# Data Descriptor Template

**Scope Guidelines**

**Data Descriptors** submitted to *Scientific Data* should provide detailed descriptions of valuable research datasets, including the methods used to collect the data and technical analyses supporting the quality of the measurements. Data Descriptors focus on helping others reuse data, rather than testing hypotheses, or presenting new interpretations, methods or in-depth analyses. Relevant datasets must be deposited in an appropriate public repository prior to Data Descriptor submission, and their completeness will be considered during editorial evaluation and peer review. The data must be made publicly available without restriction in the event that the Data Descriptor is accepted for publication (excepting reasonable controls related to human privacy issues or public safety).

### Title

Mass spectrometry peptidomics data from porcine wounds infected with *Staphylococcus aureus* and *Pseudomonas aeruginosa*

### Authors

Fredrik Forsberg1, Jitka Petrlova, Sven Kjellström, Artur Schmidtchen1, Johan Malmström1, Erik Hartman1

**Affiliations**

1. Lund University

corresponding author: Fredrik Forsberg (fredrik.forsberg@med.lu.se)

### Abstract

Recently, mass spectrometry based peptidomics studies have proven useful in the identification of biomarkers and bioactive peptide-based therapeutics. Here, we present a dataset consistent of temporal wound fluid peptidomics data from highly defined porcine models. Wound fluids from porcine wounds infected with *Staphylococcus aureus* and *Pseudomonas aeruginosa* were sampled at different timepoints of the infection. Peptides were extracted from the samples, followed by liquid chromatography tandem mass spectrometry analysis in data dependent acquisition mode. The resulting spectra are deposited, allowing for database searching and subsequent peptidomic analysis of the infected wound fluid peptidome.

### Background & Summary

Peptides are short sequences of amino acids which are naturally produced in organisms, both through the translation of mRNA but importantly also through the degradation of proteins. These peptides play important roles in different biological systems, such as being messengers in signalling pathways [<https://www.sciencedirect.com/topics/neuroscience/peptide-hormone>] and by providing antimicrobial properties during infection [<https://www.nature.com/articles/s41467-018-05242-0>, <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2020.582779/full>, <https://www.nature.com/articles/s41573-019-0058-8>]). Peptidomics is the study of large sets of peptides from biological samples and has proven to be pivotal in the characterization of peptides and protein degradation under different physiological conditions as well as in the search for therapeutic peptides. Identification and quantification of peptides is necessary to conduct peptidomic analyses. Commonly, liquid chromatographic (LC) separation followed by mass spectrometry (MS) analysis is used for this purpose [<https://link.springer.com/protocol/10.1007/978-1-0716-3646-6_1>].

During a wound infection, a combination of host and pathogen derived proteases create an environment with increased proteolytic activity. These proteases are utilized by the host to remodel tissue and fight the invasion of bacteria. It has also been speculated that similar mechanisms are utilized by the bacteria, to promote colonization [<https://onlinelibrary.wiley.com/doi/10.1046/j.1524-475X.1999.00433.x>, <https://www.magonlinelibrary.com/doi/full/10.12968/jowc.2022.31.4.352>, <https://www.mcponline.org/article/S1535-9476(20)33742-7/fulltext>]. By investigating the resulting peptidome during different wound infection conditions, further understanding of this environment and mechanism could be gained.

Chronic wounds pose a large economic burden on society and life quality burden for patients. The prevalence of these wounds is likely to increase with an aging population and more life-style related diseases [<https://pubmed.ncbi.nlm.nih.gov/28118847>]. Further, pathogens such as *Staphylococcus aureus* and *Pseudomonas aeruginosa* are two of the most prevalent bacteria in infected wounds [<https://www.mdpi.com/2079-6382/10/10/1162>]. They are considered of high and critical priority respectively by the World Health Organization due to their developed resistance against current antibiotics [<https://www.who.int/publications/i/item/WHO-EMP-IAU-2017.12>]. Therefore, identifying novel means of diagnosing and treating such wounds is of extreme importance.

This study was carried out to identify differences in the peptidomic landscape of wound fluids depending on the presence and type of pathogen, while also employing a newly developed analysis algorithm with the potential to remove large amounts of redundancy in peptidomic datasets [<https://www.nature.com/articles/s41467-024-51589-y>]. This was done by generating LC-MS/MS data from wound fluids derived from highly defined infected porcine wounds (Fig. 1). Here, we present descriptions and access to the datasets. We believe that the vastness of the peptidomic data yields many opportunities to analyze it using different methods to gain novel insights about the mechanisms underlying protein degradation in infected wounds.

