



## Data Article

# Proteomics dataset from 26th Dynasty Egyptian mummified remains sampled using minimally invasive skin sampling tape strips

Dylan H. Multari<sup>a</sup>, Prathiba Ravishankar<sup>a</sup>, Geraldine J. Sullivan<sup>a</sup>,  
Ronika K. Power<sup>b,c,d</sup>, Constance Lord<sup>e</sup>, James A. Fraser<sup>f</sup>,  
Paul A. Haynes<sup>a,c,d,\*</sup>

<sup>a</sup> School of Natural Sciences, Macquarie University<sup>1</sup>, North Ryde, NSW 2109, Australia

<sup>b</sup> Department of History and Archaeology, Macquarie University, North Ryde, NSW 2109, Australia

<sup>c</sup> Biomolecular Discovery Research Centre, Macquarie University, North Ryde, NSW 2109, Australia

<sup>d</sup> Centre for Ancient Cultural Heritage and Environment, Macquarie University, North Ryde, NSW 2109, Australia

<sup>e</sup> Chau Chak Wing Museum, University of Sydney, Camperdown, NSW 2006, Australia

<sup>f</sup> Middle East Department, British Museum, United Kingdom

## ARTICLE INFO

## Article history:

Received 23 June 2022

Revised 1 August 2022

Accepted 25 August 2022

Available online 31 August 2022

## Dataset link:

[DataInBrief\\_Deamidation \(Original data\)](#)

## Keywords:

Paleoproteomics

Deamidation

Bioarchaeology

Cultural heritage

Mass spectrometry

## ABSTRACT

Paleoproteomics typically involves the destructive sampling of precious bioarchaeological materials. This analysis aims to investigate the proteins identifiable via nanoLC-MS/MS from highly degraded 26th Dynasty Egyptian mummified human remains (NMR.29.1-8) after non-destructive sampling with commercially available dermatology-grade skin sampling tape strips. A collection of cranial and other bone fragments were sampled with the tape strips then subsequently analysed using a shotgun proteomics approach. The number of proteins identified using this method ranged from 18 to 437 at a peptide FDR of <1%. Deamidation ratios were assessed using an in-house R script, with asparagine deamidation averaging ~20–30% and glutamine deamidation averaging ~15–25%.

DOI of original article: [10.1016/j.jas.2022.105548](https://doi.org/10.1016/j.jas.2022.105548)

\* Corresponding author at: School of Natural Sciences, Macquarie University, North Ryde, NSW 2109, Australia.

E-mail address: [paul.haynes@mq.edu.au](mailto:paul.haynes@mq.edu.au) (P.A. Haynes).

Social media: [@DylanMultari](#) (D.H. Multari), [@HaynesLabMQ](#) (P.A. Haynes)

<sup>1</sup> [@MQNatSci](#)

<https://doi.org/10.1016/j.dib.2022.108562>

2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

© 2022 The Author(s). Published by Elsevier Inc.  
 This is an open access article under the CC BY-NC-ND  
 license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Specifications Table

Subject	Omics: Proteomics
Specific subject area	Modern human skin surrogate and ancient human bone surfaces sampled with dermatology-grade skin sampling strips.
Type of data	Table, Figure, Graph
How the data were acquired	nanoLC-MS/MS, bottom-up proteomics. Thermo Easy-nLC1000 system coupled to a Thermo Q-Exactive orbitrap mass spectrometer (Thermo Scientific, San Jose, CA). HALO C18 160 Å, 2.7 µm bead size, 75 µm internal diameter x 75 mm length column. A 60 min reversed-phase chromatographic linear gradient of 1–50% solvent A (0.1% formic acid) to solvent B (99.9% acetonitrile/0.1% formic acid). X! Tandem search algorithm running under Global Proteome Machine (version 3.0) software searching the SwissProt Human Protein database (20,329 proteins).
Data format	Raw, Analysed, Filtered
Description of data collection	A modern human skin surrogate, and the interior and exterior surfaces of ancient human bone fragments were sampled with D-Squame skin sampling tape strips. Sampling was performed with 3 strips for each bone fragment surface except humeral shaft and modern skin which were sampled with a total of 9 strips (later divided into groups of 1, 3, and 5).
Data source location	Institution: Macquarie University City/Town/Region: Sydney Country: Australia
Data accessibility	Data is accessible from the ProteomeXchange Consortium via the PRIDE partner repository (dataset identifier PXD029003). Additional data is available with the article. <b>Project Webpage:</b> <a href="http://www.ebi.ac.uk/pride/archive/projects/PXD029003">http://www.ebi.ac.uk/pride/archive/projects/PXD029003</a> <b>FTP Download:</b> <a href="ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2022/04/PXD029003">ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2022/04/PXD029003</a> <b>GitHub Link:</b> The R script used in this study along with the collated GPM protein and peptide outputs are publicly available on GitHub using the following link: <a href="https://github.com/dymult/DataInBrief_Deamidation">https://github.com/dymult/DataInBrief_Deamidation</a>
Related research article	Multari, D.H., Ravishankar, P., Sullivan, G.J., Power, R.K., Lord, C., Fraser, J.A., Haynes, P.A., 2022. Development of a novel minimally invasive sampling and analysis technique using skin sampling tape strips for bioarchaeological proteomics, <i>Journal of Archaeological Science</i> 139, 105548. <a href="https://doi.org/10.1016/j.jas.2022.105548">10.1016/j.jas.2022.105548</a> .

