



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Ovary and embryo proteogenomic dataset revealing diversity of vitellogenins in the crustacean *Gammarus fossarum*Judith Trapp^{a,b}, Jean-Charles Gaillard^b, Arnaud Chaumot^a, Olivier Geffard^a, Olivier Pible^b, Jean Armengaud^{b,*}^a Irstea, Unité de Recherche MALY, Laboratoire d'écotoxicologie, CS70077, F-69626 Villeurbanne, France^b CEA-Marcoule, DSV/IBITEC-S/SP/Li2D, Laboratory "Innovative technologies for Detection and Diagnostic", BP 17171, F-30200 Bagnols-sur-Cèze, France

ARTICLE INFO

Article history:

Received 11 July 2016

Received in revised form

18 July 2016

Accepted 21 July 2016

Available online 28 July 2016

Keywords:

Crustacean

Embryogenesis

Non-model organism

Oogenesis

Proteogenomics Reproduction

Vitellogenins

ABSTRACT

Ovaries and embryos from sexually mature *Gammarus fossarum* were sampled at different stages of the reproductive cycle. The soluble proteome was extracted for five biological replicates and samples were subjected to trypsin digestion. The resulting peptides were analyzed by high resolution tandem mass spectrometry with a LTQ-Orbitrap XL instrument. The MS/MS spectra were assigned with a previously described RNAseq-derived *G. fossarum* database. The proteins highlighted by proteogenomics were monitored and their abundance kinetics over the different stages revealed a large panel of vitellogenins. Criteria were i) accumulation during oogenesis, ii) decrease during embryogenesis, iii) classified as female-specific, and iv) sequence similarity and phylogenetic analysis. The data accompanying the manuscript describing the database searches and comparative analysis ("*High-throughput proteome dynamics for discovery of key proteins in sentinel species: unsuspected vitellogenins diversity in the crustacean Gammarus fossarum*" by Trapp et al. [1]) have been deposited to the ProteomeXchange via the PRIDE repository with identifiers PRIDE: PXD001002.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

DOI of original article: <http://dx.doi.org/10.1016/j.jprot.2016.07.007>

* Corresponding author. Fax: +33 466791905.

E-mail address: jean.armengaud@cea.fr (J. Armengaud).<http://dx.doi.org/10.1016/j.dib.2016.07.045>2352-3409/© 2016 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 biology</i>
More specific sub- ject area	<i>Amphipod comparative proteogenomics</i>
Type of data	<i>Figure</i>
How data was acquired	<i>Data-dependent acquisition of tandem mass spectra using a LTQ-Orbitrap-XL mass spectrometer (Thermo).</i>
Data format	<i>Raw</i>
Experimental factors	For each female, ovaries or embryos were dissected under stereomicroscope magnification, immediately frozen in liquid nitrogen and stored at -80°C until needed. Proteins were extracted and analyzed by shotgun proteogenomics. Five biological replicates were performed for each stage. Five molt stages were analyzed: post-molt stage A and B, inter-molt stage C1, inter-molt stage C2, pre-molt stage D1, and pre-molt stage D2. Accordingly, five different embryo development stages were analyzed.
Experimental features	<i>The 50 soluble protein extracts were briefly run on SDS-PAGE, followed by in-gel trypsin proteolysis. Tryptic peptides were analyzed by nanoLC-MS/MS and spectra were assigned with a RNA-seq derived protein sequence database.</i>
Data source location	CEA-Marcoule, DSV-Li2D, Laboratory “Innovative technologies for Detection and Diagnostics”, BP 17,171, F-30200 Bagnols-sur-Cèze, France
Data accessibility	Data is within this article and deposited to the ProteomeXchange via the PRIDE repository with identifier PRIDE: PXD001002 .

Value of the data

- The data represent the largest repertoire of proteins involved in reproduction established from a crustacean based on 50 samples from various stages of the reproductive cycle.
- Because gammarids are considered as sentinel organisms with great potential in the field of ecotoxicology and more specifically freshwater health monitoring, these data represent a useful resource for potential toxicological biomarkers [2].
- The data have been used to characterize the diversity of vitellogenins in amphipods. As described in detail in the companion manuscript [1], this diversity is rather unusual and calls for additional functional characterization of this crucial family of proteins.

