HAL
open science

Molecular approaches for the discovery, development and application of toxicity biomarkers in gammarids

Duarte Domingos Gouveia

To cite this version:

Duarte Domingos Gouveia. Molecular approaches for the discovery, development and application of toxicity biomarkers in gammarids. Ecotoxicology. Université de Lyon, 2017. English. NNT : 2017LYSE1318 . tel-01723905

HAL Id: tel-01723905

https://theses.hal.science/tel-01723905

Submitted on 5 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nº ordre NNT : 2017LYSE1318

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 341
(Evolution Ecosystèmes Microbiologie Modélisation)

Spécialité : Ecotoxicologie moléculaire

Soutenue publiquement le 18/12/2017, par :
Duarte Domingos Gouveia

Approches moléculaires pour la découverte, le développement et l'application de biomarqueurs de toxicité chez les gammaridés

Composition du jury :

Christophe MINIER, Professeur, Université du Havre	Rapporteur
Frédéric SILVESTRE, Professeur, Université de Namur	Rapporteur
Charlotte CORPOREAU, Chercheuse, IFREMER Bretagne	Examinatrice
Nathalie MONDY, Maître de Conférences, Université Lyon 1	Examinatrice
Olivier GEFFARD, Directeur de Recherche, Irstea Lyon	Directeur de thèse
Jean ARMENGAUD, Directeur de Recherche, CEA Marcoule	Co-directeur de thèse
Arnaud CHAUMOT, Chargé de Recherche, Irstea Lyon	Encadrant
Christine ALMUNIA, Chargé de Recherche, CEA Marcoule	Encadrant

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. le Professeur Frédéric FLEURY**
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Fabrice VALLÉE
Directrice Générale des Services	Mme Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur G.RODE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. X. PERROT
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. THEVENARD
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur G. TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. le Professeur J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. le Professeur E.PERRIN
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

ACKNOWLEDGEMENTS

Je tiens à remercier premièrement à mes encadrants de thèse Olivier Geffard, Arnaud Chaumot, Jean Armengaud et Christine Almunia. Merci de m'avoir donné l'opportunité d'effectuer ma thèse sous votre tutelle, j'ai du mal à trouver les mots pour exprimer ma gratitude pour tout ce que vous m'avez apporté. Ces trois années ont été très enrichissantes tant au niveau scientifique/professionnelle comme personnelle. Olivier, j'admire énormément tes qualités humaines, ainsi comme tes capacités de conduire l'équipe toujours pour des nouveaux challenges et faire progresser les acquis et compétences du labo. Merci pour avoir été l'encadrant que tu as été, pour ta disponibilité, pour ton positivisme, et pour m'avoir toujours guidé pour le bon déroulement de la thèse. Arnaud, j'ai beaucoup apprécié ta patience et disponibilité, les multiples conseils et toutes les heures que tu as consacré pour m'expliquer les nombreuses doutes et questions que je t'ai posés pendant ces trois ans. Merci aussi pour m'avoir appris une multitude de faits historiques lors de nos sorties terrain ou longues voyages en train pour les conférences. Jean, c'était un privilège de travailler avec un des grands experts en protéomique/spectrométrie de masse, je te remercie pour ton accueil au sein de ton équipe de recherche au CEA. Merci aussi pour te rendre disponible à chaque fois que j'ai sollicité ton aide, pour tes précieux avis sur les papiers, et pour avoir posé ta confiance en moi pour la suite. Christine, je te remercie énormément pour ton implication dans mes travaux de thèse, notamment pour m'avoir formée à l'analyse des données protéomiques, et pour tes avis précieux pour le déroulement des manips. Merci aussi pour ta bonne humeur, qui fait la journée de travail encore plus agréable. J'en garderai l'excellent souvenir des covoiturages lors de mes multiples visites au CEA.

