



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Proteomics data for characterizing *Microbacterium oleivorans* A9, an uranium-tolerant actinobacterium isolated near the Chernobyl nuclear power plant



Nicolas Gallois<sup>a</sup>, Laurie Piette<sup>a</sup>, Philippe Ortet<sup>c</sup>,  
Mohamed Barakat<sup>c</sup>, Justine Long<sup>c</sup>, Catherine Berthomieu<sup>a</sup>,  
Jean Armengaud<sup>b,\*</sup>, Virginie Chapon<sup>a</sup>, Béatrice Alpha-Bazin<sup>b</sup>

<sup>a</sup> CEA, CNRS, Aix-Marseille Université, UMR 7265 Biologie Végétale et Microbiologie Environnementales, Laboratoire des Interactions Protéine Métal, 13108 Saint-Paul-lez-Durance, France

<sup>b</sup> CEA, DRF/Joliot/SPI/Li2D, BP 17171, F-30200 Bagnols-sur-Cèze, France

<sup>c</sup> CEA, CNRS, Aix-Marseille Université, UMR 7265 Biologie Végétale et Microbiologie Environnementales, Laboratoire d'écologie microbienne de la rhizosphère et d'environnements extrêmes, 13108 Saint-Paul-lez-Durance, France

## ARTICLE INFO

## Article history:

Received 2 December 2017

Received in revised form

5 March 2018

Accepted 25 October 2018

Available online 30 October 2018

## ABSTRACT

*Microbacterium oleivorans* A9 cells were exposed or not to 10 μM uranyl nitrate as resting cells in sodium chloride solution. Bacteria exposed to U(VI) and controls were harvested after 0.5, 4, and 24 h of toxicant exposure. Bacteria were subjected to high-throughput proteomics analysis using a Q-Exactive HF high resolution tandem mass spectrometer incorporating an ultra-high-field orbitrap analyzer. MS/MS spectra were assigned with a protein sequence database derived from a draft genome obtained by Illumina sequencing and systematic six-reading frame translation of all the contigs. Proteins identified in bacteria exposed to U(VI) and controls at the three time points allow defining the proteome dynamics upon uranium stress. The data reported here are related to a published study regarding the proteome dynamics of *M. oleivorans* A9 upon uranium stress by Gallois et al. (in press) entitled “Proteogenomic insights into uranium tolerance of a Chernobyl's *Microbacterium* bacterial isolate”. The data accompanying the manuscript describing the database searches and

\* Correspondence to: CEA-Marcoule, DRF/ISVF/SPI/Li2D, Laboratory “Innovative technologies for Detection and Diagnostics”, BP 17171, F-30200 Bagnols-sur-Cèze, France. Fax: +00 33 4 66 79 19 05.

E-mail address: [jean.armengaud@cea.fr](mailto:jean.armengaud@cea.fr) (J. Armengaud).

<https://doi.org/10.1016/j.dib.2018.10.136>

2352-3409/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

comparative analysis have been deposited to the ProteomeXchange with identifier PXD005794.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications table

