne** is the first probe to globally monitor arginine in the proteome.

A set of probes to globally study nine amino acids and the N-terminus.

Through screening of 54 alkyne probes, we have identified a set of 17 probes that we currently consider ideal for profiling nine different amino acids and the *N*-terminus proteome-wide (Fig. 6). The fraction of all residues of a certain amino acid that we quantified using this probe set ranged from 2.5% of all aspartates to 29% of all cysteines encoded in the genome of *S. aureus*. In total, we were able to quantify 20,558 different sites in the proteome using our probe selection. These sites cover 1399 of 2959 proteins encoded in the *S. aureus* genome (54%) and 85% of the annotated essential proteins (301/353).⁶⁴ In this way, our probe selection allows us to gain very deep insights into the bacterial proteome.

To show that the application of our probe selection is not restricted to bacterial systems, we also applied these probes in lysates of the human cancer cell line MDA-MB-231. Except for **PCA-alkyne**, we reproduced similar selectivities for all probes in the human proteome

(Supplementary Fig. 68-76). In this way, we were able to quantify a total of 23,831 sites in 3770 proteins in this human cell line.

Conclusion

We report here on a universal, unbiased workflow to study the reactivity and selectivity of electrophilic probes in a proteome-wide setup. For this purpose, we extended and established new components of the broadly available MSFragger-based FragPipe computational platform.^{6,7} These include an optimised workflow for filtering of Open Searches to reliably identify masses of modification, a new workflow for downstream data analysis of Offset Searches to investigate selectivity at all proteogenic amino acids simultaneously and a new feature for IonQuant that now allows to quantify data from stable isotopic labelling-based MS experiments. In this way, we are for the first time able to identify the added masses of modification, their selectivity at all proteinogenic amino acids and their relative abundance in a completely unbiased fashion.

While our selection of 54 alkyne probes is certainly not comprehensive, we are convinced that we covered most chemistries that have more broadly been applied for protein labelling and that do not require additional reagents. In this way, we verified and newly identified tailored probes to study nine different amino acids as well as the *N*-terminus proteome-wide. Within our set of probes, we verify the selectivity of **IA-** and **EBX2-alkyne** for cysteines,^{2,3} **STP-alkyne** for lysines,¹⁸ **SuTeX2-alkyne** for tyrosines²³ as well as **MeTet**¹⁰ and **Az-alkyne**²⁰ for aspartates and glutamates. For several of these amino acids, we identify additional probes based on complementary chemistries like **ArSq-** and **EBA-alkyne** for lysines, **HC-alkyne** for aspartates and glutamates and **PTAD-alkyne** for tyrosines, which increase the proteome-wide coverage. For lysines, we demonstrate for the first time that proteome-wide residue-specific profiling is possible in a light-dependent fashion using **oNBA-alkyne**, which opens up exciting opportunities for in cell studies. Additionally, we developed a tailored probe (**OxMet2-alkyne**) for methionines that allows application in our workflow with high coverage due to increased stability of the protein modification. Furthermore, we transferred the concept of labelling proteins at the *N*-terminus to global residue-specific studies using **PCA-alkyne**. Strikingly, our study has enabled us to develop probes that for the first time allow studying tryptophans (**CP-**, **HMN-**, **MMP-alkyne**), histidines (**CP-alkyne**) and arginines (**PhGO-alkyne**) in a residue-specific fashion in the whole proteome.

Important residues, for which no selective broadly reactive alkyne probes could be established here, are serines and threonines as well as C-termini. For serines and threonines, bioconjugation to proteins using P(V)-based reagents for phosphorus-sulfur incorporation (PSI) has recently been established.^{65,66} Nevertheless, we were unable to detect a clear modification for **PSI-alkyne** (Supplementary Fig. 77 and 78), most likely due to instability of the conjugate during our workflow. In our hands, **STP-alkyne** is still the best way to monitor at least some serines and threonines in a residue-specific fashion. For C-termini, the aspartate- and glutamate-directed probes allowed us to quantify 179 C-termini, but increased coverage and selective probes are surely desirable here.

