

GigaScience

mspms: A Comprehensive R Package and Graphic Interface for Multiplex Substrate Profiling by Mass Spectrometry Analysis

--Manuscript Draft--

Manuscript Number:			
Full Title:	mspms: A Comprehensive R Package and Graphic Interface for Multiplex Substrate Profiling by Mass Spectrometry Analysis		
Article Type:	Technical Note		
Funding Information:	National Institute of General Medical Sciences (GM007752)	Mr Charlie Bayne Ms Brianna Hurysz	
	National Institute of Allergy and Infectious Diseases (R01AI158612)	Dr. Anthony O'Donoghue	
	National Institute of Allergy and Infectious Diseases (R21AI171824)	Dr. Anthony O'Donoghue	
	Division of Cancer Epidemiology and Genetics, National Cancer Institute (R21CA256460)	Dr. Anthony O'Donoghue	
Abstract:	Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS) is a powerful method for determining the substrate specificity of proteolytic enzymes, knowledge key for developing protease inhibitors, diagnostics, and protease-activated therapeutics. However, the complex datasets generated by MSP-MS pose significant analytical challenges. To address this, we developed mspms, a Bioconductor R package complemented by an intuitive graphical interface. Mspms streamlines MSP-MS data analysis by standardizing workflows for data preparation, processing, statistical analysis, and visualization. Designed for accessibility, it serves both advanced users via the R package and broader audiences through the web interface. We validated mspms by profiling the substrate specificity of four well-characterized cathepsins (A–D), demonstrating its ability to reliably capture expected substrate specificities. As the first publicly available platform for MSP-MS data analysis, mspms delivers comprehensive functionality, transparency, and ease of use, making it a valuable resource for the protease research community. Access to mspms is available through the Bioconductor project at https://bioconductor.org/packages/mspms , and a graphic interface is available at https://gonzalezlab.shinyapps.io/mspms_shiny/ .		
Corresponding Author:	Charlie Bayne University of California San Diego UNITED STATES		
Corresponding Author Secondary Information:			
Corresponding Author's Institution:	University of California San Diego		
Corresponding Author's Secondary Institution:			
First Author:	Charlie Bayne		
First Author Secondary Information:			
Order of Authors:	Charlie Bayne		
	Brianna Hurysz		
	David J. Gonzalez, PhD		
	Anthony O'Donoghue, PhD		
Order of Authors Secondary Information:			
Additional Information:			

Question	Response
Are you submitting this manuscript to a special series or article collection?	No
<p>Experimental design and statistics</p> <p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends.</p> <p>Have you included all the information requested in your manuscript?</p>	Yes
<p>Resources</p> <p>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?</p>	Yes
<p>Availability of data and materials</p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the "Availability of Data and Materials" section of your manuscript.</p>	Yes

<p>Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?</p>	
<p>GigaScience has policies and guidelines in place for the use of generative AI-writing tools such as ChatGPT. If you have used such writing tools to assist with writing the manuscript this must be declared and cited in the text. Authors should not list AI-writing tools and other AI-assisted technologies as an author or co-author and should acknowledge that they are fully responsible for text generated or refined by AI-writing tools.<p></p> <p>A summary of use (particularly in the introduction or among methods) needs to be included at the end of the paper, and the outputs should also be included as a supplementary file hosted in GigaDB or other open repositories. Please read our guidelines for more information. <p></p> <p>By submitting to GigaScience, you are aware of the journal's AI-writing tools policy, and if you have declared use of such tools below, you have acknowledged this where appropriate in your manuscript and have made a summary of use and outputs available. <p></p> <p>AI-assisted writing tools have been used in the preparation of this manuscript?</p>	<p>No</p>

1 *mspms*: A Comprehensive R Package and Graphic Interface for Multiplex Substrate Profiling by
2 Mass Spectrometry Analysis

3 Charlie Bayne^{1,2,3}, Brianna Hurysz^{1,2,3}, David J. Gonzalez^{2,3,4}, Anthony O'Donoghue^{2,3}

4 Institutions:

5 ¹Biomedical Sciences Graduate Program, UC San Diego, La Jolla, California 92093, USA.

6 ²Department of Pharmacology, University of California San Diego, La Jolla, California 92093,
7 USA

8 ³Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego,
9 La Jolla, California, USA

10 ⁴Center for Microbiome Innovation, University of California at San Diego, La Jolla, California
11 92093, USA

12

13 Abstract:

14 Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS) is a powerful method for
15 determining the substrate specificity of proteolytic enzymes, knowledge key for developing
16 protease inhibitors, diagnostics, and protease-activated therapeutics. However, the complex
17 datasets generated by MSP-MS pose significant analytical challenges. To address this, we
18 developed *mspms*, a Bioconductor R package complemented by an intuitive graphical interface.

19 *Mspms* streamlines MSP-MS data analysis by standardizing workflows for data preparation,
20 processing, statistical analysis, and visualization. Designed for accessibility, it serves both
21 advanced users via the R package and broader audiences through the web interface. We validated
22 *mspms* by profiling the substrate specificity of four well-characterized cathepsins (A–D),
23 demonstrating its ability to reliably capture expected substrate specificities. As the first publicly

24 available platform for MSP-MS data analysis, *mspms* delivers comprehensive functionality,
25 transparency, and ease of use, making it a valuable resource for the protease research
26 community. Access to *mspms* is available through the Bioconductor project at
27 <https://bioconductor.org/packages/mspms>, and a graphic interface is available at
28 https://gonzalezlab.shinyapps.io/mspms_shiny/.

29

30 Introduction:

31 Proteases play crucial roles in a wide range of biological processes, from digestion and immunity
32 to cancer and neurodegenerative diseases [1]. Understanding the substrate specificity of these
33 enzymes is essential for designing inhibitors, diagnostics, and protease-activated therapeutics [2].
34 One of the most effective methods for determining protease substrate specificity is Multiplex
35 Substrate Profiling by Mass Spectrometry (MSP-MS) [3]. This technique involves incubating a
36 rationally designed peptide library with a protease or protease-containing sample and using mass
37 spectrometry to identify the resulting cleavage products, revealing the enzyme's substrate
38 preferences [4].

39

40 The data produced by MSP-MS are complex and multi-dimensional. Accurate interpretation of
41 these results requires rigorous data analysis, encompassing multiple steps: preparing the data
42 (identifying cleavage motifs, and positions), processing the data (data transformation,
43 normalization, and imputation), statistical analysis, and data visualization. Historically, analysis
44 of MSP-MS data has lacked dedicated analysis tools, leaving each researcher to analyze their
45 data in an *ad-hoc* manner. This approach, while functional, results in an inherent lack of
46 reproducibility. Inconsistent and irreproducible analysis pipelines have been noted to lead to

47 grave problems in biological research [5], often a consequence of decentralized, error prone
48 evolution of codebases as personnel transitions [6]. Such absence of standardized, reproducible
49 data analysis tools for MSP-MS has limited scientific progress by creating a barrier for
50 collaboration across research groups.

51

52 To address these challenges, we developed *mspms*, an R package specifically designed for the
53 robust, reproducible analysis of MSP-MS data. Through integration into the Bioconductor
54 ecosystem [7], the *mspms* package adheres to best practices in software development and data
55 analysis, offering a transparent and portable solution for processing complex datasets.

56 Recognizing that many users may not have programming experience, we complemented the R
57 package with a user-friendly graphical interface, available both as a web application and for
58 download. This interface allows researchers to perform key MSP-MS analysis steps—data
59 preprocessing, normalization, statistical analysis, and visualization—without needing any R
60 programming knowledge.

61

62 Here, we demonstrate the functionality of *mspms* by analyzing publicly available MSP-MS data
63 for four well-characterized cathepsins, validating the package’s ability to accurately determine
64 their substrate specificities. By offering comprehensive functionality, transparency, and user
65 accessibility, *mspms* is positioned to be a valuable tool for the protease research community,
66 streamlining the analysis of MSP-MS data while promoting reproducible research.