## Value of the Data

- This dataset highlights the types of proteins that can be identified using minimally invasive sampling techniques on ancient human bone.
- The data may be of use to researchers working within paleoproteomics and those working in museums/cultural heritage organisations interested in new minimally invasive techniques.
- The data may be reused to compare with other paleoproteomics studies, in particular reference to the types of proteins identified, and degree of deamidation identified in these samples.

## 1. Data Description

This dataset represents data from a shotgun proteomics analysis of modern skin and ancient human bone samples, collected using minimally invasive skin sampling tape [1]. The raw mass spectrometric data, along with a collated Excel workbook of Global Proteome Machine (GPM) protein outputs (filename “S1 Protein Results All Samples.xlsx”) and peptide outputs (filename “S2 Peptide Results All Samples.xlsx”) filtered at a protein expectation score  $\log(e)$  of -3 have been deposited on the ProteomeXchange Consortium via the PRIDE repository [2] and are accessible using the identifier PXD029003. An Excel spreadsheet (filename “Explanation of raw files and samples.xlsx”) was also uploaded with the raw data to explain which series of raw files correlate to which sample.

The Excel spreadsheets of peptide and protein output results are also publicly available a GitHub repository, accessible via [https://github.com/dymult/DataInBrief\\_Deamidation](https://github.com/dymult/DataInBrief_Deamidation). This repository also contains the deamidation R script used to assess the deamidation ratios of the samples, as well as the input GPM peptide .xlsx files and the output .png files.

## 2. Experimental Design, Materials and Methods

### 2.1. Sampling

Modern surrogate samples were collected using a series of one, three, and five D-Square D100 Sampling tape strips (22 mm diameter; CuDerm Corporation, Texas, USA) from the forearm of a healthy consenting volunteer. Strips were gently applied to the sample surface with uniform pressure for a total of 10 s before removal and immediate storage in a fresh 1.5 mL microcentrifuge tube. Ancient samples were collected from four cranial fragments along with the mandible and humeral diaphysis (hereafter referred to as humeral shaft). Strips were applied in a similar fashion to the modern samples across the endo- and ectocranial surfaces of the four cranial fragments, the buccolabial and lingual surfaces of the mandible (for ease of reporting, these are hereafter referred to as the interior and exterior surfaces, respectively), and along the length of the humeral shaft. A total of three strips were used on each of the cranial fragment and mandible surfaces, while a total of nine strips were used on the humeral shaft. These nine strips were later split into three groups of one, three, and five strips, respectively, to mitigate issues with fitting nine strips into a single microcentrifuge tube.

### 2.2. Protein Extraction

All laboratory work was performed in accordance with established guidelines for ancient protein analysis [3]. Briefly, this involved the use of nitrile gloves, sterilised workspaces and equipment, freshly prepared reagents, mass spectrometric analysis of both analytical (1% formic acid) and procedural blanks (sampling tape strips removed from packaging and immediately transferred to 1.5 mL tube), and the use of a laminar flow hood to minimise potential exogenous contamination. Protein extraction from the sampling tape strips was performed using a modified protocol adapted from Clausen et al. [4]. Each set of strips was transferred to fresh 1.5 mL microcentrifuge tubes and covered with 1 mL of phosphate buffered saline (PBS) prior to bath sonication for 15 min at room temperature. Proteins were then precipitated with ice-cold acetone at  $-20^{\circ}\text{C}$  for 2 h and centrifuged at  $20,000 \times g$  for 10 min. The supernatant was discarded, and the pellet was briefly air-dried.