One important application of our probe selection will be to competitively study target engagement for any covalently reactive protein ligand regardless of its amino acid selectivity and adduct stability. This will be especially important for ligands, for which probe synthesis is challenging or which form covalent adducts that evade direct detection like those forming unstable modifications such as thioesters¹³ or covalently reversible protein adducts.^{14,15} It is tempting

to speculate that in the future it will be possible to use a mixture of several probes to competitively profile the selectivity of a protein ligand simultaneously at various amino acids within the same proteomic experiment as has recently been described for the combined detection of cysteines and lysines.¹³

Taking all of this together, we report on an unbiased, universal workflow to characterise amino acid selectivity of electrophilic compounds in the whole proteome. These studies have enabled the profiling of established, tailored as well as new probes for residue-specific proteomics, which now allows monitoring of a total of nine different amino acids and the *N*-terminus. We are convinced that our workflow and this probe selection will be instrumental to identifying selectively reactive groups for the design of covalent inhibitors targeting diverse amino acids. In this way, they will advance the development of TCIs for the many protein binding sites that lack a suitable cysteine residue for engagement with the currently dominant acrylamide-based chemistry.

Data availability

Mass spectrometric data for all proteomic analyses have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository⁶⁷ with the dataset identifier PXD024454. Raw data for Fig. 2-6 is available in Supplementary Tables 1-4. All other data is available in the main text or the Supplementary Information.

Code Availability

Code for the MSFragger-based FragPipe analyses is available at <https://fragpipe.nesvilab.org/>.

Acknowledgements

S.M.H. and P.R.A.Z. acknowledge funding by the Fonds der Chemischen Industrie through a Liebig Fellowship and a Ph.D. fellowship and by the TUM Junior Fellow Fund. A.I.N. acknowledges financial support from the US National Institutes of Health (R01-GM-094231 and U24-CA210967). F.D.T. thanks the Novartis Institutes for BioMedical Research and the Novartis-Berkeley Center for Proteomics and Chemistry Technologies for supporting this work. P.Z.M. thanks the American Cancer Society for a post-doctoral fellowship (PF-18-132-01-CDD). M.Z. gratefully acknowledges funding by the Studienstiftung des Deutschen Volkes through a Ph.D. fellowship. K.L. is grateful for support by the collaborative research centre SFB1035 (German Research Foundation DFG, Sonderforschungsbereich 1035, Projektnummer 201302640, project B10). C.J.C. acknowledges funding by the NIH (ES 28096 and GM139245). We gratefully acknowledge Stephan A. Sieber and his group for their generous support. We thank the group of Jérôme Waser (EPFL) for providing **EBX1-alkyne**, Ku-Lung Hsu and Jeffrey W. Brulet (University of Virginia) for providing **SuTex1-alkyne** and **SuTex2-alkyne** and the group of Phil S. Baran (Scripps Research) for providing **PSI-alkyne**. We thank Simon Kollmannsberger and Maximilian Iglhaut (Technical University of Munich) for assisting with the synthesis of probes as well as Katja Bäuml and Mona Wolff (Technical University of Munich) for technical assistance. We thank Daniel Geiszler, Felipe da Veiga Leprevost and Dmitry Avtonomov for useful discussions and technical assistance regarding new developments in FragPipe.

Author Contributions

P.R.A.Z. and S.M.H. designed the research, planned experiments, performed and analysed proteomics experiments. F.Y. and A.I.N. designed, developed and benchmarked data analysis software. P.R.A.Z. synthesised the majority of probes. P.Z.M., L.L., M.Z., K.K., D.M., P.R., T.E.M., and M.C. contributed the synthesis of individual probes. K.L., C.J.C. and F.D.T. designed individual probes. P.R.A.Z. and S.M.H. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper in the online version.

Correspondence and requests for materials should be addressed to S.M.H.