67

68 Materials and Methods

69 Data Used for Study

70 Raw data from a previously reported MSP-MS study was acquired from the MassIVE Repository
71 (accession number MSV00008595) [8]. In brief, these data were generated from a study that
72 utilized a 228-member peptide library that was incubated with either 18.4 nM cathepsin A, 2.64
73 nM cathepsin B, 19.6 nM cathepsin C or 100 nM cathepsin D. The concentration of each peptide
74 in the library was 0.5 µM. After incubation at 37°C for defined time points, the reaction was
75 quenched by addition of 6.4 M guanidine hydrochloride. These samples were desalted with C18
76 spin columns, and ~0.4 µg of each sample was subjected to LC-MS/MS analysis using an
77 Ultimate 3000 HPLC and Q-Exactive mass spectrometer. LC and MS parameters were as
78 previously reported [8].

79

80 **Upstream Proteomic Software**

81 Peptides/proteins were identified and quantified using PEAKS Studio [9], Proteome Discoverer
82 [10], and FragPipe [11]. The database used in each search was the 228-member peptide library
83 described previously [3] (Supplementary File 1).

84

85 **PEAKS Studio**

86 Data from all .raw files were processed using PEAKS Studio v8.5 software, using a customized
87 template (Supplementary File 2). For each sample experiment specific parameters were set as
88 follows: Q-Exactive instrument, HCD fragmentation, no enzyme. Scans were merged with a
89 retention time window of 0.8 min, and precursor m/z error tolerance of 10 ppm. Precursor mass
90 was corrected. Scans were filtered to include retention time between 0 and 95 min with a
91 precursor mass tolerance of 10 ppm. For identification, a precursor mass tolerance of 20 ppm
92 using monoisotopic mass and a fragmentation ion of 0.01 Da was specified. No PTMs were

93 included in the search. FDR was estimated using decoy-fusion strategy. Label free quantification
94 was performed with a mass error tolerance of 9 ppm, and retention time shift tolerance of 3
95 minutes. Replicate samples were added to new groups.

96

97 The peaks_protein-peptides-lfq.csv file was prepared by navigating to the quantification options
98 setting the normalization factor to “No normalization”, changing peptide filters to include all
99 peptides (quality ≥ 0 , Avg.Area ≥ 0 , Peptide ID Count ≥ 0 , charge +1 - +10, and at least 1
100 confident sample). Protein filters were changed so no filtering occurs (Significance ≥ 0). Data
101 was then exported as the peptides-lfq.csv file.

102

103 Data in figures is derived from the PEAKS software, unless otherwise specified.

104

105 **Proteome Discoverer**

106 Data from all .raw files was also processed using Proteome Discoverer V 2.5.0.400 using a
107 customized processing and consensus workflow. (Supplemental Files 3 and 4). Briefly, min
108 precursor mass was specified as 350 Da, max precursor mass was specified as 5000. Enzyme
109 was set to be unspecific, with a min peptide length of 5, and max peptide length of 14. Precursor
110 mass tolerance was set to 10 ppm, and fragment mass tolerance was set to 0.6. Percolator Target
111 FDR was set to 0.01.

112

113 **FragPipe**

114 FragPipe V22.0. MSFragger version 4.1, IonQuant version 1.10.27, and Python version 3.9.13
115 were used to process all .raw files with a customized analysis workflow derived from the MBR-

116 LFQ workflow template (Supplementary File 5). Briefly, decoys were added to the peptide
117 library database, cleavages were set to nonspecific; peptide length was set at 5-14, and 350 Da to
118 5000 Da, match between runs was enabled, and top runs was set to 3 (as there were 4 biological
119 replicates in each group except for cathepsin D at time zero).

120

121 **Preprocessing**

122 To prepare MSP-MS data for analysis, *mspms* preprocesses an exported file from the user's
123 proteome software of choice. In this process, the data is converted to a standardized format and
124 loaded as a QFeatures object [12] containing a SummarizedExperiment [13] object named
125 "peptides", which contains the detected peptide intensities. Cleavage motifs of a user specified
126 length to the left and right of the scissile bond are calculated, and the numerical location of the
127 cleavage site (via reference to the member of the library it was derived from) is determined.
128 These peptide centric features are then loaded as the rowData corresponding to the QFeatures
129 object. The colData composing the QFeatures experiment contains sample metadata describing
130 the experiment and must include descriptors core to every MSP-MS experiment: "group",
131 "condition", and "time".

132

133 **Data Processing**

134 Peptide values are subjected to \log_2 transformation followed by a median centered normalization
135 utilizing the center.median method. Due to the left-censored nature of MSP-MS data, imputation
136 is subsequently performed using the QRILC method. Lastly, the data is reverse \log_2 transformed.
137 All data manipulation is performed using MScoreutils [14]. Data resulting from each step of
138 data processing is stored within the resulting QFeatures object as SummarizedExperiment

139 objects named “log2_peptides”, “log2_peptides_norm”, “log2_peptides_norm_imputed”, and
140 “peptides_norm” respectively.

141

142 Statistics

143 *Mspms* calculates the \log_2 fold change relative to the user specified denominator in colData and
144 then performs pairwise t-tests with FDR p adjustment as implemented in the Rstatix [15]
145 package using the peptides_norm values. By default, significant peptides are denoted as having
146 a p.adj < 0.05 and \log_2 fold change > 3 relative to time 0.

147

148 Visualizations

149 *Mspms* supports ggplot2 [16] based plotting of several types of static visualizations including
150 quality control, PCA, volcano, time course, and iceLogo plots. Interactive heatmaps are plotted
151 using the plotly [17] based heatmaply [18] library .

152

153 iceLogo Analysis

154 iceLogo [19] analysis calculates the chance of occurrence (p-value) of every amino acid
155 surrounding the scissile bond in the experimental data relative to a user-defined reference set of
156 possible scissile bond locations. We implemented this approach in R, using the previously
157 described Java implementation as a reference. The underlying logic is described below.

158

159 First, a count of the number of times an amino acid at each position is calculated for an
160 experimental and reference set.

161

162 Then the frequency of each amino acid at each position is calculated.

163

164 Then the standard deviation (σ) is calculated using the frequency of an amino acid in the

165 reference set ($f\%$)

$$\sigma = \sqrt{\frac{f\%}{N}}$$

167

168 These calculated standard deviations are subsequently used to calculate significances by

169 conversion to p-values using the Wichura algorithm. Only p values \leq to the user specified p

170 value threshold are retained for subsequent visualization.

171

172 The height of each amino acid at each position is then visualized using the ggseqlogo R package

173 [20] using the user's choice of percent change or fold change to represent the height of each

174 amino acid letter.

175

176 The percent change (PC) is calculated from the experimental set frequency (F^+) and reference

177 set frequency (F^-) as follows:

$$178 \quad PC = F^+ - F^-$$

179 The fold change (FC) is calculated from the experimental set frequency (F^+) and reference set

180 frequency (F^-) as follows:

$$181 \quad FC = \frac{F^+}{F^-}$$

182 In the event that the fold change is smaller than one, it is transformed into the converted fold
183 change ($FCcon$) in order to allow the comparison of height with positively regulated amino
184 acids.

185
$$FCcon = \frac{1}{FC} \times -1$$

186

187 If only one amino acid is found to be significant at a given position and the calculated amino acid
188 size is infinite, the height of the amino acid is represented as the maximal height that can be
189 visualized in the iceLogo plot.

190

191 If several amino acids are found to be significant at a given position and all have infinite
192 calculated amino acid sizes, the height of the amino acids combined is represented as the
193 maximal height that can be visualized in the iceLogo plot.

194

195 **Report Generation**

196 *Mspms* supports the production of a generic *mspms* report (Supplementary File 6). This function
197 produces a generic self-contained .html report with embedded downloadable data frames
198 (containing normalized data and results of statistics), and figures. This report is produced by
199 leveraging the *mspms* R package inside of a parameterized rmarkdown [21] template
200 incorporating the downloadthis [22] package.