### Methods

**Study design**

The data generated in this study was utilized to analyze the peptidomic landscape of infected wounds in Hartman et al. (2024). Briefly, wounds were generated on Göttingen minipigs, whereafter they were infected with either *S. aureus* or *P. aeruginosa*, or not infected and used as control. Wound dressings were placed on the wounds, which absorbed the wound fluid generated in the wound healing process (Fig. 1a). The dressings were collected every 24 hours for 2-3 days depending on the sample type. Four of the wounds infected by *S. aureus* on day 1 were infected with *P. aeruginosa* on day 2, creating a double infection. Additionally, some of the wounds infected by *S. aureus* were contaminated by *P. aeruginosa* on the infection day, creating a set of samples which undergo an accidental double infection (Fig. 1b).

The proteomic content of the samples was discarded by filtration (Fig. 1c), and the subsequent peptidomic sample was analyzed with liquid chromatography tandem mass spectrometry (LC-MS/MS) in data dependent acquisition mode on an Evosep One LC (Evosep, Denmark) coupled to a timsTOF Pro MS (Bruker, USA). The data was searched with PEAKS X and deposited to ProteomeXchange (Fig. 1d). A stratified blinded random subset of samples from day 1, containing 4 samples from each single infection and control condition, were re-analyzed. The re-analysis was conducted 8 months after the original samples using the same conditions except for that the MS had been changed to a timsTOF HT (Bruker, USA) (Fig 1b).

**Pig acute wound fluid collection**

Partial thickness wounds were induced to Göttingen minipigs, which were then either infected with *S. aureus* or *P.aeruginosa*, or kept uninfected as control samples. Polyurethane dressings were used to dress the wounds and were changed after 24 and 48 hours. The old dressings were soaked in 10 mM Tris, and the fluids were extracted using a syringe. A subset of the *S. aureus* infected wounds was accidentally cross contaminated with *P. aeruginosa*. Another set of the *S. aureus* infected wounds were infected with *P. aeruginosa* after 24 hours, and these dressings were also collected and extracted 72 hours after the initial *S. aureus* infection. Extracted fluids were then supplemented with Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific, USA) and kept at -80°C until further use.

**Wound fluid peptide extraction**

500 μg of protein (determined with Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, USA) as per provided instructions) was diluted to 100 μl with 10 mM Tris at pH 7.4. 300 μl of 8 M urea diluted in 10 mM Tris at pH 7.4, supplemented with 0.067% RapiGest SF (Waters, USA) was then added followed by an incubation at room temperature for 30 minutes. 100 μl 6 M urea in 10 mM Tris at pH 7.a was added to Microcon - 30 centrifugal filter units and centrifuged at 10000 RCF for 15 minutes at room temperature (RT). After this the wound fluid samples were added to the centrifugal filters, and centrifuged at 10000 RCF for 30 minutes at RT. Finally, another 100 μl of 6 M urea in 10 mM Tris at pH 7.4 was centrifuged through the filter units at 10000 RCF for 5 minutes at RT and the filtrate was stored at -20°C.

**Acidification and solid phase extraction**

1 μl 100% formic acid (FA) was added to 60 μl of each extracted peptide sample. 100 μl 100% acetonitrile (ACN) + 0.1% FA was added to UltraMicro Spin Columns (The Nest Group, USA) which were centrifuged at 800 RCF for 1 minute at room temperature. All further centrifugation steps in this section were performed this way. 100 μl 2% ACN + 0.1% trifluoroacetic acid (TFA) was centrifuged through the columns, twice, before adding the samples and performing an additional centrifugal step. Lastly, 100 μl 70% ACN + 0.1% TFA was centrifuged through the columns to elute the sample which was then dried in an Eppendorf Concentrator plus (Eppendorf, Germany).

**LC-MS/MS**

The dried peptide samples were then dissolved in 30 μl 2% ACN + 0.1% FA, before being loaded onto Evotip Pure columns (Evosep, Denmark) according to the provided instructions, apart from that the samples were not dissolved in 20 µl 0.1% FA before loading. The samples were analyzed by LC/MS-MS on an Evosep One LC (Evosep, Denmark) coupled with a timsTOF Pro mass spectrometer (Bruker, USA). The LC used a EV1137 Performance Column - 15 cm x 150 µm, with 1.5 µm ReproSil-Pur C18 beads (Evosep, Denmark). The accompanying 30 samples per day program was used for separation. The MS used the DDA PASEF mode, with 10 PASEF scans every acquisition cycle. Accumulation and ramp times were set to 100 ms, precursors with a +1 charge were ignored, and target intensity was set to 20000, with dynamic exclusion active, at 0.4 min. Isolation width was set to 2 at 700 Th and 3 at 800 Th.

