# 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

Peptidomic data from *Staphylococcus aureus* and *Pseudomonas aeruginosa* infected porcine wound fluids

### 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

Wound fluids from pig wounds infected with *Staphylococcus aureus* and *Pseudomonas aeruginosa* were collected at different timepoints of the infection. The wounds fluids had their peptides extracted and then analyzed by liquid chromatography with tandem mass spectrometry. The collected data could be searched for interesting peptides relevant to infection and immune response based on the infecting bacteria.

### Background & Summary

Peptidomics is the study of all peptides found in biological samples. The data for these studies are most commonly derived from a liquid chromatographic (LC) separation followed by analysis using mass spectrometry (MS) [<https://link.springer.com/protocol/10.1007/978-1-0716-3646-6_1>]. As peptides can play a vast number of roles in biological systems (such as being messengers [<https://www.sciencedirect.com/topics/neuroscience/peptide-hormone>] and having antimicrobial properties [<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>]) there is a huge amount of information that can be explored in peptidomic datasets.

During wound infection, a mix of host and pathogen derived proteases create an environment with increased proteolytic activity, utilized by both the host and the pathogen [<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 looking at the resulting peptidome during different wound infection conditions, further understanding of this environment could be gained.

As chronic wounds have a large economic burden on society and life quality burden for patients, which are likely to increase with an aging population and more life-style related diseases [<https://pubmed.ncbi.nlm.nih.gov/28118847>], aiding the progress of understanding and treating is of high value. Another concern that needs addressing is pathogens such as *Staphylococcus aureus* and *Pseudomonas aeruginosa* being two of the most prevalent bacteria in wound infection [<https://www.mdpi.com/2079-6382/10/10/1162>], and being ranked as being 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>].

This study was carried out to identify differences in the peptidomic landscape of wound fluids depending on the presence and type of pathogenic bacteria, while also employing a newly developed analysis algorithm with the potential to remove large amounts of redundancy in peptidomic datasets [cite clustering paper]. This was done by generating MS data from wound fluids derived from infected pig wounds (figure 1 [fig 1.a in clustering paper]). Analyzing the differences between the sample types in this dataset could potentially give insight into pathogen-specific infection mechanisms and immune responses.

### Methods

**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 [cite pig model paper]. 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.

**Human chronic wound fluid collection**

Patients suffering from chronic venous wounds had their wounds dressed with polyurethane dressings for 48 to 72 hours before being collected, extracted and stored the same way as the pig samples.

**Wound fluid peptide extraction**

500 μg of protein for pig wound fluids, and 100 μg of protein for human wound fluids (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. Human samples were searched using the UniProt reviewed (Swiss-prot) human protein list (organism\_id:9606) (downloaded September 29th, 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 as well as the database search of the data have been uploaded to ProteomeXchange as a part of the public dataset PXD048892 along with a design file for sample identification (<https://dx.doi.org/10.6019/PXD048892>).

### Technical Validation

To get an understanding of the mass spec data, general characteristics of the different groups were summarized (figure 2). Peptide length distribution pattern was similar for all sample types (fig 2.a). To look for inherent patterns within the dataset, the data was logarithmized and standardized before being dimensionally reduced using UMAP (fig 2.b). It can be seen that the data cluster based on infection type and sampling day.

[Peptide length distribution in samples]

[UMAP of present peptides in samples]

To validate that our peptide extraction and mass spectrometry methods are reproducible, 4 day-1 wound fluid samples from each of the groups *S. aureus*, *P. aeruginosa*, and control were randomly selected and had their peptides extracted again. Once extracted, the samples were blinded and analyzed by LC-MS/MS as the previous samples, but with an [add new column name] equipped. These samples were then analyzed the same way as the previous samples for comparison (figure 2.x)

### Usage Notes

### Code Availability

To generate figures [Figure numbers length distribution and UMAP], python scripts which employ the matplotlib.pyplot package were utilized.

### Acknowledgements

### Author contributions

### 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.