

PepFoot: a software package for semi-automated processing of protein footprinting data

Jeddiah Bellamy-Carter^{1*} and Neil J. Oldham^{1*}

¹*School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD, UK*

*E-mail: jeddiah.bellamy-carter@nottingham.ac.uk, neil.oldham@nottingham.ac.uk

ABSTRACT

Covalent footprinting of proteins using reactive intermediates such as radicals and carbenes is emerging as a valuable tool for mapping surface accessibility, and hence binding sites of proteins. The approach generates a significant amount of mass spectrometry (MS) data, which can be time-consuming to process manually. PepFoot, a software package that allows semi-automated processing of MS data from footprinting experiments, is described. By using the open source .mz5 file format, it is able to accept data from all the major instrument manufacturers. Following manual user interrogation of one data file within a user-friendly GUI, the software then automates determination of the degree of fractional modification (f_m) with the footprinting agent across a batch of experimental data. This greatly increases efficiency and throughput compared to manual analysis of each file, and provides initial scrutiny and confidence compared to fully-automated analysis. Histogram plots of f_m for each peptide from the footprinted protein may be displayed within PepFoot and mapped onto an imported protein structure to reveal differential labeling patterns and hence binding sites. The software has been tested on data from carbene and hydroxyl radical labeling experiments to demonstrate its broad utility. PepFoot is released under the LGPL version 3 license, and is available for Windows, MacOS and Linux systems at github.com/jbellamycarter/pepfoot.

KEYWORDS protein footprinting, covalent modification, carbene, hydroxyl radical, mass spectrometry, protein interactions, workflow, visualization, software, python

INTRODUCTION

Protein footprinting techniques are emerging as fast and reliable methods for investigating protein-protein or protein-small molecule interactions, which are central to biochemical processes. In order to characterize these interactions, an array of high-resolution techniques are frequently used, such as X-ray crystallography, nuclear

magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM), but these methods are time-consuming and require relatively large quantities of sample. Mass spectrometry (MS) based techniques in contrast are both rapid and sensitive,¹ introducing a defined mass shift which is then used to quantify surface accessibility. One of the earliest and most widely used MS techniques is hydrogen-deuterium exchange (HDX), which probes solvent exposure of a protein's surface through the uptake of deuterium.² HDX has been used to investigate both binding interactions and conformational dynamics. While the method provides promising results it suffers from back-exchange to hydrogen in the common protic solvents used in liquid chromatography-mass spectrometry (LC-MS)³ and H/D scrambling during collisional activation for MS/MS, which can make sub-peptide level analysis challenging.⁴

Hydroxyl radical footprinting (HRF) is a very promising method that relies on irreversible oxidation of residue side-chains by fast-reacting $\cdot\text{OH}$. A crucial advantage of this approach over HDX is the permanent nature of the oxidative modification, which remains throughout subsequent processing steps. The original technique, developed by Chance,^{5,6} used synchrotron radiation to ionize water over millisecond to minute timescales. A variant developed by Gross,^{7,8} fast photochemical oxidation of proteins (FPOP) uses ultraviolet laser irradiation to generate the hydroxyl radical from hydrogen peroxide with microsecond exposure times allowing native state conformations to be sampled rapidly.⁹ The reactivity of $\cdot\text{OH}$ with amino acid side chains ranges over three orders of magnitude,⁶ leaving several amino acids (Gly, Ala, Asp, Asn, Gln, and Glu) unreactive under optimal conditions for the sulfur containing and aromatic side chains. FPOP has been used to sample conformational unfolding states of Im7 (folded, partially unfolded and globally unfolded) and reveal differences in side-chain accessibility.¹⁰ Recently a similar method producing trifluoromethyl radical ($\cdot\text{CF}_3$)¹¹ has been reported to give excellent coverage of residues.

Carbene footprinting is an alternative to the hydroxyl radical approach. It utilizes a highly reactive carbene generated *in situ*, typically from a diazirine reagent irradiated with a near-ultraviolet laser.¹²⁻¹⁵ The rate of this insertion is on the order of nanoseconds,¹⁶ much faster than secondary structure folding events, making carbenes an attractive reactive intermediate for protein study. While initial studies used diazirine gas to generate methylene,¹² more recent studies have employed trifluoromethylaryldiazirines due to their stability, solubility and efficacy.^{14,17-19}

As footprinting methods introduce an irreversible covalent modification, unlike HDX, labeling and hence structural information is retained well throughout proteolytic and chromatographic steps. The extent of peptide modification is expressed as fractional modification (f_m). Differential experiments compare the fractional

modification of peptides between treatments to identify sites of interaction on a protein. Sub-peptide-level modification can then be determined with MS/MS experiments to pinpoint key residues.