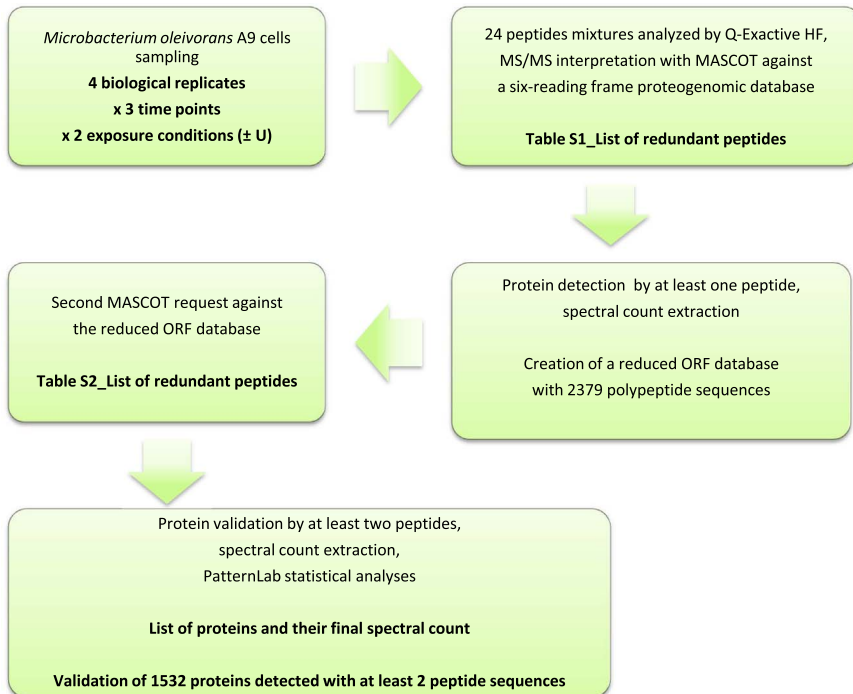
Subject area	<i>Environmental microbiology</i>
More specific subject area	<i>Actinobacteria comparative proteogenomics</i>
Type of data	<i>Figure, mass spectrometry raw files, Excel tables</i>
How data was acquired	<i>Data-dependent acquisition of tandem mass spectra using a Q-Exactive HF tandem mass spectrometer (Thermo).</i>
Data format	<i>Raw and processed</i>
Experimental factors	<i>Cells, at the exponential growth phase, were harvested and the resulting cell pellets were resuspended in NaCl solution with 0 (control) or 10 <math>\mu</math>M uranyl nitrate. For each of the three sampling time points, 0.5, 4 and 24 h, four biological replicates were performed.</i>
Experimental features	<i>The 24 proteomes were briefly run on SDS-PAGE, followed by trypsin proteolysis. Tryptic peptides were analyzed by nanoLC-MS/MS and spectra were assigned with a draft genome-derived protein sequence database.</i>
Data source location	<i>CEA-Marcoule, DRF-Li2D, Laboratory “Innovative technologies for Detection and Diagnostics”, BP 17171, F-30200 Bagnols-sur-Cèze, France</i>
Data accessibility	<i>Data are within this article and deposited to the ProteomeXchange via the PRIDE repository with identifier PRIDE: <a href="https://proteomecentral.proteomexchange.org/protein/PXD005794">PXD005794</a>.</i>

## Value of the data

- The data are an interesting resource regarding the proteome content of a soil bacterium with extreme tolerance to heavy metals.
- The proteogenomics strategy described here allows a quick identification of proteins based on a draft genome of this high GC content organism.
- The data have been used to define the proteome changes in response to uranium stress in an Actinobacteria isolate from the trench T22 located near the Chernobyl nuclear power plant. As described in detail in the accompanying manuscript [1], the uranyl stress perturbed the phosphate and iron metabolic pathways.

## 1. Data

This report contains the complete list of peptide-to-spectrum assignments for the control (not treated) samples and Uranium-treated samples of *Microbacterium oleivorans* A9 in the first round of the proteogenomics cascade search (Supplementary Table S1), in the second round of the cascade search (Supplementary Table S2). A total of 746,092 MS/MS spectra were assigned in the first search round and 747,621 MS/MS spectra were interpreted in the second search round. All the tandem mass spectrometry characteristics (measured mass, charge, protease cleavage, post-translational modifications, and retention time) are indicated.



**Fig. 1.** Flowchart of experiments, data processing and refined data tables.

**Fig. 1** shows the schematic flowchart of experiments, data processing and results that were obtained and formatted in.xls tables. The proteome data from four independent biological replicates per time point (0.5, 4 and 24 h) upon uranium exposure or control, i.e. from 24 samples, were assigned to tryptic peptides against the *M. oleivorans* A9 ORF database described by Gallois *et al.* [1] following a proteogenomic approach [2–4]. The deposited data comprised the 24 raw files and the interpreted files. [Supplementary Tables S1 and S2](#) list the peptide-to-spectrum matches with all the tandem mass spectrometry characteristics in the first query round against the large proteogenomics database and second query round against the reduced ORF database, respectively.