201

202 **Helper Functions**

203 Only a subset of functions are exported to the user in order to maintain an intuitive application
204 programming interface. Helper functions can be found in the helper_functions.R file named
205 corresponding to the type of functions it assists with.

206

207 **Graphic Interface**

208 A graphic interface to the *mspms* R package is implemented using the R shiny framework [23].
209 This interface is accessible on the web at https://gonzalezlab.shinyapps.io/mspms_shiny/ or
210 downloadable from <https://github.com/bayne2/mspms-shiny/>.

211

212 **Results:**

213 The *mspms* R package was developed to provide a dedicated tool to analyze MSP-MS data,
214 focusing on reproducibility, ease of use, and robustness. It includes modular functions to handle
215 key steps generalizable to any MSP-MS analysis: data preparation, processing, statistical
216 analysis, and visualization (Figure 1).

217

218 **Data Quality Evaluation**

219 To assess the quality of the MSP-MS data from the cathepsin A, B, C, and D experiments, we
220 applied the quality control functions of *mspms*. We found that over 90% of the full-length
221 peptide library was detected in all samples at time zero (T0), and more than 95% of the library,
222 including cleavage products, was detected across the dataset (Supplementary Figure 4A). Only
223 five peptides from the library were consistently missing across all samples, suggesting high-
224 quality data with minimal loss (Supplementary Figure 4B).

225

226 **Evaluation of Global Data Patterns**

227 Next, we examined global patterns in the dataset using principal component analysis (PCA) and
228 unsupervised hierarchical clustering. PCA demonstrated tight clustering of replicates within each
229 experimental group (condition and timepoint), as shown by the 95% confidence intervals
230 surrounding each group (Figure 2A). Near-perfect clustering of replicates from identical
231 conditions was observed, indicating high experimental consistency. Differential peptide
232 abundance between groups was evident, supporting distinct activity for each cathepsin over time
233 (Figure 2B).

234

235 **Significant Peptide Changes and Cleavage Position Preferences**

236 We analyzed significant peptide differences for each cathepsin relative to T0 using t-tests and
237 log₂ fold change calculations. The results, visualized as volcano plots, revealed numerous
238 significantly upregulated peptides (log₂ fold change ≥ 3 , p.adj ≤ 0.05) at various timepoints
239 following incubation (Figure 3B). The number of significantly different peptides increased
240 progressively with time for each cathepsin, highlighting the dynamic substrate cleavage behavior
241 (Supplementary Figure 5).

242

243 To evaluate protease activity relative to reported substrate specificities (Table 3A), we
244 investigated cleavage site preferences within the 14-mer peptides. Cathepsin A showed clear
245 carboxypeptidase activity through the high number of cleavage sites at the C-terminus (Figure
246 3C) and overrepresentation of X (corresponding to no amino acid) at P2', P3', and P4' (Figure
247 3D). Cathepsin B displayed dipeptidyl carboxypeptidase activity, with most cleavage sites
248 occurring at position 12 and a lesser amount at position 10, suggesting sequential removal of

249 dipeptides from the C-terminus (Figure 3C). The enrichment of X at P3' and P4' evident in the
250 iceLogo plot further supported this dipeptidyl carboxypeptidase activity (Figure 3D).

251
252 Assessment of cathepsin C revealed a high number of cleavage sites at position 2, followed by a
253 smaller number at position 4, suggesting the sequential removal of dipeptides from the N-
254 terminus (Figure 3C). An overrepresentation of X at P4 and P3 in the iceLogo also confirmed the
255 dipeptidyl aminopeptidase activity (Figure 3D). Lastly, cathepsin D showed endopeptidase
256 activity as observable by the Gaussian peak centered around position 8 of the cleavage position
257 plots (Figure 3C). The iceLogo plot showed that X was not enriched at any of the sites from P4
258 to P4' further validating this endopeptidase activity (Figure 3D).

259

260 **Amino Acid Preferences**

261 To visualize the amino acid preferences at cleavage sites, we performed an iceLogo analysis
262 using *mspms*, focusing on the eight positions surrounding the cleavage site (P4 to P4').

263

264 Cathepsin A showed a preference for the removal of hydrophobic amino acids (such as Phe and
265 Leu) from the C-terminus of substrates, when additional hydrophobic residues occupy the P1
266 position (Figure 3D). Cathepsin B favored substrates with positively charged (Arg, Lys) or
267 hydrophobic residues in P1 and P2 (Figure 3D). Cathepsin C showed a limited preference at the
268 P1 and P1' position (Figure 3D). Cathepsin D showed a preference for Phe, Tyr, and norleucine
269 at the P1 and P1' position (Figure 3D).

270

271 **Comparison of Results Across Different Upstream Proteomics Software**

272 To demonstrate the compatibility of MSP-MS with a range of upstream proteomics software, we
273 analyzed MSP-MS data using PEAKS Studio, Proteome Discoverer (PD), and FragPipe
274 separately, comparing their outputs (Supplementary Figure 1).

275

276 Processed data corresponding to peptides detected across all three approaches correlated well,
277 with R² values ranging from 0.70 (FragPipe to PD) to 0.81 (FragPipe to PEAKS)
278 (Supplementary Figure 1A). When assessing peptides identified as significantly different from
279 time zero, approximately 30% of significant peptides from cathepsins A, B and C were shared
280 across all three platforms (Supplementary Figure 1B). For cathepsin D, the only endopeptidase in
281 our study, there was only 17% agreement across the three software tools. However, an increase
282 in shared peptide identities was observed between FragPipe and PEAKS for cathepsin D (27%)
283 compared to cathepsins A–C (~10%) (Supplementary Figure 1B).

284

285 Positional specificity, as evaluated through cleavage position plots, was highly consistent for
286 cathepsins A, B and C across all tools. However, PD showed limited ability to capture the
287 expected endopeptidase activity of cathepsin D, while FragPipe results displayed more N-
288 terminal cleavage than observed in PEAKS data (Supplementary Figure 1C).

289

290 The specificity profiles at positions P4–P4', evaluated using iceLogos, were comparable across
291 all tools, with subtle differences depending on the software. Major discrepancies were observed
292 in the cathepsin D results, consistent with findings from cleavage position plots (Supplementary
293 Figure 2).

294

295 Since cathepsin D was the only endopeptidase included in this study, we hypothesized that the
296 observed discrepancies arose from differences in each tool's ability to detect shorter peptides,
297 which are more commonly generated by endopeptidases. Analyzing the distribution of
298 significant peptides by length revealed that PD systematically detected fewer peptides shorter
299 than eight amino acids compared to PEAKS and FragPipe (Supplementary Figure 3). This likely
300 explains PD's poorer performance on cathepsin D compared to the other tools.

301

302 Discussion:

303 Before the development of the *mspms* package, MSP-MS data analysis relied heavily on ad hoc
304 developed R scripts, which posed significant challenges. These scripts were fragmented, poorly
305 documented, and difficult to adapt, making reproducibility a concern. Researchers without
306 specialized programming skills struggled with customizing these workflows to accommodate
307 different experimental designs, limiting the broader utility of MSP-MS data analysis.
308 Furthermore, only data exported from the proteomic search engine PEAKS was compatible,
309 hindering usability across different research groups.

310

311 The *mspms* R package effectively addresses key limitations in MSP-MS data analysis through its
312 modular, reproducible, user-friendly approach, and compatibility with a wide range of
313 proteomics software. It provides self-contained functions for data preparation, processing,
314 statistical analysis, and visualization, ensuring ease of maintenance, extensibility, and usability.

315

316 One significant feature is the integration of functionality from the widely cited iceLogo tool
317 within R, which allows for the analysis of nonstandard amino acids, such as norleucine, and

318 positions marked by “X.” Additionally, *mspms* includes a graphical user interface, accessible
319 both online and via local download, enabling researchers without R programming experience to
320 leverage its core functionalities. Moreover, *mspms* integrates smoothly with the Bioconductor
321 ecosystem. By employing established S4 classes internally, it offers R users the flexibility to tap
322 into Bioconductor’s extensive analytical resources, further enabling advanced data exploration,
323 statistical analysis, and visualizations. This makes *mspms* a versatile and adaptable tool that
324 meets the diverse needs of the protease research community.