### 2.3. SDS-PAGE and Trypsin In-Gel Digestion

Protein pellets were resuspended in 50  $\mu\text{L}$  2x SDS loading buffer (100 mM Tris-HCl, 4% SDS, 0.2% bromophenol blue, 20% glycerol, 200 mM dithiothreitol), and heated to 95°C for exactly 5 min with agitation. SDS-PAGE separation was performed on a precast gel (Bio-Rad 10% Mini-PROTEAN® TGX™, 10  $\times$  50  $\mu\text{L}$  wells) at 100 V for 1 h. A protein marker standard (10  $\mu\text{L}$ ; Bio-Rad Precision Plus Unstained Marker) was also loaded on each gel. Proteins were visualised using colloidal Coomassie blue staining protocols, and sample lanes were excised and fractionated into eight equal fractions for subsequent in-gel digestion.

Gel fractions were finely chopped and transferred to fresh 1.5 mL microcentrifuge tubes, and re-equilibrated in 200  $\mu\text{L}$  of 100 mM  $\text{NH}_4\text{HCO}_3$  before three destaining washes of 200  $\mu\text{L}$  of 50% acetonitrile (ACN)/50 mM  $\text{NH}_4\text{HCO}_3$  for 10 min each, then dehydrated with a wash of 100% ACN. The ACN was removed and gel fractions were rehydrated and reduced in 50  $\mu\text{L}$  of a reducing solution consisting of 10 mM dithiothreitol (DTT) in 100 mM  $\text{NH}_4\text{HCO}_3$  for 1 h at room temperature. DTT solution was removed and replaced with an alkylating solution of 55 mM iodoacetamide in 100 mM  $\text{NH}_4\text{HCO}_3$  and incubated in the dark for 1 h at ambient temperature. Iodoacetamide solution was decanted, and gel pieces were re-equilibrated with a wash of 100  $\mu\text{L}$  of 100 mM  $\text{NH}_4\text{HCO}_3$  for 10 min, then two washes with 200  $\mu\text{L}$  of 50% ACN/50 mM  $\text{NH}_4\text{HCO}_3$ , prior to dehydration in 100% ACN. In-gel trypsin digestion was performed by rehydrating gel pieces in 30  $\mu\text{L}$  of trypsin solution (10 ng/ $\mu\text{L}$  Promega Sequencing Grade Modified Trypsin in 50 mM  $\text{NH}_4\text{HCO}_3$ ) for 30 min at 4°C, then incubating overnight at 37°C. Resultant peptides were extracted in 30  $\mu\text{L}$  of 50% ACN/2% formic acid (FA), followed by two more extractions with 70% ACN/2% FA, and 90% ACN/2% FA, respectively. Peptide extracts were vacuum centrifuged to dryness and reconstituted in 10  $\mu\text{L}$  of 1% FA for mass spectrometric analysis via nanoflow liquid chromatography – tandem mass spectrometry (nanoLC-MS/MS).

### 2.4. NanoLC-MS/MS of Extracted Peptides

A shotgun proteomics workflow was applied with minor modification, as previously described [1,5,6]. Briefly, peptides were analysed using a Thermo Q-Exactive orbitrap mass spectrometer coupled to a Thermo Easy-nLC1000 system (Thermo Scientific, San Jose, CA). A reversed-phase chromatographic separation was performed using a C18 HALO column (2.7  $\mu\text{m}$  bead size, 160 Å pore size; 75  $\mu\text{m}$  internal diameter  $\times$  75 mm length) and employed a 60 min linear gradient of 1–50% solvent A (0.1% FA) to solvent B (99.9% ACN/0.1% FA). Data-dependent acquisition mode was configured to automatically switch from Orbitrap MS to MS/MS mode. Peptide spectra were acquired within a  $m/z$  range of 350–1600 amu at a resolution of 35,000 with an isolation window of 3.0  $m/z$ . Higher energy collisional dissociation (HCD) fragmentation was performed at 30% normalised HCD collision energy on the top ten most abundant ions. Dynamic exclusion of target ions was set for 20 s, and fragment ions were analysed in the orbitrap at a resolution of 17,500. Analytical blanks were analysed between samples to minimise system carryover of abundant peptides.