1. Data

Fig. 1 shows the schematic flowchart of experiments, data processing and interpreted results that were presented in four large tables and published recently [1]. The proteome data from five independent biological replicates per stages (5 different oocyte molt stages and 5 different embryo development stages), i.e. from 50 proteome samples, were assigned to tryptic peptides against the *G. fossarum* RNAseq derived database described by Trapp et al. [3] following a proteogenomic strategy [4–6]. The Deposited data comprised the 50 raw files, the protein sequence database, and the interpreted files.

2. Experimental design, materials and methods

2.1. Sampling of animals and preparation of biological samples

Female *G. fossarum* amphipods were collected as described [1]. Based on limb inter-tegmental change criteria, molt stages of female gammarids were classified into five different categories: post-molt stages A

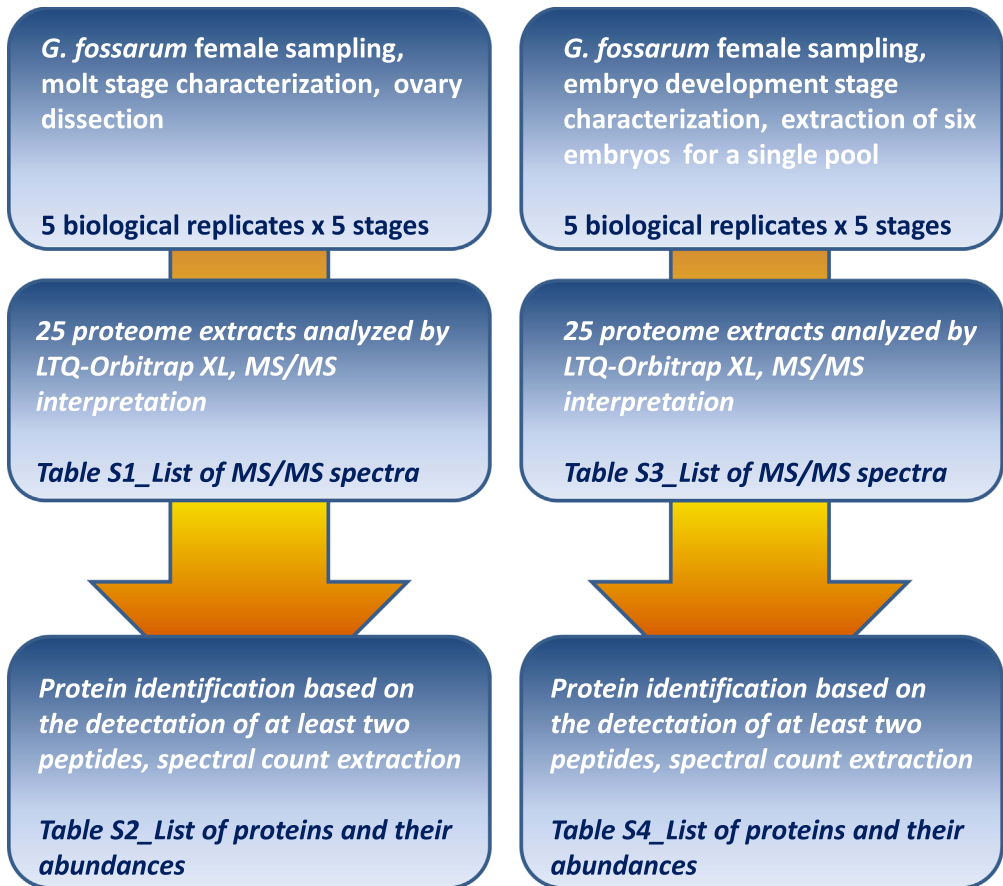


Fig. 1. Flowchart of experiments, data processing and refined data tables.

and B, inter-molt stage C1, inter-molt stage C2, pre-molt stage D1, and pre-molt stage D2, as previously described [7]. Accordingly, five different embryo development stages were also delineated. For each female, six embryos were collected from the ventral pouch and the ovaries were excised under stereomicroscope magnification as described by Lacaze et al. [8]. For each stage, five biological replicates were performed, immediately frozen in liquid nitrogen and stored at -80°C until needed. Proteins were processed as previously described [1,9].

2.2. Tandem mass spectrometry

Peptide identification was performed by nanoLC-MS/MS with a LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings) [10,11]. Peptides were resolved on a nanoscale C18 PepMapTM 100-capillary column (LC Packings) prior to injection into the ion trap mass spectrometer as previously described [12]. Full-scan mass spectra were measured from m/z 300 to 1800 with the LTQ-Orbitrap XL mass spectrometer in data-dependent mode with a scan cycle initiated with a full scan of high mass accuracy in the Orbitrap followed by MS/MS scans in the linear ion trap on the three most abundant ions.