325

326 In the present study, we applied *mspms* to profile the substrate specificity of four well-
327 characterized cathepsin proteases, namely cathepsin A, B, C, and D. In doing so, we demonstrate
328 features of *mspms* that are broadly applicable to any MSP-MS experiment, while rigorously
329 benchmarking the software by evaluating its ability to detect substrate specificities accepted to
330 be the ground truth.

331

332 A critical but often overlooked step in proteomic analysis is conducting a thorough quality
333 control assessment to ensure data quality is sufficient for drawing biologically meaningful
334 conclusions [24]. Since each MSP-MS experiment is based on a known peptide library, an
335 effective quality control measure involves evaluating the percentage of the un-cleaved peptide
336 library detected in each sample. Ideally, 100% of the un-cleaved peptide library should be
337 detectable at T0; however, due to limitations in mass spectrometry performance, this is rarely
338 achieved in practice. When we applied this quality control check to our cathepsin experiment, we
339 observed no indication of data quality issues, confirming the reliability of our results.

340

341 The next step of an MSP-MS analysis is to explore global data patterns present in the data. This
342 approach allows users the ability to evaluate the data as a whole, determine whether the
343 experiment was technically successful (by verifying that the positive and negative controls
344 performed as expected), and identify interesting patterns. *Mspms* allows the user to easily create
345 PCA and interactive heatmap plots. In our cathepsin MSP-MS experiment, PCA and heatmap
346 analyses revealed tight clustering among replicate conditions, indicating minimal variability
347 across biological replicates. Moreover, distinct clustering of different experimental groups based
348 on cathepsin type and time point highlighted the unique substrate specificities of each enzyme.
349 These visualizations underscore the reliability of the experiment and the capacity of *mspms* to
350 facilitate robust, comprehensive data exploration.

351
352 Once the MSP-MS experiment is confirmed as a technical success, the main objective—
353 determining the enzymes' substrate specificity—is readily achievable using the *mspms* tool.
354 First, *mspms* computes the \log_2 fold change and FDR-corrected p-values from t-tests conducted
355 on normalized and imputed intensity values. The specific features of significantly altered
356 peptides are then examined via cleavage location plots, which illustrate the enzyme's positional
357 specificity, and iceLogo plots, which reveal amino acid preferences.

358
359 When applied to the cathepsin data, our analysis near perfectly captured previously reported
360 substrate specificities for each of the enzymes profiled. We show that:
361 1. Cathepsin A is a carboxypeptidase that preferentially removes hydrophobic amino acids
362 (such as Phe and Leu) from the C-terminus of substrates, especially when additional
363 hydrophobic residues occupy the P1 position, as previously reported [25].

364 2. Cathepsin B is a dipeptidyl carboxypeptidase that cleaves dipeptides from the C-
365 terminus, favoring substrates with positively charged (Arg, Lys) or hydrophobic residues
366 in P1 and P2, as previously reported [26].

367 3. Cathepsin C functions as a dipeptidyl aminopeptidase, cleaving dipeptides from the N-
368 terminus with broad specificity, as previously reported [8].

369 4. Cathepsin D is an endopeptidase that cleaves between hydrophobic amino acids,
370 including Phe, Leu, and Tyr, as previously reported [27].

371 By validating the specific activities of these cathepsins, *mspms* confirms its capability to
372 accurately identify expected substrate specificities, establishing its value as a powerful tool in
373 protease research. Beyond the enzymes evaluated in this study, the package's modular and
374 flexible design enables it to be readily applied to analyze the substrate specificity of virtually any
375 protease mixture, allowing for diverse applications across the protease field.

376
377 Moreover, *mspms* is built to integrate seamlessly with future advancements in MSP-MS assays.
378 As peptide synthesis becomes more cost-effective and mass spectrometer technology advances, it
379 will be feasible to expand the MSP-MS assay by incorporating significantly larger peptide
380 libraries than the current 228-member set. The ability to support any peptide library, coupled
381 with its reproducible workflows and user-friendly features, secures *mspms* as an enduring and
382 valuable asset to the protease research community.

383
384 To enhance accessibility, MSP-MS is designed to be compatible with three major proteome
385 search engines: PEAKS Studio, Proteome Discoverer (PD), and FragPipe. Compatibility was
386 validated by independently analyzing MSP-MS data using each software tool. Substrate

387 specificity interpretations for cathepsins A, B, and C were remarkably consistent across all
388 platforms.

389
390 For cathepsin D, both PEAKS and FragPipe effectively detected endopeptidase activity, but
391 FragPipe identified a higher frequency of significant N-terminal cleavages relative to PEAKS.
392 Determining which profile is more biologically accurate would require further validation using
393 orthogonal assays. Proteome Discoverer, in contrast, struggled to convincingly identify cathepsin
394 D's endopeptidase activity, likely due to its reduced sensitivity for shorter peptides relative to
395 PEAKS and FragPipe.

396
397 If endopeptidase activity is a primary focus, we recommend using PEAKS Studio or Fragpipe,
398 which provided the best detection of cathepsin D activity in our tests. We note that FragPipe is
399 an attractive option, particularly as it is an open-source tool freely available for academic use and
400 demonstrated analysis speeds at least an order of magnitude faster than the paid software
401 solutions. We caution against using Proteome Discoverer with the search settings applied in this
402 study unless the parameters are further optimized, which may be able to increase the ability to
403 detect smaller peptides. If endopeptidase activity is not of interest, our results indicate that all
404 three software platforms perform comparably.

405

406 **Conclusion**

407 In summary, *mspms* streamlines MSP-MS data analysis, providing a reliable, reproducible, and
408 adaptable platform for protease substrate profiling. The combination of its powerful analytical
409 capabilities and intuitive design enables researchers to extract biologically meaningful insights

410 from complex datasets with minimal technical barriers. Given its flexibility and broad
411 applicability, *mspms* is positioned to become a standard tool in protease research, offering
412 significant advancements in the study of proteolytic enzymes and their roles in health and
413 disease.

414

415 Availability of Source Code and Requirements:

- 416 • Project name: *mspms*
- 417 • Operating system(s): Linux, macOS, Windows
- 418 • Programming language: R
- 419 • Other requirements: R 4.4.0, QFeatures, SummarizedExperiment, magrittr, rlang, dplyr,
420 purrr, stats, tidyverse, stringr, ggplot2, ggseqlogo, heatmaply, readr, rstatix, tibble, ggpubr.
- 421 • License: MIT
- 422 • Bioconductor home page: <https://bioconductor.org/packages/mspms>
- 423 • GitHub home page: <https://github.com/bayne2/mspms>
- 424 • Shiny app instance: https://gonzalezlab.shinyapps.io/mspms_shiny/
- 425 • Shiny app repository: <https://github.com/bayne2/mspms-shiny>
- 426 • Vignette :
427 https://bioconductor.org/packages-devel/bioc/vignettes/mspms/inst/doc/mspms_vignette.html
- 429 • Manuscript repository: <https://github.com/bayne2/mspms-manuscript>.

430

431 Availability of Supporting Data and Materials:

432 All data used to build this manuscript can be found in the GitHub repository for the manuscript
433 (https://github.com/bayne2/mspms_manuscript). Mass spectrometry data in. raw format is
434 available from MassIVE Repository under accession number MSV00008595.