References

- Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug. Discov.* **10**, 307-317 (2011).
- Backus, K. M. *et al.* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570-574 (2016).
- Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790-795 (2010).
- Zanon, P. R. A., Lewald, L. & Hacker, S. M. Isotopically Labeled Desthiobiotin Azide (isoDTB) Tags Enable Global Profiling of the Bacterial Cysteinome. *Angew. Chem. Int. Ed.*, 2829-2836 (2020).
- Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596-2599 (2002).
- Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513-520 (2017).
- Yu, F. *et al.* Identification of modified peptides using localization-aware open search. *Nat. Commun.* **11**, 4065 (2020).
- Yu, F. *et al.* Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger and IonQuant. *Mol. Cell. Proteom.* **19**, 1575-1585 (2020).
- Sutanto, F., Konstantinidou, M. & Dömling, A. Covalent inhibitors: a rational approach to drug discovery. *RSC Med. Chem.* **11**, 876-884 (2020).
- Bach, K., Beerkens, B. L. H., Zanon, P. R. A. & Hacker, S. M. Light-Activatable, 2,5-Disubstituted Tetrazoles for the Proteome-Wide Profiling of Aspartates and Glutamates in Living Bacteria. *ACS Cent. Sci.* **6**, 546-554 (2020).
- Gehring, M. & Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **62**, 5673-5724 (2019).
- Parker, C. G. & Pratt, M. R. Click Chemistry in Proteomic Investigations. *Cell* **180**, 605-632 (2020).
- Cao, J. *et al.* Multiplexed CuAAC Suzuki-Miyaura Labeling for Tandem Activity-Based Chemoproteomic Profiling. *Anal. Chem.* **93**, 2610-2618 (2021).
- Lee, K. M., Le, P., Sieber, S. A. & Hacker, S. M. Degrasyn exhibits antibiotic activity against multi-resistant *Staphylococcus aureus* by modifying several essential cysteines. *Chem. Commun.* **56**, 2929-2932 (2020).
- Senkane, K. *et al.* The Proteome-Wide Potential for Reversible Covalency at Cysteine. *Angew. Chem. Int. Ed.* **58**, 11385-11389 (2019).
- Motiwala, H. F., Kuo, Y.-H., Stinger, B. L., Palfey, B. A. & Martin, B. R. Tunable Heteroaromatic Sulfones Enhance in-Cell Cysteine Profiling. *J. Am. Chem. Soc.* **142**, 1801-1810 (2020).
- Abegg, D. *et al.* Proteome-Wide Profiling of Targets of Cysteine reactive Small Molecules by Using Ethynyl Benzodioxolone Reagents. *Angew. Chem. Int. Ed.* **54**, 10852-10857 (2015).
- Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181-1190 (2017).
- Ward, C. C., Kleinman, J. I. & Nomura, D. K. NHS-Esters As Versatile Reactivity-Based Probes for Mapping Proteome-Wide Ligandable Hotspots. *ACS Chem. Biol.* **12**, 1478-1483 (2017).
- Ma, N. *et al.* 2H-Azirine-Based Reagents for Chemoselective Bioconjugation at Carboxyl Residues Inside Live Cells. *J. Am. Chem. Soc.* **142**, 6051-6059 (2020).
- Lin, S. *et al.* Redox-based reagents for chemoselective methionine bioconjugation. *Science* **355**, 597-602 (2017).
- Taylor, M. T., Nelson, J. E., Suero, M. G. & Gaunt, M. J. A protein functionalization platform based on selective reactions at methionine residues. *Nature* **562**, 563-568 (2018).
- Hahn, H. S. *et al.* Global targeting of functional tyrosines using sulfur-triazole exchange chemistry. *Nat. Chem. Biol.* **16**, 150-159 (2019).
- Brulet, J. W., Borne, A. L., Yuan, K., Libby, A. H. & Hsu, K.-L. Liganding Functional Tyrosine Sites on Proteins Using Sulfur-Triazole Exchange Chemistry. *J. Am. Chem. Soc.* **142**, 8270-8280 (2020).
- Horsburgh, M. J. *et al.* sigmaB modulates virulence determinant expression and stress resistance: characterization of a functional rsbU strain derived from *Staphylococcus aureus* 8325-4. *J. Bacteriol.* **184**, 5457-5467 (2002).
- Benns, H. J., Wincott, C. J., Tate, E. W. & Child, M. A. Activity- and reactivity-based proteomics: Recent technological advances and applications in drug discovery. *Curr. Opin. Chem. Biol.* **60**, 20-29 (2021).