**Database search**

The data from the LC-MS/MS runs were searched with PEAKS X. UniProtKB reviewed (Swiss-Prot) protein list of pig (organism\_id:9823) proteins was used as a database, but with fibrinogen alpha chain (FIBA\_PIG) and fibrinogen beta chain (FIBB\_PIG) being changed to the UniProt KB unreviewed (TrEMBL) versions F1RX36\_PIG and F1RX37\_PIG. The list was downloaded May 11th, 2023. Data refinement was set to merge scans and correct precursor based on mass and charge states with charges between 1 and 4. It was also set to associate features between 2 and 8. Precursor tolerance was set to 20.0 ppm using monoisotopic mass and fragment tolerance was set to 0.03 Da. 1 modification per peptide was allowed with oxidation (M, +15.99) being the only possible modification. Results were filtered at 1% FDR with ≥1 unique peptide for each protein. FDR was set to be estimated with decoy-function.

### Data Records

Both the raw mass spectrometry data (.d folders generated by Bruker Compass) as well as the database search of the data (.mgf and .mzid.gz generated by PEAKS X) have been uploaded to ProteomeXchange as a part of the public dataset PXD048892 along with a design file for sample identification (.csv) (<https://dx.doi.org/10.6019/PXD048892>). Additionally, similar files from the blinded re-run of samples have been uploaded to ProteomeXchange under the identifier PXD055074 (<https://dx.doi.org/10.6019/PXD055074>).

### Technical Validation

To get an understanding of the data, general characteristics of the different groups were summarized (Fig. 2). Unique peptide overlap was summarized, with more unique peptides appearing in the *S. aureus* (2115)and *P. aeruginosa* (4762) groups compared to the control group (370), while the number of peptides shared by all samples was 1774 (Fig 2.a). The number of identified peptides decreases over time (Fig 2.b). The mean length distribution weighted by the log2 intensities were similar for all sample types (Fig 2.c). The log2 intensities were scaled to a mean of 0 and a unit variance. Thereafter, the dimensionality of the data was reduced using Uniform Manifold Approximation Projection (UMAP). Default settings were used as per the umap-learn python package (Fig 2.d). The data cluster based on infection type and sampling day, suggesting that there are differences in the peptidomes depending on infection type and time point.

To validate that our peptide extraction and mass spectrometry methods are reproducible, 4 wound fluid samples from day 1 and each of the groups *S. aureus*, *P. aeruginosa*, and control were randomly selected and had their peptides extracted again (Fig. 1b). Once extracted, the samples were blinded and analyzed by LC-MS/MS as the previous samples, but using a timsTOF HT (Bruker, USA) MS. It was seen that the unique peptides identified in each sample group corresponded well with peptides previously seen in the group, with fewer observed unique peptides in the blind group due to missing values (Fig 3.a, b, c), and that unique peptide overlap is similar for blinded samples compared to their respective group (Fig 3.d). It can also be seen that the peptide length distribution is retained (Fig 3.e), and that the samples cluster together with the other samples from the same group and day when reducing dimensionality using UMAP (Fig 3.f), showing that the method is robust and replicable.

### Usage Notes

The data was supplied as both raw output files and result files searched with PEAKS X. The raw files can be used for searching the data with different softwares or parameters than the one presented in this study. There are many ways to analyze the data in the resulting result files. MS intensities follow a lognormal distribution, so for comparative analyses the log of the intensities is computed prior to assuming normal distributions. The data contains missing values, and one can deal with these in different ways, e.g. through imputation. Further, to remove technical bias effects, normalization of the peptide intensities is commonly applied.

### Code Availability

No custom code was used to generate the data.

### Acknowledgements

We thank the Swedish National Infrastructure for Biological Mass Spectrometry (BioMS) for performing the LC-MS/MS analysis. We acknowledge support by grants from the Swedish Research Council (projects 2017-02341, 2018-05916 and 2020-02016 (A.S.), and 2023-02107 (J.M.)), Edvard Welanders Stiftelse and Finsenstiftelsen (Hudfonden) (A.S.) the Royal Physiographic Society (A.S.), the Österlund Foundation (A.S.), and the Swedish Government Funds for Clinical Research (ALF) (A.S.).

### Author contributions

FF and EH wrote the manuscript. FF, SK and JP performed the sample preparation. FF and SK performed the mass spectrometry analysis. AS and JM supervised the project.