While the data for these methods are rich, they are typically complex, making data processing and interpretation a major bottleneck to high-throughput investigation of protein interactions. Some software tools have been developed to aid data processing. ProtMapMS²⁰, for example, is a proprietary compilation of automated algorithms for identification, dose-response and rates of oxidative labeling experiments written by the Chance group, and a commercial software Byologic (www.proteinmetric.com) can identify protein modifications and quantify oxidation from these experiments. However, these tools are not designed to provide immediate comparative visualization and mapping of f_m for convenient interpretation of footprinting experiments. Furthermore these tools require MS/MS data for all putative peptides for assignment; covalent labels, arylidiazirines in particular, can exhibit wide retention windows across isomers and may be missed by non-targeted MS/MS.

Here we present PepFoot, an open source cross-platform, semi-automated software tool for processing footprinting data. The software can accept data files from all major vendors by conversion to an open source file format and allows tailored identification of covalent modifications followed by batch processing and visualization to facilitate rapid and reliable interpretation of footprinting experiments. The output of the software can be shared with other researchers improving scrutiny and collaborative efforts. A semi-automated approach was chosen to provide the scrutiny of initial manual verification with the reliability and speed of computational methods.

METHODS

Input Data Formats Previously published protein footprinting data^{18,19} were used to access the efficacy of PepFoot for semi-automated processing. A variety of open source data formats exist for storing MS data, the most popular of which is .mzML,²¹ a standardized XML framework for which there are many libraries for data access. While use of this format is wide-spread, compared to raw vendor formats its read-write speeds are low and file sizes are large; a variant of this format based on the HDF5 (Hierarchical Data Format Version 5) dubbed .mz5 overcomes these shortcomings.^{22,23} Input data were converted from raw vendor formats to the .mz5 file format through the msconvert program packaged with ProteoWizard.²⁴ High-level access to this program is available through PepFoot itself on machines running Windows or Wine emulator.

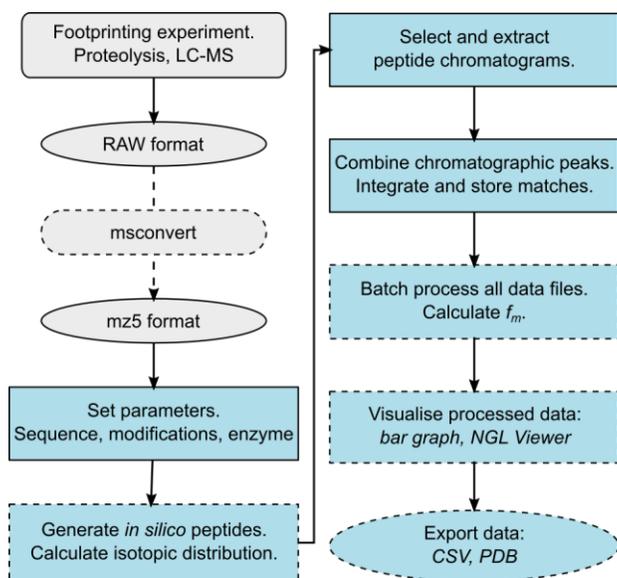


Figure 1. Schematic representation of data processing workflow with PepFoot. Processes performed outside and within PepFoot are colored gray and blue respectively. Processes automated by software have dashed borders.

Data processing. PepFoot was written in Python 3.6 with a graphical user interface built using Qt 5.11. In addition to standard packages, PepFoot handles data through the h5py, numpy and scipy packages; processes theoretical peptides with the pyteomics^{25,26} package; and visualizes data with the PyQt5, matplotlib²⁷ python packages and NGL Viewer^{28,29} for 3-dimensional visualization. The .mz5 files are parsed using a custom python class. In the .mz5 format, m/z data are stored in a delta mass representation for storage efficiency which must be converted to m/z arrays prior to analysis, for convenience PepFoot allows optional parsing and storage of these m/z arrays as a ragged list of numpy arrays rather than ad hoc parsing. This can greatly increase user interface response at the cost of RAM.