435

436 Figure Legends

437 **Figure 1. Overview of the *mspms* R package and MSP-MS profiling of cathepsin proteases.**

438 Schematic of the functions contained within the *mspms* R package.

439

440 **Figure 2. Global visualization of MSP-MS data.**

441 (A) Principal component analysis displaying PC1 and PC2. Samples are colored by time, while
442 the shape and line type show the type of cathepsin with eclipses representing the 95% confidence
443 interval. (C) Heatmap showing the results of the experiment as clustered using unsupervised
444 hierarchical clustering. Rows of the heatmap represent the samples while columns represent the
445 peptides. Color of the heatmap cells represent the normalized, column centered, and scaled
446 values. Colored bars to the right of the heatmap indicate the cathepsin and time of the samples in
447 each row. Colored bars corresponding to each peptide in the columns display whether the
448 corresponding peptide is a full-length peptide belonging to the 228-member peptide library (non
449 -cleaved, dark blue) or a cleavage product (cleaved, blue)

450

451 **Figure 3. Differentially abundant peptide cleavages over time.**

452 (A) Summarized substrate specificities for cathepsin A- D as reported in the literature.

453 (B) Volcano plots displaying the log₂-fold change of the timepoint as indicated by color relative
454 to and -log₁₀ FDR corrected p values for each cathepsin. (C) Plot showing the number of

455 significant cleavage events at each position of the peptide library (as defined as having a \log_2
456 fold change ≥ 3 and FDR adjusted p values ≤ 0.05) (D) IceLogo plots as implemented in the
457 *mspms* package. Amino acid residues (with X representing positions past the terminus) four
458 positions to the left and right of the cleavage site are displayed. Only residues with significantly
459 higher proportions relative to the proportion of all possible cleavage sequences present in the
460 initial peptide library ($pval \leq 0.05$) are shown, with the height representing the percentage
461 differences.

462

463 Supplementary Figure Legends

464 **Supplementary Figure 1. Comparison of Results from Upstream Proteomic Software**

465 **Compatible with *mspms*.**

466 (A) Correlation analysis of shared peptides detected by PEAKS Studio, Proteome Discoverer
467 (PD), and FragPipe. (B) Venn diagram showing peptides identified as significantly different by
468 each software tool. (C) Cleavage site plots illustrating peptide cleavage patterns across
469 comparisons made using the three tools.

470

471 **Supplementary Figure 2. Comparison of IceLogos from Upstream Proteomic Software**

472 **Compatible with *mspms*.**

473 IceLogo generated using peptide data exported from PEAKS Studio, Proteome Discoverer (PD),
474 and FragPipe.

475

476 **Supplementary Figure 3. Comparison of Detected Peptide Lengths Using Different**

477 **Upstream Proteomic Software Compatible with *mspms*.**

478 (A) Count of significant peptides detected by length, categorized by the upstream proteomic
479 software tool used.
480 (B) Count of total detected peptides by length categorized by the upstream proteomic software
481 tool used.

482

483 **Supplementary Figure 4. Quality Control Evaluation.**

484 (A) Histogram displaying the count of samples for each sample grouping at time 0 as a function
485 of the percentage of undetected full-length and cleavage product peptides mapping to the 228-
486 peptide library. (B) Percent of samples that indicated member of the 228-peptide library was
487 undetected in when considering only full-length or cleavage product peptides.

488

489 **Supplementary Figure 5. Number of Significant Differences Relative to T0 as a Function of**
490 **MSP-MS Incubation Time.** The number of significantly enriched peptides relative to time 0 as
491 shown per duration of time incubated with the indicated cathepsin.

492

493 **Supplementary Figure 6. IceLogo Analyses with Extended Cleavage Motifs.** IceLogo
494 analysis of peptide cleavage motifs containing the 6 amino acids before and after the
495 significantly enriched peptides with detected cleavage sites relative to all possible within the
496 background of the 228-peptide library used for the experiment.

497

498 **Supplementary Figure 7. Screenshots of *mspms* graphic interface.**

499 (A) About page. (B) File upload page. Once files are uploaded, subsequent pages become
500 available to the user. (C) Processed data page containing normalized and imputed data. (D) Page

501 containing quality control plots. (E) Stats page containing results of log₂ fold change and FDR
502 corrected t-tests relative to time 0 for each condition. User selected peptides can be interactively
503 plotted with either the normalized imputed data, or the raw intensities. (F) DataViz page
504 containing PCA, interactive heatmap, volcano plot, or iceLogo plots. (G) Page containing button
505 to generate a self-contained *mspms* html report.

506

507 Supplementary Files

508 **Supplementary File 1. Peptide library fasta file used as proteomics search database.**

509 **Supplementary File 2. Peptide library .csv file corresponding to Supplementary file 1.**

510 **Supplementary File 3. Screenshot of parameters used in PEAKS studio search.**

511 **Supplementary File 4. Proteome Discoverer analysis template used for *mspms* analysis.**

512 **Supplementary File 5. Proteome Discoverer consensus template used for *mspms* analysis.**

513 **Supplementary File 6. Fragpipe workflow used for *mspms* analysis.**

514 **Supplementary File 7. Generic *mspms* .html report for cathepsin A-D data.**

515

516 Abbreviations:

517 MSP-MS: Multiplex Substrate Profiling by Mass Spectrometry. T0: Time zero. PCA: Principal

518 Component Analysis. PC1: Principal component 1. PC2: Principal component 2. FDR: False

519 Discovery Rate. PD: Proteome Discoverer.

520

521 Competing Interests:

522 The authors declare that they have no competing interests.

523

524 Funding:

525 Charlie Bayne and Brianna Hurysz were supported in part by the UCSD Graduate Training
526 Program in Cellular and Molecular Pharmacology through an institutional training grant from the
527 National Institute of General Medical Sciences, T32 GM007752. Dr. O'Donoghue would like to
528 acknowledge the following NIH funding to support this research, R01AI158612, R21AI171824
529 and R21CA256460. This study was also supported by the UCSD Collaborative Center of
530 Multiplexed Proteomics.

531

532 Author contributions

533 C.B wrote the R package, shiny app, documentation, and manuscript; B.H made substantial
534 contributions to the conception and design of the work, while D.J.G. and A.J.O provided
535 funding, oversaw the project and provided contributions to the conception and design of the
536 work. All authors edited and approved the final version of the manuscript.

537

538 Acknowledgements

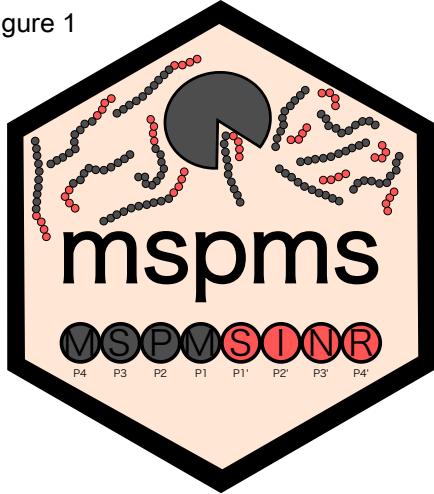
539 We would like to acknowledge Dr. Zhenze Jiang for his initial conception of previous R scripts,
540 as well as Dr. Lawrence Liu and Dr. Michael Yoon for maintaining and updating the scripts. We
541 also thank Dr. Jiang and Dr. Yoon for running the MSP-MS experiment analyzed in this
542 manuscript. Finally, we would like to acknowledge Diego F. Trujillo for suggestions to improve
543 *mspms* and for educating new users how to perform the data analysis.