### Competing interests

### The authors declare no competing interests.

### Figures

### Figure Legends

### Tables

### References

### Additional Formatting Information

**Referencing Figures, Tables, and other content**

**The Word document may reference Figures (e.g. Fig. 1), Tables (e.g. Table 1), and Supplementary Information (e.g. Supplementary Table 1, or Supplementary File 2, etc.).**

**Citation format**

All references should be numbered sequentially, first throughout the text, then in tables, followed by figures and, finally, boxes; that is, references that only appear in tables, figures or boxes should be last in the reference list. Only one publication is given for each number. Only papers that have been published or accepted by a named publication or recognized preprint server should be in the numbered list; preprints of accepted papers in the reference list should be submitted with the manuscript. Published conference abstracts, numbered patents, and archived code with an assigned DOI may be included in the reference list. Grant details and acknowledgments are not permitted as numbered references. Footnotes are not used.

Scientific Data uses standard Nature referencing style. All authors should be included in reference lists unless there are six or more, in which case only the first author should be given, followed by ‘et al.’. Authors should be listed last name first, followed by a comma and initials (followed by full stops, '.') of given names. Article titles should be in Roman text; only the first word of the title should have an initial capital and the title should be written exactly as it appears in the work cited, ending with a full stop. Book titles should be given in italics and all words in the title should have initial capitals. Journal names are italicized and abbreviated (with full stops) according to common usage. Volume numbers and the subsequent comma appear in bold. The full page range should be given where appropriate. See the examples below:

**Journal Article**:

1. Schott, D. H., Collins, R. N. & Bretscher, A. Secretory vesicle transport velocity in living cells depends on the myosin V lever arm length. *J. Cell Biol*. **156**, 35‐39 (2002).

**Book** ‐ Book titles should be given in italics and all words in the title should have initial capitals:

1. Hogan, B. *Manipulating The Mouse Embryo: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, 1994)

**Publicly available preprint:**

1. Babichev, S. A., Ries, J. & Lvovsky, A. I. Quantum scissors: teleportation of single-mode optical states by means of nonlocal single photon. Preprint at http://arXiv.org/quant-ph/0208066 (2002).

**Code:**

1. Gallotti, R. & Barthélemy, M. Source code for: The multilayer temporal network of public transport in Great Britain. *Figshare* https://dx.doi.org/10.6084/m9.figshare.1249862.v1 (2014).

**Online material** ‐ Stable documents hosted on the web may be cited in the main reference list, using the format below. Websites or dynamic web resources should be cited by embedding the URL in the main article text:

1. Manaster, J. Sloth squeak. *Scientific American Blog Network* http://blogs.scientificamerican.com/psi-vid/2014/04/09/sloth-squeak (2014).

**Technical or government report:**

1. Akutsu, T. *Total Heart Replacement Device.* Report No. NIH-NHLI-69 2185-4 (National Institutes of Health, 1974).

## Citing Data

In line with emerging [industry-wide standards for data citation](https://www.nature.com/articles/sdata2018259), references to all datasets described or used in the manuscript should be cited in the text with a superscript number and listed in the ‘References’ section in the same manner as a conventional literature reference.

An author list (formatted as above) and title for the dataset should be included in the data citation, and should reflect the author(s) and dataset title recorded at the repository. If author or title is not recorded by the repository, these should not be included in the data citation. The name of the data-hosting repository, URL to the dataset and year the data were made available are required for all data citations. For DOI-based (e.g. figshare or Dryad) repositories the DOI URL should be used. For repositories using accessions (e.g. SRA or GEO) an [identifiers.org](https://identifiers.org/) URL should be used where available. For first submissions, authors may choose to include just the accession number. Scientific Data staff will provide further guidance after peer-review. Please refer to the following examples of data citation for guidance:

1. Zhang, Q-L., Chen, J-Y., Lin, L-B., Wang, F., Guo, J., Deng, X-Y. Characterization of ladybird Henosepilachna vigintioctopunctata transcriptomes across various life stages. figshare <https://doi.org/10.6084/m9.figshare.c.4064768.v3> (2018).
2. NCBI Sequence Read Archive <http://identifiers.org/ncbi/insdc.sra:SRP121625> (2017).
3. Barbosa, P., Usie, A. and Ramos, A. M. Quercus suber isolate HL8, whole genome shotgun sequencing project. GenBank<http://identifiers.org/ncbi/insdc:PKMF00000000> (2018).
4. DNA Data Bank of Japan <http://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA004814> (2016).

**Depositing your data to an appropriate repository**

Your *Scientific Data* manuscript will not be sent to review unless the dataset(s) described therein have been deposited in an appropriate public repository ([please see our policy on this](http://www.nature.com/sdata/policies/repositories)). Should a subject specific repository not be available for your field or data-type, or should the repository of your choice not permit confidential peer-review, please use a [generalist repository](https://www.nature.com/sdata/policies/repositories#general). Integrated submission systems are available for both figshare and Dryad.