The main window provides a panel with three tabs: for inspecting peptides (Peptide Level), analyzing fractional modification (Analysis) and mapping onto a pdb structure (NGL viewer), respectively. Within the Peptide Level tab, the right-hand side parameter bar consists of a list of data files to be processed, an input for single-letter amino acid protein sequences with modX²⁵ support, a selection of fixed modifications, the labeling probe, and standard in silico digestion settings. These settings are used to generate a list of theoretical peptides to search for labeled and unlabeled variants. For each of these theoretical peptides a modX sequence is generated for all fixed modifications for both unlabeled and labeled forms, and the elemental composition of these used to generate theoretical isotopic distributions using the binomial method, the most abundant of which are used for further analysis. In the Peptide Level tab, shown in Figure 2, peptides may be selected and

extracted ion chromatograms are generated for the theoretical unlabeled and singly labeled m/z peptide ions simultaneously. The user may combine chromatographic peaks to return mass spectra, to which the theoretical isotopic distribution is fitted with an m/z error to aid verification of positive matches. Suitable matches can be integrated under the chromatographic peaks and the assigned m/z and retention time ranges for that peptides are stored, these parameters are called upon by the batch processor to automate the analysis of remaining data files (approximately 10 s per file) provided the chromatography is reproducible. Combined mass spectra are simply summed, in rare case, possibly due to FT-ICR instrumental instability, we have observed peak splitting (see Figure S1); in Xcalibur, spectra are summed with a tolerance, which results in ‘smooth’ peaks.

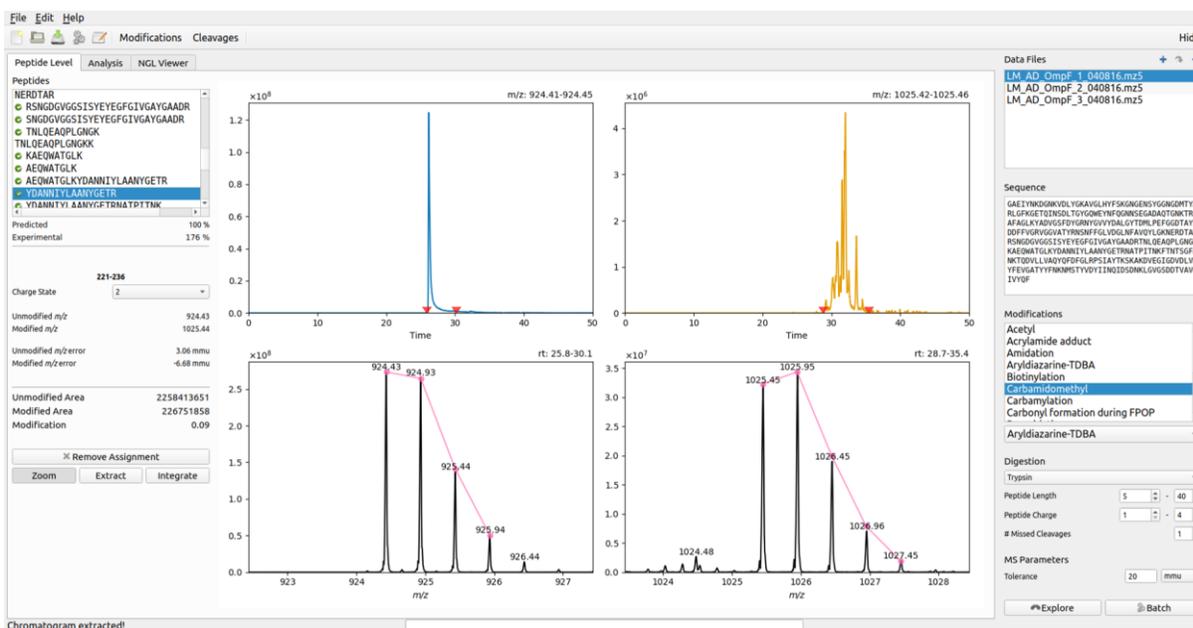


Figure 2. Graphical User Interface for inspection of peptide-level footprinting. A sidebar (*right*) for user defined parameters to generate theoretical peptides. The Peptide Level tab allows selection of potential peptides for a given charge state (*top-left*) followed by generation of extracted ion chromatograms for unlabeled (blue, *center-left*) and labeled (orange, *center-right*) ions in the central panel. Extraction across chromatographic peaks yields mass spectra fitted with the theoretical isotopic distribution (red markers and lines), with m/z errors to assess confidence. Integration of EIC peaks produces peak area and f_m for the selected ion (*lower left*). The EIC of the modified peptide shows multiple peaks due to labeling at different sites within the peptide.