544

545 References

- 546 1. López-Otín C, Bond JS. Proteases: Multifunctional Enzymes in Life and Disease. *Journal of*
547 *Biological Chemistry*. 2008; doi: 10.1074/jbc.R800035200.
- 548 2. Leung D, Abbenante G, Fairlie DP. Protease Inhibitors: Current Status and Future Prospects. *J*
549 *Med Chem*. 2000; doi: 10.1021/jm990412m.
- 550 3. Rohweder PJ, Jiang Z, Hurysz BM, O'Donoghue AJ, Craik CS. Multiplex substrate profiling
551 by mass spectrometry for proteases. *Methods in Enzymology*. Elsevier;
- 552 4. O'Donoghue AJ, Eroy-Reveles AA, Knudsen GM, Ingram J, Zhou M, Statnekov JB, et al..
553 Global identification of peptidase specificity by multiplex substrate profiling. *Nat Methods*.
554 2012; doi: 10.1038/nmeth.2182.
- 555 5. Miller G. A Scientist's Nightmare: Software Problem Leads to Five Retractions. *Science*.
556 2006; doi: 10.1126/science.314.5807.1856.
- 557 6. Casadevall A, Steen RG, Fang FC. Sources of error in the retracted scientific literature.
558 *FASEB j.* 2014; doi: 10.1096/fj.14-256735.
- 559 7. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al.. Bioconductor:
560 open software development for computational biology and bioinformatics. *Genome Biology*.
561 2004;
- 562 8. Jiang Z, Lietz CB, Podvin S, Yoon MC, Toneff T, Hook V, et al.. Differential
563 Neuropeptidomes of Dense Core Secretory Vesicles (DCSV) Produced at Intravesicular and
564 Extracellular pH Conditions by Proteolytic Processing. *ACS Chem Neurosci*. 2021; doi:
565 10.1021/acschemneuro.1c00133.
- 566 9. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, et al.. PEAKS: powerful
567 software for peptide *de novo* sequencing by tandem mass spectrometry. *Rapid Comm Mass*
568 *Spectrometry*. 2003; doi: 10.1002/rcm.1196.
- 569 10. Orsburn BC. Proteome Discoverer—A Community Enhanced Data Processing Suite for
570 Protein Informatics. *Proteomes*. 2021; doi: 10.3390/proteomes9010015.
- 571 11. Yu F, Haynes SE, Nesvizhskii AI. IonQuant Enables Accurate and Sensitive Label-Free
572 Quantification With FDR-Controlled Match-Between-Runs. *Molecular & Cellular Proteomics*.
573 2021; doi: 10.1016/j.mcpro.2021.100077.
- 574 12. Gatto L, Vanderaa C. QFeatures: Quantitative features for mass spectrometry data.
- 575 13. Morgan M, Obenchain V, Hester J, Pagès H. SummarizedExperiment.
- 576 14. Rainer J, Vicini A, Salzer L, Stanstrup J, Badia JM, Neumann S, et al.. A Modular and
577 Expandable Ecosystem for Metabolomics Data Annotation in R. *Metabolites*. 2022; doi:
578 10.3390/metabo12020173.
- 579 15. Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests.

- 580 16. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al.. *ggplot2: Create*
581 *Elegant Data Visualisations Using the Grammar of Graphics.*
- 582 17. Plotly Technologies Inc.. *Collaborative data science.* Montréal, QC: Plotly Technologies Inc;
- 583 18. Galili T, O'Callaghan A, Sidi J, Sievert C. *heatmaply: an R package for creating interactive*
584 *cluster heatmaps for online publishing.* Wren J, editor. *Bioinformatics.* 2018; doi:
585 10.1093/bioinformatics/btx657.
- 586 19. Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Improved visualization of
587 protein consensus sequences by iceLogo. *Nat Methods.* 2009; doi: 10.1038/nmeth1109-786.
- 588 20. Wagih O. *ggseqlogo: a versatile R package for drawing sequence logos.* Hancock J, editor.
589 *Bioinformatics.* 2017; doi: 10.1093/bioinformatics/btx469.
- 590 21. Allaire JJ, Xie Y, Dervieux C, McPherson J, Luraschi J, Ushey K, et al.. *rmarkdown:*
591 *Dynamic Documents for R.*
- 592 22. Maturana FM. *downloadthis: Implement Download Buttons in “rmarkdown.”*
- 593 23. Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, et al.. *shiny: Web Application*
594 *Framework for R.*
- 595 24. Rozanova S, Uszkoreit J, Schork K, Serschnitzki B, Eisenacher M, Tönges L, et al.. Quality
596 *Control—A Stepchild in Quantitative Proteomics: A Case Study for the Human CSF Proteome.*
597 *Biomolecules.* 2023; doi: 10.3390/biom13030491.
- 598 25. Cuervo AM. Cathepsin A regulates chaperone-mediated autophagy through cleavage of the
599 *lysosomal receptor.* *The EMBO Journal.* 2003; doi: 10.1093/emboj/cdg002.
- 600 26. Yoon MC, Solania A, Jiang Z, Christy MP, Podvin S, Mosier C, et al.. Selective Neutral pH
601 *Inhibitor of Cathepsin B Designed Based on Cleavage Preferences at Cytosolic and Lysosomal*
602 *pH Conditions.* *ACS Chem Biol.* 2021; doi: 10.1021/acschembio.1c00138.
- 603 27. Ivry SL, Sharib JM, Dominguez DA, Roy N, Hatcher SE, Yip-Schneider MT, et al.. Global
604 *Protease Activity Profiling Provides Differential Diagnosis of Pancreatic Cysts.* *Clinical Cancer*
605 *Research.* 2017; doi: 10.1158/1078-0432.CCR-16-2987.
- 606



Data Preparation

Preparing data exported from proteomics software

- prepare_peaks()
- prepare_pd()
- prepare_fragpipe()

Calculating all possible cleavages of peptide library

- calculate_all_cleavages()

Data Processing

Data transformation, normalization, and imputation

- process_qf()

Convert to tidy format

- mspms_tidy()

Statistics

Calculate log2 fold change and perform T-tests

- log2fc_t_test()

Visualization

Quality control

- plot_nd_peptides()
- plot_qc_check()

PCA

- plot_volcano()

Heatmap

- plot_heatmap()

Volcano plot

- plot_volcano()

Time course

- plot_time_course()

Cleavage position plot

- plot_cleavage_pos()

iceLogo

- plot_icelogo()