### 2.5. Peptide-to-Spectrum Using X! Tandem Algorithm

Raw spectral files were converted to mzXML format using the MSConvert graphical user interface [7,8] for peptide-to-spectrum matching (PSM) using the X! Tandem algorithm operating under the GPM interface software (version 3.0, <https://www.thegpm.org/>) [9,10]. Files were searched against the curated SwissProt Human Protein database (downloaded January 2019; 20,329 proteins) and the Common Repository of Adventitious Proteins (cRAP) database containing sequence data for common laboratory contaminants and protein standards [11]. PSM was performed with the following specifications: Orbitrap method including  $\pm 20$  ppm parent ion

mass tolerance,  $\pm 0.1$  Da fragment ion mass tolerance, trypsin as digestion enzyme allowing for a maximum of two missed cleavages, peptide search length of between 6 and 50 amino acids with allowance for up to 4+ charge states, carbamidomethylation of cysteine/selenocysteine as a fixed modification, and oxidation of methionine and tryptophan, hydroxylation of proline, and deamidation of asparagine and glutamine as potential variable modifications. PSM was also conducted against a reversed database for assessment of false discovery rates (FDR). Each of the eight sample fractions were processed sequentially then merged to generate a collated non-redundant sample output file. The 'HumeralShaft\_total' data represents all 24 gel fractions compiled into a single GPM search. GPM protein and peptide output files were exported from the graphical user interface as .xlsx files. Data were manually filtered at a protein expectation score  $\log(e)$  cut-off value of -3, including retention of high-quality single peptide-based protein assignments, to achieve a peptide-level FDR of <1%. Details of the filtered protein and peptide-level data can be found in the Supplementary Data of the original publication [1] and on GitHub ([https://github.com/dymult/DataInBrief\\_Deamidation](https://github.com/dymult/DataInBrief_Deamidation)).

## 2.6. Assessment of Deamidation Ratios

The deamidation ratios of the ancient and modern samples were assessed using the R coding language (version 3.6.3) [12] operating in the RStudio integrated development environment (version 1.3.1073). The R script, input data files, and graphical outputs can be found in Supplementary Data S1 and on GitHub ([https://github.com/dymult/DataInBrief\\_Deamidation](https://github.com/dymult/DataInBrief_Deamidation)). An explanation of how the R script works is as follows:

Empty data frames called *deamtable*, *ker* and *nker* were created to store deamidation ratios overall and within protein groups (In this instance: keratins and non-keratin). All Excel files within the current directory are read in using **readxl::read\_excel** and put into the variable *files*. Sample names are read from the Excel file names, removing the .xlsx extension.

Each file is run through a loop to calculate and collate the deamidation ratios. Four columns are extracted from the file and stored in *input*: peptide sequence (sequence), observed modifications (modifications), protein accession ID (protein), protein description (description). *Contam* contains a list of common contaminants (proteins identified from the cRAP database) and reversed identifications which are removed in the next command using **grepl** and stored in *filt*. The GPM modifications column contains strings separated by semicolons, and we use **strsplit** with a ';' separator and **unlist** to create a large character vector (*mod*) of modifications. We subset this character string into *deam* using **grepl** searching for "deamidated" to find all instances of deamidation within the peptide sequences. The number of deamidated glutamine and asparagine residues are counted with **grepl** and stored in variables *eventQ* and *eventN* respectively. The total number of glutamine and asparagine residues within the peptide sequences column are counted with **stringr::str\_count** and stored in *numQ* and *numN*, respectively. The ratios of deamidated glutamine and asparagine residues are calculated and stored in *ratioQ* and *ratioN*, respectively. For each file, a row is created in *deamtable* with the respective file name, and deamidation ratios for glutamine and asparagine.