Data visualization and interpretation. Fractional modification (f_m) is determined as the ratio of chromatographic peak areas for a singly labeled peptide to the sum of labeled and singly labeled peptides and displayed in the form of an interactive bar graph in the ‘Analysis’ tab, as shown in Figure 3a. For differential

experiments the data files can be interactively grouped by treatment and the difference in f_m between with and without binding partner treatments visualized with significance determined by Students' t-test (significant if p -value is below user-set threshold). Labeling can be mapped onto a 3-dimensional structure of the protein in a .pdb file by assigning b-factor for labeling below or above a user set threshold respectively; for single treatment experiments the extent of labeling is mapped, while for differential experiments, with and without a binding partner, the significance as defined above is mapped. The modified .pdb file is then loaded and visualized using embedded NGL Viewer²⁸ providing interactive visualization as shown in Figure 3b. User provided .pdb files are parsed by a custom python class that interfaces with PepFoot output to assign b-factor, this is only applied to chains that exactly match the user defined amino acid sequence, non-matches are assigned a null b-factor. The modified .pdb files may be processed in all common molecular visualization software and colored by b-factor. B-factor is assigned a value of -2, 0, 1 or -1 for residues that are not detected, insignificantly labeled, significantly labeled and significantly exposed (for differential experiments only) respectively. Optional scaling by extent of modification or change between these values is available in the 'NGL Viewer' tab.

The parameters and extracted data are stored in a JSON (JavaScript Object Notation) file with the .pfoot extension. This human-readable file may be shared for reuse with or parameters imported into PepFoot for analogous experiments, the results may also be exported to .csv for convenience.