- Chang, H.-Y. *et al.* Crystal-C: A Computational Tool for Refinement of Open Search Results. *J. Proteome Res.* **19**, 2511-2515 (2020).
- Geiszler, D. J. *et al.* PTM-Shepherd: analysis and summarization of post-translational and chemical modifications from open search results. *Mol. Cell. Proteom.* (2020).
- Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383-5392 (2002).
- Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646-4658 (2003).
- da Veiga Leprevost, F. *et al.* Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* **17**, 869-870 (2020).
- Lenčo, J., Khalikova, M. A. & Švec, F. Dissolving Peptides in 0.1% Formic Acid Brings Risk of Artificial Formylation. *J. Proteome Res.* **19**, 993-999 (2020).
- Teo, G. C., Polasky, D. A., Yu, F. & Nesvizhskii, A. I. Fast Deisotoping Algorithm and Its Implementation in the MSFragger Search Engine. *J. Proteome Res.* **20**, 498-505 (2021).
- Fu, L. *et al.* A quantitative thiol reactivity profiling platform to analyze redox and electrophile reactive cysteine proteomes. *Nat. Protoc.* **15**, 2891-2919 (2020).
- Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).
- Chi, H. *et al.* Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat. Biotechnol.* **36**, 1059-1061 (2018).
- Abo, M. & Weerapana, E. A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells. *J. Am. Chem. Soc.* **137**, 7087-7090 (2015).
- Embaby, A. M., Schoffelen, S., Kofoed, C., Meldal, M. & Diness, F. Rational Tuning of Fluorobenzene Probes for Cysteine-Selective Protein Modification. *Angew. Chem. Int. Ed.* **57**, 8022-8026 (2018).
- Frei, R. *et al.* Fast and Highly Chemoselective Alkynylation of Thiols with Hypervalent Iodine Reagents Enabled through a Low Energy Barrier Concerted Mechanism. *J. Am. Chem. Soc.* **136**, 16563-16573 (2014).
- Tessier, R. *et al.* "Doubly Orthogonal" Labeling of Peptides and Proteins. *Chem* **5**, 2243-2263 (2019).
- Tessier, R. *et al.* Ethynylation of Cysteine Residues: From Peptides to Proteins in Vitro and in Living Cells. *Angew. Chem. Int. Ed.* **59**, 10961-10970 (2020).
- O'Shea, J. P. *et al.* pLogo: a probabilistic approach to visualizing sequence motifs. *Nat. Methods* **10**, 1211-1212 (2013).
- Pettinger, J., Jones, K. & Cheeseman, M. D. Lysine-Targeting Covalent Inhibitors. *Angew. Chem. Int. Ed.* **56**, 15200-15209 (2017).
- Hirata, T. *et al.* Synthesis and reactivities of 3-indocyanine-green-acyl-1,3-thiazolidine-2-thione (ICG-ATT) as a new near-infrared fluorescent-labeling reagent. *Bioorg. Med. Chem.* **6**, 2179-2184 (1998).
- Tamura, T. *et al.* Rapid labelling and covalent inhibition of intracellular native proteins using ligand-directed N-acyl-N-alkyl sulfonamide. *Nat. Commun.* **9**, 1870 (2018).
- Ivancová, I., Pohl, R., Hubálek, M. & Hocek, M. Squaramate-Modified Nucleotides and DNA for Specific Cross-Linking with Lysine-Containing Peptides and Proteins. *Angew. Chem. Int. Ed.* **58**, 13345-13348 (2019).
- Deng, J.-R. *et al.* N-Terminal selective modification of peptides and proteins using 2-ethynylbenzaldehydes. *Commun. Chem.* **3**, 67 (2020).
- Guo, A.-D. *et al.* Light-induced primary amines and o-nitrobenzyl alcohols cyclization as a versatile photoclick reaction for modular conjugation. *Nat. Commun.* **11**, 5472 (2020).
- Onoda, A., Inoue, N., Sumiyoshi, E. & Hayashi, T. Triazolecarbaldehyde Reagents for One-Step N-Terminal Protein Modification. *ChemBioChem* **21**, 1274-1278 (2020).
- MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **11**, 326-331 (2015).
- Zengeya, T. T. *et al.* Co-opting a Bioorthogonal Reaction for Oncometabolite Detection. *J. Am. Chem. Soc.* **138**, 15813-15816 (2016).