Within the same loop, the files are separated into keratins (*ker*) and non-keratins (*notker*) using **grepl**. Both sets of data are listed into *dataframes*. A second loop is run over the two data frames. Within this loop, deamidation ratios for each data set are calculated and collated, separately from the overall deamidation ratios using the same method as described above. For each data set, a row is created in *ker* and *nker* with the file name, and deamidation ratios for glutamine and asparagine. The keratin and non-keratin glutamine and asparagine deamidation ratios per file are stored in *pgdeam*. The per-file loop finishes here.

The *deamtable* dataframe with the overall deamidation ratios was converted from wide to long format using **reshape2::melt**, with sample name as ID variable and ratio as value name. The ratio column was converted to a numeric variable and the second column was renamed 'amino\_acid'. Two descriptive columns called 'age' and 'residue' were generated to differentiate

between ancient and modern samples and glutamine and asparagine residues using **grepl**. A descriptive column called 'sampleID' was generated to simplify the plotting labels for the X axis.

The *pgdeam* dataframe with the protein group deamidation ratios was converted from wide to long format as described above (*pgdeammelt*), and the 'amino\_acid', 'age', 'residue' and 'sampleID' columns were added in. Additionally, a descriptive column called 'protein' was generated using **grepl** to differentiate between keratin and non-keratin ratios.

The samples in this study were the sampled interior and exterior surfaces of various cranial and other bones. To assess a sample-specific deamidation value, the data for both surfaces of each bone sample and all three modern samples were respectively averaged as follows: A summary function (*summary\_func*) was written to calculate the mean and standard deviation within the samples to produce an overall picture of the deamidation of non-keratin proteins. The keratin proteins were filtered out of *pgdeammelt* using **subset** to create *nkdeammelt*. The *summary\_func* function was applied over *nkdeammelt* using **plyr::ddply** with columns 'sampleID' and 'residue' as the groupings and stored in *avdeammelt*.

## 2.7. Data Visualisation

A PNG wrapped ggplot2 grouped bar chart was created using the *deammelt* data with X axis 'sample\_name', Y axis 1-'ratio' (percentage of non-deamidated residues) and coloured by residue. The chart was faceted by the 'age' column. The output file was called 'bulkdeam.png'.

A PNG wrapped ggplot2 grouped bar chart was created using the *pgdeammelt* data with X axis 'sample\_name', Y axis 1-'ratio' (percentage of non-deamidated residues) and coloured by residue. The chart was faceted by both the 'protein' and 'age' columns. The output file was called 'pgdeam.png'.

A PNG wrapped ggplot2 grouped bar chart was created using the *avdeammelt* data with X axis 'sampleID', Y axis 1-'mean' (average percentage of non-deamidated residues) and coloured by residue. Error bars of one standard deviation were used and plotted using **ggplot2::geom\_errorbar**. The output file was called 'avdeamnonkeratin.png'.

Examples of these graphical outputs can be found in the Supplementary Data of the original publication [1], and also at the GitHub repository referred to above.

## Ethics Statements

Ethics clearance was not required for the ancient samples analysed in this study as the remains were known to be several thousand years old and no living relatives are known. A letter of exemption was issued by the Macquarie University Ethics Committee (Reference No. 5201849176758). Modern skin samples were by DHM from his own forearm and are exempt from ethics approval on the grounds of self-experimentation.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability

[DataInBrief\\_Deamidation \(Original data\)](#) (GitHub).

[Development of a novel minimally invasive sampling and analysis technique for bioarchaeological proteomics \(Original data\)](#) (PRIDE proteomics database).

## CRedit Author Statement

**Dylan H. Multari:** Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Writing – original draft, Visualization; **Prathiba Ravishankar:** Conceptualization, Writing – review & editing; **Geraldine J. Sullivan:** Software, Formal analysis, Visualization, Writing – original draft; **Ronika K. Power:** Supervision, Writing – review & editing; **Constance Lord:** Resources, Writing – review & editing; **James A. Fraser:** Resources, Writing – review & editing; **Paul A. Haynes:** Supervision, Conceptualization, Resources, Writing – review & editing, Project administration.

## Acknowledgments

DHM acknowledges scholarship support from Macquarie University. PAH acknowledges funding support from Macquarie University and the Australian Research Council. The mass spectrometric analysis performed in this study was facilitated by access to the Australian Proteome Analysis Facility (APAF) funded under the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS)/Education Investment Fund. The authors would like to thank the curatorial and collection management teams of the Nicholson Collection, Chau Chak Wing Museum (The University of Sydney), with particular thanks to Candace Richards for support and access to materials.