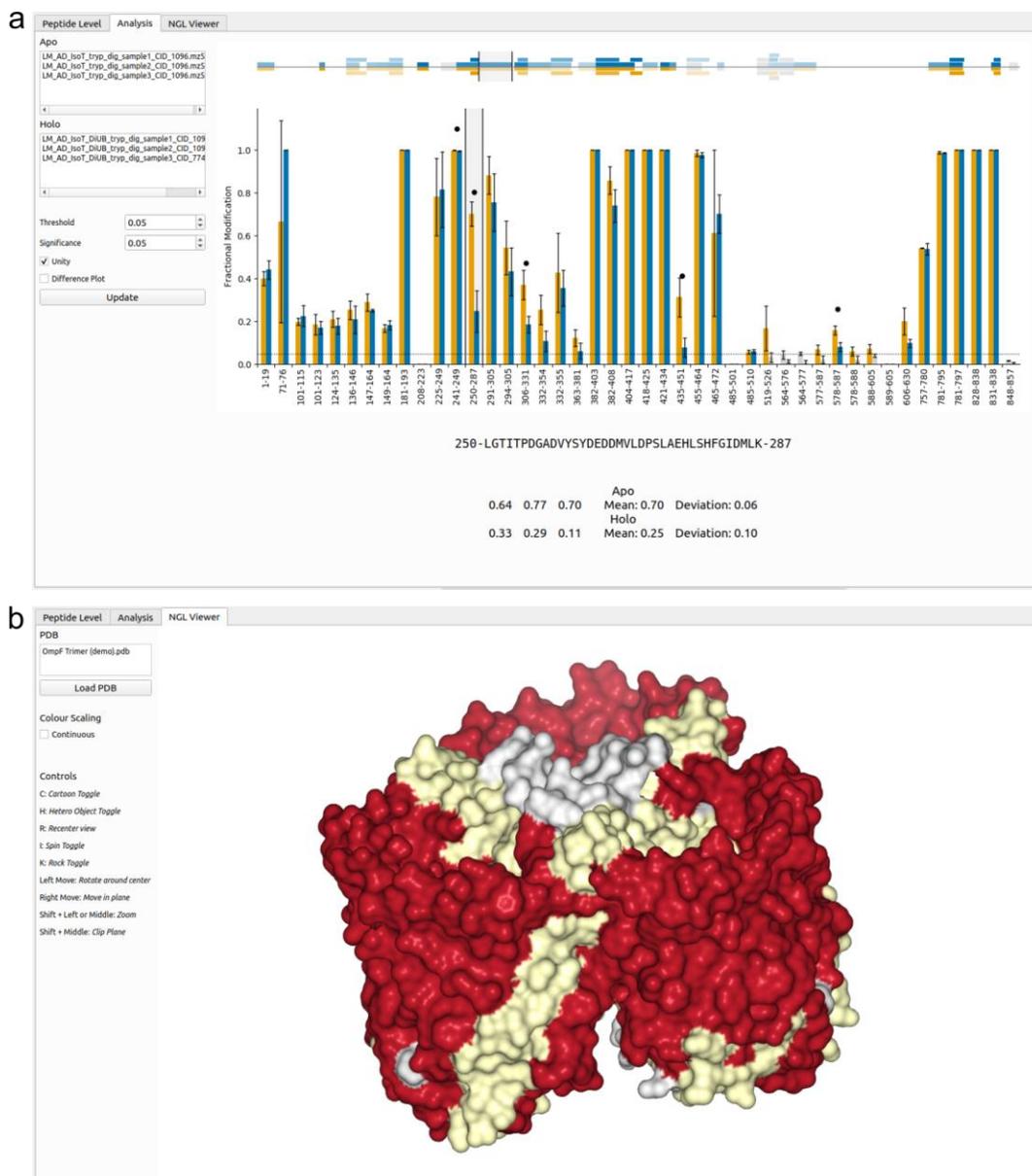


Figure 3. Graphical User Interface tabs for visualization of PepFoot output. **(a)** Analysis tab with interactive bar graph for fractional modification data. Bars can be selected to display f_m data for the selected peptide and shows location on a 2D map of the protein. Panels are available to group data files by treatment (*top-left*) and update bar graph and NGL Viewer. The view can be set to display a difference plot for grouped data. **(b)** Embedded NGL Viewer with selection for local .pdb file and described controls for interaction. User selected PDB files are parsed by PepFoot using a custom class and exact sequence matches are assigned a b-factor reflecting masking. The parsed .pdb file is then overwritten on disk and loaded into NGL Viewer with a b-factor color scheme as follows: non-detected peptides (-2, light-gray), insignificant masking (0, wheat), significant labeling (1, red) and significant exposure (-1, blue). The ‘Continuous’ checkbox enables color scaling.

RESULTS AND DISCUSSION

Analysis and interpretation of protein footprinting data is typically time consuming and the primary bottleneck to high throughput experiments. Analysis of a single-protein footprinting experiment can take days as each file must be processed manually. Protein footprinting provides a valuable tool for interrogating protein interactions, but improving the efficiency and reproducibility of analysis is essential for the field to grow.

Revealing membrane protein interfaces. Footprinting of OmpF in 1% octyl-glucoside with the aryldiazirine-type carbene probe sodium 4-(3-trifluoromethyl)-3H-diazirin-3-yl)benzoate reveals membrane-binding and protein-trimer interfaces.¹⁹ Data for membrane binding interfaces of the *E. coli* membrane protein OmpF were downloaded from PRIDE data set [PXD007207](#), converted (Thermo RAW → mz5, remove zeros) and processed with PepFoot (aryldiazirine-TDBA, peptide length 5–40, charge range 1–4, missed cleavages 1 and mass tolerance 20 mmu). Tryptic peptides yielding 91% coverage of the protein sequence were identified with 16/27 peptides having significant labeling by the diazirine probe. This shows an increase in identification compared with the original study,¹⁹ which found 86% coverage with 12/20 peptides showing significant labeling. Of the seven newly identified peptides, four peptides (18-26, 18-43, 221-244 and 237-254) accounted for previously unidentified regions of OmpF (18-26 and 237-244). These peptides were manually verified (see Table S1). All previously identified peptides except 134-141 showed remarkably similar f_m and deviations, see Figure 4a, S2 and Table S1. Peptide 134-141 shows consistent f_m when processed through PepFoot but showed an anomalously low f_m from one data file in the published study, upon further inspection this appears to be the result of manual error in recording the peak area. Manual recalculation of the peak areas for this peptide yielded improved deviation and close match to PepFoot output.

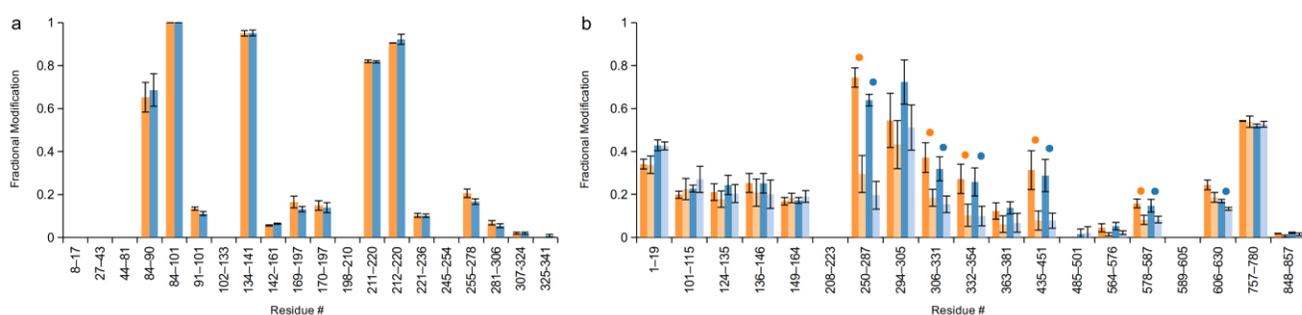


Figure 4. Comparative bar charts of peptide fractional modification determined by PepFoot and from the published data for OmpF¹⁹ (a) and USP5¹⁸ (b). PepFoot data are shown in orange and published manually-processed data in blue. For USP5, data are grouped by presence or absence of diubiquitin (dark shades without diubiquitin, light shades with). Results show highly similar f_m between PepFoot and manually processed data. Error bars indicate \pm s.d. ($n=3$) and significant masking ($p < 0.05$) is displayed with a dot.