2.3. Data interpretation and monitoring of proteome dynamics

MS/MS spectra were assigned against the GFOSS protein database, created from RNA-seq data acquired on *G. fossarum*. This database comprises 1,311,444 entries totaling 289,084,257 amino acids. Molecular ion peak lists were extracted as described previously by Christie-Oleza et al. [13]. Peptide assignment with MASCOT and protein validation were done with the parameters described previously [1]. The number of spectra recorded per protein (spectral counts) was extracted from the spectra-to-peptide data set for each protein and each experimental condition. The proteome dynamics along the different stages was assessed with the TrendQuest module of the PatternLab program [14].

Acknowledgements

We thank the Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (France), the Commissariat à l'Energie Atomique et aux Energies Alternatives (France) through the transversal toxicology program (PPTOX), and the Agence Nationale de la Recherche program "ProteoGam" (ANR-14-CE21-0006-02) for financial support.

Transparency document. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.07.045>.

References

- [1] J. Trapp, J. Armengaud, J.-C. Gaillard, O. Pible, A. Chaumot, O. Geffard, High-throughput proteome dynamics for discovery of key proteins in sentinel species: unsuspected vitellogenins diversity in the crustacean *Gammarus fossarum*, *J. Proteom.* 146 (2016) 207–214.
- [2] J. Trapp, J. Armengaud, A. Salvador, A. Chaumot, O. Geffard, Next-generation proteomics: toward customized biomarkers for environmental biomonitoring, *Environ. Sci. Technol.* 48 (2014) 13560–13572.
- [3] J. Trapp, O. Geffard, G. Imbert, J.-C. Gaillard, A.-H. Davin, A. Chaumot, J. Armengaud, Proteogenomics of *Gammarus fossarum* to document the reproductive system of amphipod, *Mol. Cell. Proteom.* 13 (2014) 3612–3625.
- [4] J. Armengaud, E.M. Hartmann, C. Bland, Proteogenomics for environmental microbiology, *Proteomics* 13 (2013) 2731–2742.
- [5] E.M. Hartmann, J. Armengaud, N-terminomics and proteogenomics, getting off to a good start, *Proteomics* 14 (2014) 2637–2646.
- [6] 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.
- [7] O. Geffard, B. Xuereb, A. Chaumot, A. Geffard, S. Biagianni, C. Noel, K. Abbaci, J. Garric, G. Charmantier, M. Charmantier-Daures, Ovarian cycle and embryonic development in *Gammarus fossarum*: application for reproductive toxicity assessment, *Environ. Toxicol. Chem.* 29 (2010) 2249–2259.
- [8] E. Lacaze, O. Geffard, S. Bony, A. Devaux, Genotoxicity assessment in the amphipod *Gammarus fossarum* by use of the alkaline Comet assay, *Mutat. Res. -Gen. Tox. En.* 700 (2010) 32–38 59.
- [9] 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.
- [10] A. Dedieu, J.-C. Gaillard, T. Pourcher, E. Darrouzet, J. Armengaud, Revisiting iodination sites in thyroglobulin with an organ-oriented shotgun strategy, *J. Biol. Chem.* 286 (2011) 259–269.
- [11] E.M. Hartmann, J. Armengaud, Shotgun proteomics suggests involvement of additional enzymes in dioxin degradation by *Sphingomonas wittichii* RW1, *Environ. Microbiol.* 16 (2014) 162–176.
- [12] J.A. Christie-Oleza, G. Miotello, J. Armengaud, High-throughput proteogenomics of *Ruegeria pomeroyi*: seeding a better genomic annotation for the whole marine *Roseobacter* clade, *BMC Genom.* 13 (2012) 73.
- [13] J.A. Christie-Oleza, J.M. Piña-Villalonga, R. Bosch, B. Nogales, J. Armengaud, Comparative proteogenomics of twelve *Roseobacter* exoproteomes reveals different adaptive strategies among these marine bacteria, *Mol. Cell. Proteom.* 11 (2012), M111.013110.
- [14] P.C. Carvalho, D.B. Lima, F.V. Leprevost, M.D. Santos, J.S. Fischer, P.F. Aquino, J.J. Moresco, J.R. Yates 3rd, V.C. Barbosa, Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0, *Nat. Protoc.* 11 (2016) 102–117.