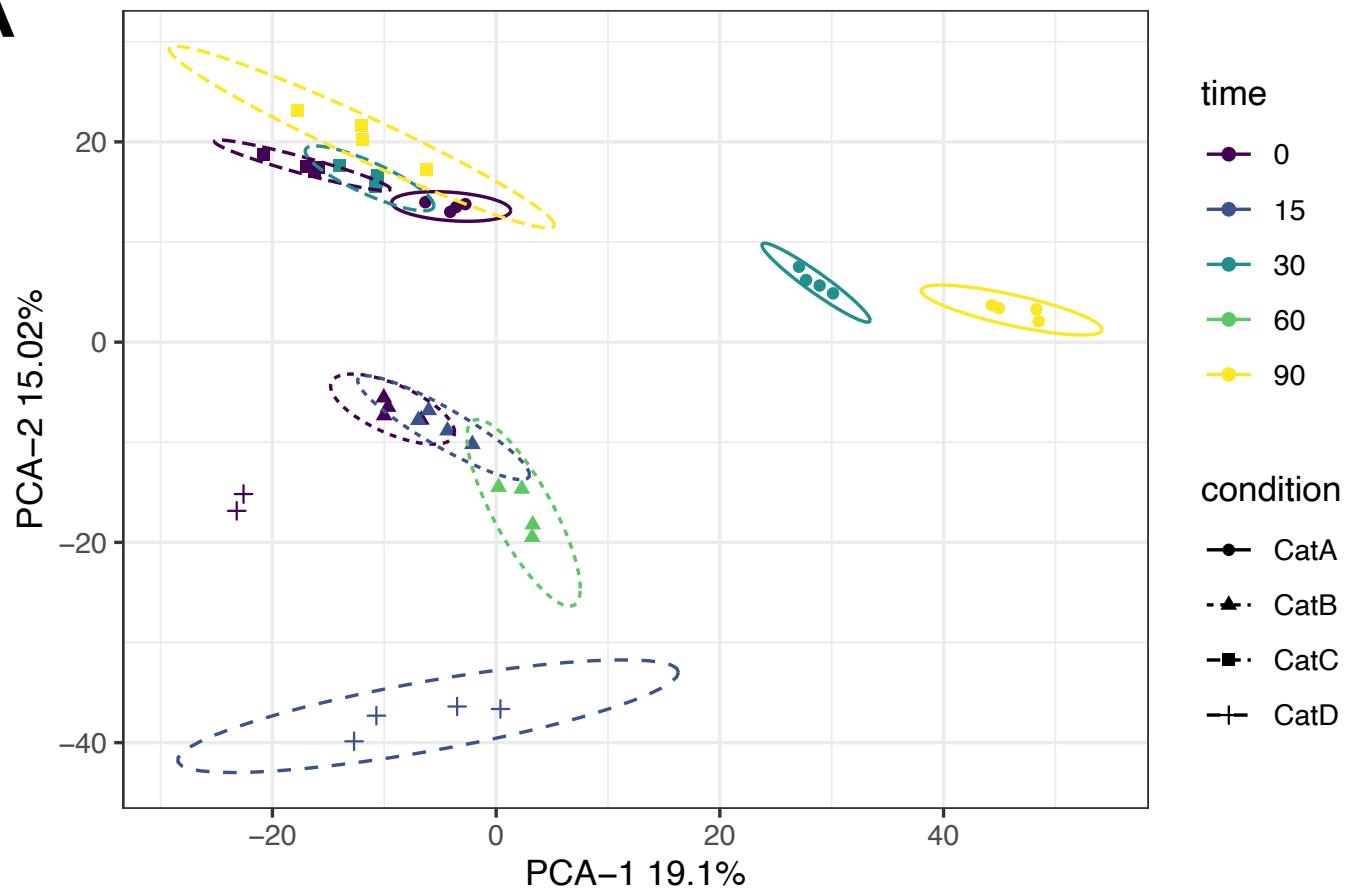
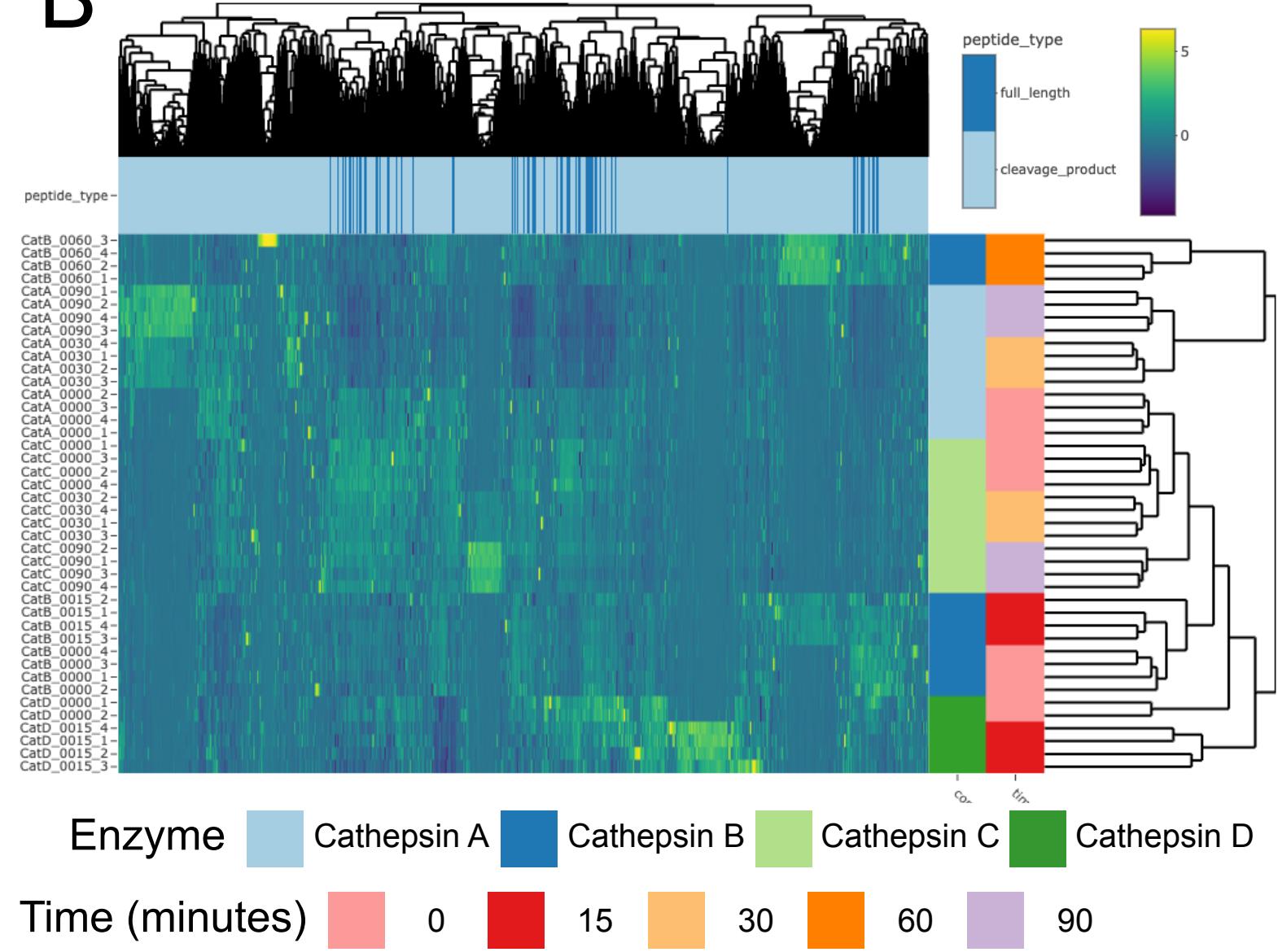
- plot_all_icelogos()

Report

- generate_report()

Only a subset of functions are exported to the user in an effort to make the package API intuitive.

Helper functions can be found in the helper_functions.R file name corresponding to the type of functions it assists with.