Identifying protein-protein binding sites. Differential footprinting of deubiquitinating enzyme (DUB) ubiquitin specific protease 5 (USP5) alone or in complex with diubiquitin revealed interaction of the diubiquitin with the ZnF-UBP and catalytic domains, and showed a biologically significant conformation not accessible through X-ray crystallography.¹⁸

Data for diubiquitin binding of USP5 (C335A) were downloaded from PRIDE data set [PXD004971](#), converted (Thermo RAW → mz5, remove zeros) and processed with PepFoot (aryldiazirine-TDBA, carbamidomethyl fixed modification, peptide length 5–40, charge range 1–4, missed cleavages 1 and mass tolerance 20 mmu). The data were then grouped by treatment with diubiquitin.

PepFoot analysis of USP5 data yielded an additional 25 putative peptides to those from the published study accounting for an additional 20% protein coverage. These peptides were manually verified (see Table S2). A majority of these were found to be highly labeled and may have been previously overlooked for this reason. Of the peptides common to both data-sets there was little difference between the output from PepFoot and the reported values (see Figure 4b, S3 and Table S2,3), with the one exception of peptide 606-630, which is found to be statistically insignificantly different through PepFoot. Even by manual analysis, the difference between the two treatments was small. Thus, grouping feature in PepFoot allowed for reliable and rapid assignment of significant masking events.

Processing FPOP data. To test the performance of PepFoot with hydroxyl radical footprinting data, we were kindly provided with a data-set for myoglobin from an FPOP experiment by the Ashcroft group. The data were converted (MassLynx RAW → mz5, remove zeros) and processed with PepFoot (Oxidation, peptide length 5–40, charge range 1–4, missed cleavages 1 and mass tolerance 20 mmu). As PepFoot can only handle a single variable probe mass shift per analysis, the data were processed independently with three variants of oxidation (+16, +32 and +14 Da). An appreciable amount of labeling was found for +16 and +32 oxidation with no significant labeling for +14 Da (see Figure S4,5), as would be predicted by residue reactivity. As +16 oxidation is the predominant modification it is appropriate to use PepFoot for characterization of these areas for it alone, with additional investigation for peptides prone to +32 oxidation. PepFoot was able to handle the data efficiently, demonstrating its broad utility for analyzing data from covalent labeling experiments.

The FPOP experiments also provided the opportunity to test data acquired on a time-of-flight mass spectrometer. Due to the higher noise-to-signal acquired from ToF instruments, the file sizes for LC-MS runs are large and are cumbersome for user interaction with the data. We have found that applying an absolute

threshold of 500 counts per scan to filter the data during conversion to .mz5 greatly improves this with little cost to data integrity (see Figure S6,7). This would typically be performed for initial inspection of the data before batch processing with the unfiltered data.

CONCLUSIONS

The spread of covalent footprinting techniques for studying protein interactions is hampered by complex data, and the laborious manual processing that comes with it. To exploit the potential of the field fully, effective processing tools are required. Herein we have described semi-automated software, with an interactive user interface, to make handling footprinting data accessible to non-expert users, and to open the way for higher throughput analysis of the complex biomolecular interactions governing biology. The software provides a generic platform for investigating all covalent labeling techniques with consistent and shareable output. The current iteration of PepFoot is focused towards peptide-level analysis and simple interpretation of footprinting data, developments to allow residue level interrogation via MS/MS and filtering potential interaction models are underway.