## References

- [1] D.H. Multari, P. Ravishankar, G.J. Sullivan, R.K. Power, C. Lord, J.A. Fraser, P.A. Haynes, Development of a novel minimally invasive sampling and analysis technique using skin sampling tape strips for bioarchaeological proteomics, *J. Archaeol. Sci.* 139 (2022) 105548, doi:[10.1016/j.jas.2022.105548](https://doi.org/10.1016/j.jas.2022.105548).
- [2] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaíno, The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucleic Acids Res.* 47 (D1) (2019) D442–D450, doi:[10.1093/nar/gky1106](https://doi.org/10.1093/nar/gky1106).
- [3] J. Hendy, F. Welker, B. Demarchi, C. Speller, C. Warinner, M.J. Collins, A guide to ancient protein studies, *Nat. Ecol. Evol.* 2 (5) (2018) 791–799, doi:[10.1038/s41559-018-0510-x](https://doi.org/10.1038/s41559-018-0510-x).
- [4] M.L. Clausen, H.C. Slotved, K.A. Krogfelt, T. Agner, Tape stripping technique for stratum corneum protein analysis, *Sci. Rep.* 6 (2016) 19918–19918, doi:[10.1038/srep19918](https://doi.org/10.1038/srep19918).
- [5] M. Mirzaei, N. Soltani, E. Sarhadi, I.S. George, K.A. Neilson, D. Pascovici, S. Shahbazian, P.A. Haynes, B.J. Atwell, G.H. Salekdeh, Manipulating root water supply elicits major shifts in the shoot proteome, *J. Proteome Res.* 13 (2) (2014) 517–526, doi:[10.1021/pr400696u](https://doi.org/10.1021/pr400696u).
- [6] S. Muralidharan, E. Thompson, D. Raftos, G. Birch, P.A. Haynes, Quantitative proteomics of heavy metal stress responses in Sydney rock oysters, *Proteomics* 12 (6) (2012) 906–921, doi:[10.1002/pmic.201100417](https://doi.org/10.1002/pmic.201100417).
- [7] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T.A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R.L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, *Nat. Biotechnol.* 30 (10) (2012) 918–920, doi:[10.1038/nbt.2377](https://doi.org/10.1038/nbt.2377).
- [8] D. Kessner, M. Chambers, R. Burke, D. Agus, P. Mallick, ProteoWizard: open source software for rapid proteomics tools development, *Bioinformatics* 24 (21) (2008) 2534–2536, doi:[10.1093/bioinformatics/btn323](https://doi.org/10.1093/bioinformatics/btn323).
- [9] R. Craig, J.P. Cortens, R.C. Beavis, Open source system for analyzing, validating, and storing protein identification data, *J. Proteome Res.* 3 (6) (2004) 1234–1242, doi:[10.1021/pr049882h](https://doi.org/10.1021/pr049882h).
- [10] R. Craig, R.C. Beavis, TANDEM: matching proteins with tandem mass spectra, *Bioinformatics* 20 (9) (2004) 1466–1467, doi:[10.1093/bioinformatics/bth092](https://doi.org/10.1093/bioinformatics/bth092).
- [11] J. Shin, Y. Kwon, S. Lee, S. Na, E.Y. Hong, S. Ju, H.G. Jung, P. Kaushal, S. Shin, J.H. Back, S.Y. Choi, E.H. Kim, S.J. Lee, Y.E. Park, H.S. Ahn, Y. Ahn, M.H. Kabir, S.J. Park, W.S. Yang, J. Yeom, O.Y. Bang, C.W. Ha, J.W. Lee, U.B. Kang, H.J. Kim, K.S. Park, J.E. Lee, J.E. Lee, J.Y. Kim, K.P. Kim, Y. Kim, H. Hirano, E.C. Yi, J.Y. Cho, E. Paek, C. Lee, Common Repository of FBS Proteins (cRFP) to be added to a search database for mass spectrometric analysis of cell secretome, *J. Proteome Res.* 18 (10) (2019) 3800–3806, doi:[10.1021/acs.jproteome.9b00475](https://doi.org/10.1021/acs.jproteome.9b00475).
- [12] R Core Team, R: a language and environment for statistical computing., 2020.