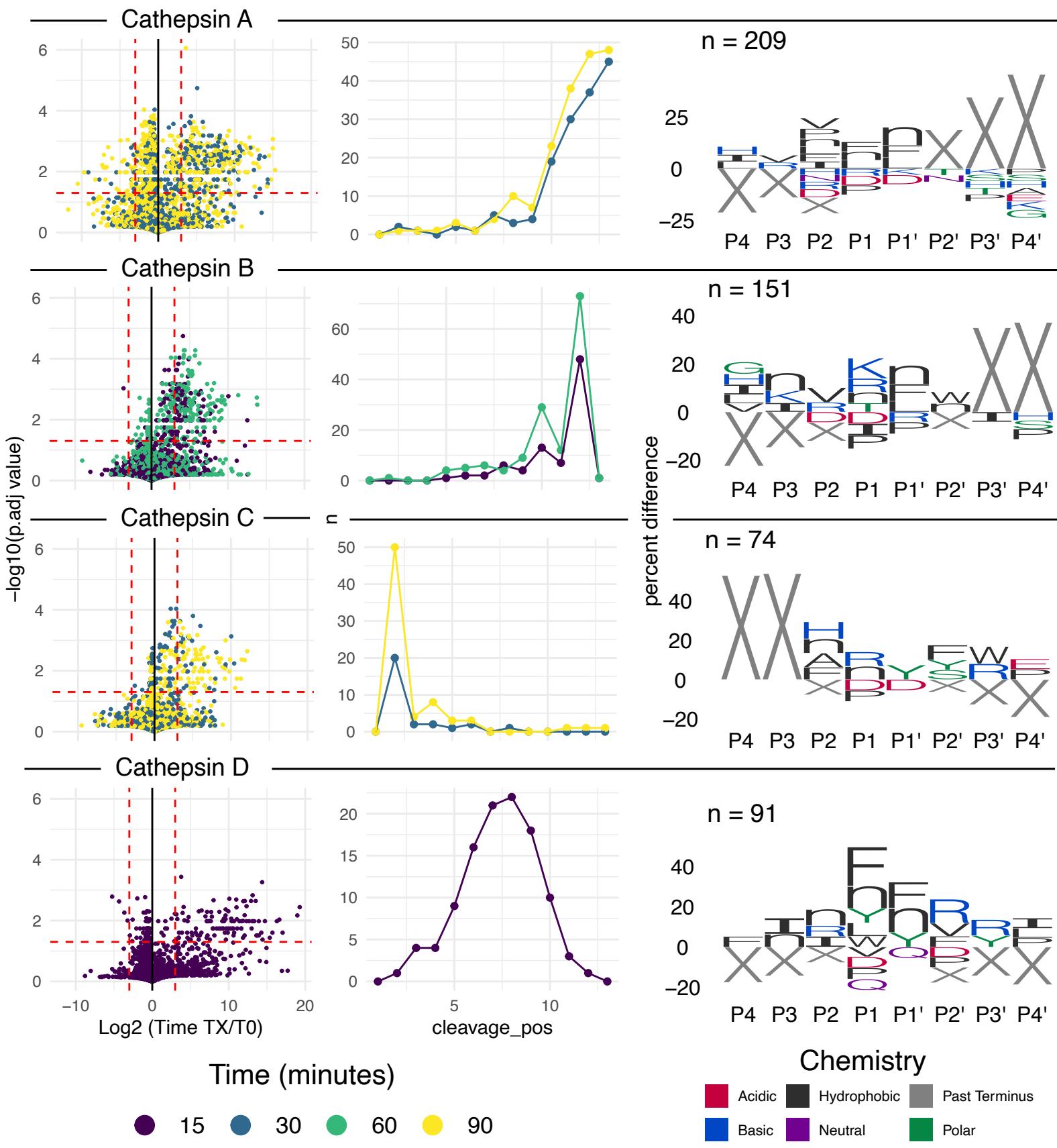
A**B**

Protease	Positional Specificity	Amino Acid Specificity	Ref
Cathepsin A	Carboxypeptidase	Removes hydrophobic amino acids (e.g. Phe and Leu) from the C-terminus of substrates when additional hydrophobic amino acids are present in P1	²⁵
Cathepsin B	Dipeptidyl carboxypeptidase	Removes dipeptides from the C-terminus when either positively charged (Arg, Lys) or hydrophobic amino acids are in the P1 and P2 position	²⁶
Cathepsin C	Dipeptidyl aminopeptidase	Removes dipeptides from the N-terminus with broad specificity	⁸
Cathepsin D	Endopeptidase	Cleaves between hydrophobic amino acids (e.g. Phe, Leu, Tyr)	²⁷

B

C

D





Click here to access/download
Supplementary Material
S1.pdf



Click here to access/download
Supplementary Material
S2.pdf





Click here to access/download
Supplementary Material
S3.pdf





Click here to access/download
Supplementary Material
S4.pdf



Click here to access/download
Supplementary Material
S5.pdf



Click here to access/download
Supplementary Material
S6.pdf



Click here to access/download
Supplementary Material
S7.pdf





Click here to access/download
Supplementary Material
SF1-peptide_library.fasta



Click here to access/download
Supplementary Material
SF2-peptide_library.csv



Click here to access/download
Supplementary Material
SF3.png



Click here to access/download
Supplementary Material
SF4-mspms.pdConsensusWF



Click here to access/download
Supplementary Material
SF5-mspms.pdProcessingWF



Click here to access/download
Supplementary Material
SF6-LFQ-MBR(MSP-MS).workflow



Click here to access/download
Supplementary Material

[**SF7-generic_mspms_analysis_of_cathepsin_A-D.html**](#)



UNIVERSITY of CALIFORNIA, SAN DIEGO

SCHOOL OF MEDICINE

Charlie Bayne

Biomedical Science Graduate Student
chnayne@health.ucsd.edu

6 December 2024

Dear Editorial Board,

We are pleased to resubmit our manuscript, "***mspms: A Comprehensive R Package and Graphic Interface for Multiplex Substrate Profiling by Mass Spectrometry Analysis***," for consideration in *GigaScience*.

Since our initial submission, the *mspms* software has successfully completed peer review at Bioconductor and is now an accepted part of its ecosystem. This process led to significant enhancements, including a more intuitive user interface and seamless integration with Bioconductor's extensive analytical tools.

In response to the editorial feedback, we have substantially revised the manuscript. These revisions include a more clearly articulated rationale for the work, an improved logical flow, and a restructured methods section to ensure clarity and accessibility. We believe these updates have strengthened the manuscript significantly.

We have also revitalized the *mspms* graphical interface. Key improvements include enhanced error-handling, loading indicators, refined graphics, and new features such as user-defined thresholds, publication-quality plot downloads, and interactive feature selection.

The core message of our manuscript remains unchanged: we present *mspms*, a novel, open-source R package now available through Bioconductor, designed to streamline and democratize the analysis of multiplex substrate profiling by mass spectrometry (MSP-MS) data. Benchmarking against established ground-truth data for cathepsins A-D demonstrates that *mspms* reliably infers substrate specificities across diverse proteases. With its robust functionality and user-friendly design, *mspms* is poised to become a standard tool in protease research, significantly advancing the study of proteolytic enzymes and their roles in health and disease.

We chose to resubmit this manuscript to *GigaScience* because we believe it is the ideal venue for our work, given the journal's emphasis on high-quality research that prioritizes reproducibility, usability, and practical utility. Additionally, the open-access model and global reach of *GigaScience* align with our mission to make the *mspms* package broadly accessible to the research community.

We have no conflicts of interest to declare. All authors have approved the manuscript for submission, and the content has not been published or submitted for publication elsewhere.

Thank you for considering our manuscript for publication. We look forward to the opportunity to share our research with the scientific community through *GigaScience*.

A handwritten signature in black ink that reads "Charlie Bayne".

Charlie Bayne

DEPARTMENT OF PHARMACOLOGY
9500 GILMAN DRIVE, LA JOLLA, CALIFORNIA 92093-0636 Telephone (858) 822-1218 Fax (858) 822-0041

Date: 11 Dec 2024
To: "Charlie Bayne" chbayne@health.ucsd.edu; baynec2@gmail.com
From: "GigaScience Editorial Office" editorial@gigasciencejournal.com
Subject: Decision on your submission to GigaScience - GIGA-D-24-00557

GIGA-D-24-00557

mspms: A Comprehensive R Package and Graphic Interface for Multiplex Substrate Profiling by Mass Spectrometry Analysis
Charlie Bayne; Brianna Hurysz; David J. Gonzalez; Anthony O'Donoghue
GigaScience

Dear Mr Bayne,

Thank you for submitting your manuscript, "mspms: A Comprehensive R Package and Graphic Interface for Multiplex Substrate Profiling by Mass Spectrometry Analysis" (GIGA-D-24-00557) to GigaScience.

The editors have assessed your manuscript and regret to inform you that it cannot be accepted for publication in GigaScience.

As you may know, we decline a substantial proportion of manuscripts without sending them to referees, so that they may be sent elsewhere without delay. Our editorial judgements are based on such considerations as the relevance to journal scope, the degree of advance provided, the breadth of potential interest to researchers, reproducibility, openness and timeliness.

We appreciate the work you have presented; however, we have concerns with the novelty of the pipeline/algorithms in your research. We feel it is more suitable for our GigaByte journal (<https://gigabytejournal.com/>), that focuses on software reports that have more limited conceptual narrative than GigaScience. Community use and reuse of data and tools being the primary aim, crediting still useful research objects that might be otherwise ignored, but that are of value to more specialist communities. This work would be a great fit for this type of article, so if you would like us to consider this manuscript for GigaByte instead, please let us know and you can submit your paper following our link here: <https://gigabyte-review.rivervalleytechnologies.com/>.

GigaByte is the winner of the 2022 ALPSP Innovation in Publishing Award and you can read more on the journal here:
<https://gigabytejournal.com/information-for-authors>

I wish you every success with your research.

Best wishes,

Qing Lan
GigaScience

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: <https://www.editorialmanager.com/giga/login.asp?a=r>). Please contact the publication office if you have any questions.