ASSOCIATED CONTENT

SUPPORTING INFORMATION

The Supporting Information is available free of charge on the [ACS Publications website](#) at DOI:

- Description of peak splitting, comparison of OmpF and USP5 carbene footprinting data-sets and FPOP results (PDF).
 - **Figure S1.** Mass Spectrum combine between Xcalibur and PepFoot
 - **Figure S2.** Additional peptides for OmpF mapped to structure
 - **Figure S3.** Peptides detected for USP5 mapped to structure
 - **Figure S4.** Oxidation of holo-myoglobin via FPOP
 - **Figure S5.** Oxidation of holo-myoglobin via FPOP on structure
 - **Figure S6.** Effect of absolute threshold on f_m for Myoglobin
 - **Figure S7.** Effect of absolute threshold on spectra quality
 - **Table S1.** Comparison of f_m for OmpF
 - **Table S2.** Comparison of f_m for USP5 in absence of diubiquitin
 - **Table S3.** Comparison of f_m for USP5 in presence of diubiquitin
- Video showing typical usage of PepFoot (AVI).
- Example .pfoot files and accompanying .pdb files for OmpF and USP5 (ZIP)

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

This work was supported by a Biotechnology and Biological Science Research Council Doctoral Training Programme studentship (project number 1644216 related to BB/M008770/1) for JBC. The authors would like to thank the Ashcroft group of the University of Leeds for providing FPOP data.

REFERENCES

- (1) Konermann, L.; Vahidi, S.; Sowole, M. A. Mass Spectrometry Methods for Studying Structure and Dynamics of Biological Macromolecules. *Anal. Chem.* **2014**, *86*, 213–232. <https://doi.org/10.1021/ac4039306>.
- (2) Konermann, L.; Pan, J.; Liu, Y. H. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40*, 1224–1234. <https://doi.org/10.1039/c0cs00113a>.
- (3) Sheff, J. G.; Rey, M.; Schriemer, D. C. Peptide-Column Interactions and Their Influence on Back Exchange Rates in Hydrogen/Deuterium Exchange-MS. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1006–1015. <https://doi.org/10.1007/s13361-013-0639-4>.
- (4) Ferguson, P. L.; Pan, J.; Wilson, D. J.; Dempsey, B.; Lajoie, G.; Shilton, B.; Konermann, L. Hydrogen/Deuterium Scrambling during Quadrupole Time-of-Flight MS/MS Analysis of a Zinc-Binding Protein Domain. *Anal. Chem.* **2007**, *79*, 153–160. <https://doi.org/10.1021/ac061261f>.
- (5) Maleknia, S. D.; Brenowitz, M.; Chance, M. R. Millisecond Radiolytic Modification of Peptides by Synchrotron X-Rays Identified by Mass Spectrometry. *Anal. Chem.* **1999**, *71*, 3965–3973. <https://doi.org/10.1021/ac990500e>.
- (6) Xu, G.; Chance, M. R. Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* **2007**, *107*, 3514–3543. <https://doi.org/10.1021/cr0682047>.
- (7) Hambly, D. M.; Gross, M. L. Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 2057–2063. <https://doi.org/10.1016/j.jasms.2005.09.008>.
- (8) Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L. Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal. Chem.* **2009**, *81*, 6563–6571. <https://doi.org/10.1021/ac901054w>.
- (9) Chen, J.; Rempel, D. L.; Gau, B. C.; Gross, M. L. Fast Photochemical Oxidation of Proteins and Mass Spectrometry Follow Submillisecond Protein Folding at the Amino-Acid Level. *J. Am. Chem. Soc.* **2012**, *134*, 18724–18731. <https://doi.org/10.1021/ja307606f>.
- (10) Calabrese, A. N.; Ault, J. R.; Radford, S. E.; Ashcroft, A. E. Using Hydroxyl Radical Footprinting to Explore the Free Energy Landscape of Protein Folding. *Methods* **2015**, *89*, 38–44. <https://doi.org/10.1016/j.ymeth.2015.02.018>.
- (11) Cheng, M.; Zhang, B.; Cui, W.; Gross, M. L. Laser-Initiated Radical Trifluoromethylation of Peptides and Proteins: Application to Mass-Spectrometry-Based Protein Footprinting. *Angew. Chemie - Int. Ed.* **2017**, *56*, 14007–14010. <https://doi.org/10.1002/anie.201706697>.
- (12) Richards, F. M.; Lamed, R.; Wynn, R.; Patel, D.; Olack, G. Methylene as a Possible Universal Footprinting Reagent That Will Include Hydrophobic Surface Areas: Overview and Feasibility: Properties of Diazirine as a Precursor. *Protein Sci.* **2000**, *9*, 2506–2517. <https://doi.org/10.1110/ps.9.12.2506>.