- 52 Dong, J., Krasnova, L., Finn, M. G. & Sharpless, K. B. Sulfur(VI) Fluoride Exchange (SuFEx): Another Good Reaction for Click Chemistry. *Angew. Chem. Int. Ed.* **53**, 9430-9448 (2014).
- 53 Ban, H., Gavriluyk, J. & Barbas, C. F. Tyrosine bioconjugation through aqueous ene-type reactions: a click-like reaction for tyrosine. *J. Am. Chem. Soc.* **132**, 1523-1525 (2010).
- 54 Ban, H. *et al.* Facile and Stable Linkages through Tyrosine: Bioconjugation Strategies with the Tyrosine-Click Reaction. *Bioconjug. Chem.* **24**, 520-532 (2013).
- 55 Zhang, J., Ma, D., Du, D., Xi, Z. & Yi, L. An efficient reagent for covalent introduction of alkynes into proteins. *Org. Biomol. Chem.* **12**, 9528-9531 (2014).
- 56 Nothling, M. D. *et al.* Bacterial Redox Potential Powers Controlled Radical Polymerization. *J. Am. Chem. Soc.* **143**, 286-293 (2021).
- 57 Naveen, N., Sengupta, S. & Chandrasekaran, S. Metal-Free S-Arylation of Cysteine Using Arenediazonium Salts. *J. Org. Chem.* **83**, 3562-3569 (2018).
- 58 Christian, A. H. *et al.* A Physical Organic Approach to Tuning Reagents for Selective and Stable Methionine Bioconjugation. *J. Am. Chem. Soc.* **141**, 12657-12662 (2019).
- 59 Tower, S. J., Hetcher, W. J., Myers, T. E., Kuehl, N. J. & Taylor, M. T. Selective Modification of Tryptophan Residues in Peptides and Proteins Using a Biomimetic Electron Transfer Process. *J. Am. Chem. Soc.* **142**, 9112-9118 (2020).
- 60 Jia, S., He, D. & Chang, C. J. Bioinspired Thiophosphorodichloridate Reagents for Chemoselective Histidine Bioconjugation. *J. Am. Chem. Soc.* **141**, 7294-7301 (2019).
- 61 Dovgan, I. *et al.* Arginine-selective bioconjugation with 4-azidophenyl glyoxal: application to the single and dual functionalisation of native antibodies. *Org. Biomol. Chem.* **16**, 1305-1311 (2018).
- 62 Jones, A. X. *et al.* Improving mass spectrometry analysis of protein structures with arginine-selective chemical cross-linkers. *Nat. Commun.* **10**, 3911 (2019).
- 63 Thompson, D. A., Ng, R. & Dawson, P. E. Arginine selective reagents for ligation to peptides and proteins. *J. Pept. Sci.* **22**, 311-319 (2016).
- 64 Chaudhuri, R. R. *et al.* Comprehensive identification of essential *Staphylococcus aureus* genes using Transposon-Mediated Differential Hybridisation (TMDH). *BMC Genomics* **10**, 291 (2009).
- 65 Vantourout, J. C. *et al.* Serine-Selective Bioconjugation. *J. Am. Chem. Soc.* **142**, 17236-17242 (2020).
- 66 Knouse, K. W. *et al.* Unlocking P(V): Reagents for chiral phosphorothioate synthesis. *Science* **361**, 1234-1238 (2018).
- 67 Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442-d450 (2019).