## 2. Experimental design, materials and methods

### 2.1. Preparation of *Microbacterium oleivorans* A9 samples

*M. oleivorans* A9 cells were isolated from Chernobyl trench T22 soil [5] and exposed as described [1]. Briefly, cells were harvested and the cell pellets were resuspended in 0.1 M NaCl pH 5.0 with 0 (control) or 10  $\mu$ M uranyl nitrate. Four independent biological replicates were performed for statistical purpose. Fractions of 1 ml of cell suspension were taken after 0.5, 4, and 24 h for each condition. After centrifugation, the resulting supernatants were removed and the pellets were conserved at  $-80^{\circ}\text{C}$  until proteomic analysis. Proteogenomics were carried out as previously described [1,6] taking into account the recent draft genome of the strain.

## 2.2. Protein extracts and tandem mass spectrometry

The 24 peptide mixtures were performed using a Q-Exactive HF mass spectrometer (Thermo-Fisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings) in similar conditions as those previously described. Peptide mixtures (10 µl) were loaded and desalted on-line on a reverse phase precolumn (Acclaim PepMap 100 C18) from LC Packings. Peptides were then resolved onto a reverse phase Acclaim PepMap 100 C18 column and injected into the Q-Exactive HF mass spectrometer. The Q-Exactive HF instrument was operated according to a Top20 data-dependent acquisition method as previously described [7].

## 2.3. Protein sequence database for proteogenomic MS/MS assignment

The recorded MS/MS spectra for the 24 samples were searched against our home-made ORF database with the parameters described previously [1]. This database contains 30,853 polypeptide sequences, for a total of 4,903,573 amino acids with an average of 159 amino acids per polypeptide. The number of MS/MS spectra per protein (spectral counts) was determined for the four replicates of each of the three time points for both conditions. The statistical protein variation was compared for each time point between the uranyl exposure and the control conditions using the T-Fold option of PatternLab 2.0 software [8].

## Acknowledgements

Nicolas Gallois was financially supported by the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA) by means of the program "Thèse Amont-Aval" directed by Yves Bréchet, Haut-Commissaire à l'Énergie Atomique. The project was supported by CEA through the Toxicology program (BEnUr).

## Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.10.136>.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.10.136>.

## References

- [1] N. Gallois, B. Alpha-Bazin, P. Ortet, M. Barakat, L. Piette, J. Long, C. Berthomieu, J. Armengaud, V. Chapon, Proteogenomic insights into uranium tolerance of a Chernobyl's Microbacterium bacterial isolate, *J. Proteom.* 177 (2018) 148–157.
- [2] E.M. Hartmann, J. Armengaud, N-terminomics and proteogenomics, getting off to a good start, *Proteomics* 14 (23–24) (2014) 2637–2646.
- [3] J. Armengaud, E.M. Hartmann, C. Bland, Proteogenomics for environmental microbiology, *Proteomics* 13 (18–19) (2013) 2731–2742.
- [4] J. Armengaud, J. Trapp, O. Pible, O. Geffard, A. Chaumot, E.M. Hartmann, Non-model organisms, a species endangered by proteogenomics, *J. Proteom.* 105 (2014) 5–18.
- [5] V. Chapon, L. Piette, M.-H. Vesvres, F. Coppin, C.L. Marrec, R. Christen, N. Theodorakopoulos, L. Février, S. Levchuk, A. Martin-Garin, C. Berthomieu, C. Sergeant, Microbial diversity in contaminated soils along the T22 trench of the Chernobyl experimental platform, *Appl. Geochem.* 27 (7) (2012) 1375–1383.
- [6] E.M. Hartmann, F. Allain, J.C. Gaillard, O. Pible, J. Armengaud, Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria, *Methods Mol. Biol.* 1197 (2014) 275–285.

- [7] G. Klein, C. Mathe, M. Biola-Clier, S. Devineau, E. Drouineau, E. Hatem, L. Marichal, B. Alonso, J.C. Gaillard, G. Lagniel, J. Armengaud, M. Carriere, S. Chedin, Y. Boulard, S. Pin, J.P. Renault, J.C. Aude, J. Labarre, RNA-binding proteins are a major target of silica nanoparticles in cell extracts, *Nanotoxicology* 10 (10) (2016) 1555–1564.
- [8] P.C. Carvalho, J. Hewel, V.C. Barbosa, J.R. Yates III, Identifying differences in protein expression levels by spectral counting and feature selection, *Genet. Mol. Res.* 7 (2) (2008) 342–356.