- (13) Gómez, G. E.; Monti, J. L. E.; Mundo, M. R.; Delfino, J. M. Solvent Mimicry with Methylene Carbene to Probe Protein Topography. *Anal. Chem.* **2015**, *87*, 10080–10087. <https://doi.org/10.1021/acs.analchem.5b02724>.
- (14) Jumper, C. C.; Schriemer, D. C. Mass Spectrometry of Laser-Initiated Carbene Reactions for Protein Topographic Analysis. *Anal. Chem.* **2011**, *83*, 2913–2920. <https://doi.org/10.1021/ac102655f>.
- (15) Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. High-Resolution Mapping of Carbene-Based Protein Footprints. *Anal. Chem.* **2012**, *84*, 4411–4418. <https://doi.org/10.1021/ac300120z>.
- (16) Platz, M.; Modarelli, D.; Morgan, S.; White, W.; Mullins, M.; Celebi, S.; Toscano, J. Lifetimes of Alkyl and Dialkyl Carbenes in Solution. *Prog. React. Kinet.* **1994**, *19*, 93–137.
- (17) Kotzyba-Hibert, F.; Kapfer, I.; Goeldner, M. Recent Trends in Photoaffinity Labeling. *Angew. Chemie Int. Ed. English* **1995**, *34*, 1296–1312. <https://doi.org/10.1002/anie.199512961>.
- (18) Manzi, L.; Barrow, A. S.; Scott, D.; Layfield, R.; Wright, T. G.; Moses, J. E.; Oldham, N. J. Carbene Footprinting Accurately Maps Binding Sites in Protein-Ligand and Protein-Protein Interactions. *Nat. Commun.* **2016**, *7*, 1–9. <https://doi.org/10.1038/ncomms13288>.
- (19) Manzi, L.; Barrow, A. S.; Hopper, J. T. S.; Kaminska, R.; Kleanthous, C.; Robinson, C. V.; Moses, J. E.; Oldham, N. J. Carbene Footprinting Reveals Binding Interfaces of a Multimeric Membrane-Spanning Protein. *Angew. Chemie - Int. Ed.* **2017**, *56*, 14873–14877. <https://doi.org/10.1002/anie.201708254>.
- (20) Kaur, P.; Kiselar, J. G.; Chance, M. R. Integrated Algorithms for High-Throughput Examination of Covalently Labeled Biomolecules by Structural Mass Spectrometry. *Anal. Chem.* **2009**, *81*, 8141–8149. <https://doi.org/10.1021/ac9013644>.
- (21) Deutsch, E. MzML: A Single, Unifying Data Format for Mass Spectrometer Output. *Proteomics* **2008**, *8*, 2776–2777. <https://doi.org/10.1002/pmic.200890049>.
- (22) Wilhelm, M.; Kirchner, M.; Steen, J. A. J.; Steen, H. Mz5 : Space- and Time-Efficient Storage of Mass Spectrometry Data Sets. *Mol. Cell. Proteomics* **2012**, *11*, O111.011379. <https://doi.org/10.1074/mcp.O111.011379>.
- (23) Askenazi, M.; Ben Hamidane, H.; Graumann, J. The Arc of Mass Spectrometry Exchange Formats Is Long, but It Bends toward HDF5. *Mass Spectrom. Rev.* **2017**, *36*, 668–673. <https://doi.org/10.1002/mas.21522>.
- (24) Holman, J. D.; Tabb, D. L.; Mallick, P. Employing ProteoWizard to Convert Raw Mass Spectrometry Data. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014; p 13.24.1-13.24.9. <https://doi.org/10.1002/0471250953.bi1324s46>.
- (25) Goloborodko, A. A.; Levitsky, L. I.; Ivanov, M. V.; Gorshkov, M. V. Pyteomics - A Python Framework for Exploratory Data Analysis and Rapid Software Prototyping in Proteomics. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 301–304. <https://doi.org/10.1007/s13361-012-0516-6>.

- (26) Levitsky, L. I.; Klein, J. A.; Ivanov, M. V.; Gorshkov, M. V. Pyteomics 4.0: Five Years of Development of a Python Proteomics Framework. *J. Proteome Res.* **2019**, *18*, 709–714. <https://doi.org/10.1021/acs.jproteome.8b00717>.
- (27) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- (28) Rose, A. S.; Hildebrand, P. W. NGL Viewer: A Web Application for Molecular Visualization. *Nucleic Acids Res.* **2015**, *43*, W576–W579. <https://doi.org/10.1093/nar/gkv402>.
- (29) Rose, A. S.; Bradley, A. R.; Valasatava, Y.; Duarte, J. M.; Prlić, A.; Rose, P. W. NGL Viewer: Web-Based Molecular Graphics for Large Complexes. *Bioinformatics* **2018**, *34*, 3755–3758. <https://doi.org/10.1093/bioinformatics/bty419>.

For Table of Contents Only