Je remercie à l'ensemble des membres du jury qui ont accepté d'évaluer mon travail : Dr. Frédéric Silvestre, Dr. Christophe Minier, Dr. Charlotte Corporeau et Dr. Nathalie Mondy. C'est un grand honneur d'avoir mon travail jugé par vos compétences scientifiques.

Je remercie également les membres de mon comité de suivi de thèse : François Bonneton, Arnaud Salvador, Charlotte Lecureuil et Gilles Comte. J'ai beaucoup apprécié les échanges scientifiques lors de nos réunions, vous nous avez beaucoup aidés dans la prise de décisions pour le bon déroulement de ma thèse. Un remerciement plus particulier à Arnaud Salvador et François Bonneton, pour m'avoir donné l'opportunité de travailler au sein de vos laboratoires de recherche, et pour l'inestimable orientation que vous m'avez donnée lors de ces travaux.

Ayant eu la chance de réaliser ma thèse entre plusieurs laboratoires, je voulais remercier tous mes collègues, plus particulièrement à ces de l'Irstea, pour avoir contribué à un super environnement scientifique et de travail. Hervé, je te remercie pour ta constante bonne humeur contagieuse, tes nombreuses leçons de l'argot français, et surtout pour ton amitié et générosité. Je te suis éternellement reconnaissant pour ton aide tout au long de ces trois années. Adeline, je te remercie aussi pour ta bonne humeur, pour ta disponibilité en m'aider avec mes appels téléphoniques en français, et surtout pour ton amitié. J'en garderai que des bons souvenirs. Un grand merci à mes collègues thésards avec qui j'ai eu le plaisir de partager le bureau : Judith pour avoir pris du temps en fin de thèse pour m'aider avec le sujet de thèse ; Amandine pour les bons moments au bureau et pour m'avoir aidé avec le français ; Paul pour les nombreuses pauses gouter et les bons moments passés au tour d'un kinder country et une tasse de

camomille; Alice Vidal pour sa bonne humeur et pour transmettre toujours des bonnes « vibes » (et aussi pour avoir la patience de m'expliquer mes interminables doutes sur la langue française) ; Aude et Pauline c'était un plaisir de partager le bureau avec vous. Et bon courage pour vos thèses en cours. Je remercie aussi tous les autres membres actuels ou passés du labo qui m'ont aidé à planifier et mettre en place les travaux expérimentaux, et avec qui j'ai pu passer de très bons moments de convivialité : Patrice, Renaud, Khédidja, Laura, Nicolas, Lucile, Alexandre Ciliberti (merci pour les bonnes aventures lors de nos sorties RMC), Guillaume Jubeaux. Merci aussi aux autres chercheurs du laboratoire avec qui j'ai eu des échanges très enrichissants : Marc, Hélène, Jeanne, et Davide (dommage que tu n'es pas arrivé plus tôt, j'ai beaucoup apprécié travailler avec toi pendant ces derniers mois). Merci aux stagiaires de master avec qui j'ai eu l'opportunité de travailler avec, et qu'ont beaucoup contribué pour le bon déroulement des manips au laboratoire : Laurent, Alice Cottereau, et Maxime.

Au personnel du CEA, ISA et IGFL : Aurore Charnot pour les longues journées de préparation des échantillons pour la masse ; Abderrahman Khila et Séverine Viala pour la précieuse orientation lors de mon début en biologie moléculaire ; Guylaine pour me prêter sa paillasse et micropipettes ; Céline et Laetitia pour avoir rendu disponible une place pour moi dans leur bureau au CEA; Gérard pour m'avoir formé aux analyses et à la manipulation de ARN au laboratoire; Olivier, Fabrice, Yannick, Béatrice, Jean Charles, Florent, Sophie, Laurent Bellanger, Charlotte, Cédric, Virginie.

A ma « famille Lyonnaise » : Rémi, Alexandre Decamps, Quentin, Estelle, Sandra, Emilia, Pierre-Etienne, Thiago, Marcelo, Diego, Tiago. J'aurais besoin de trop de pages pour vous remercier de toutes les choses que vous avez faites pour moi. En bref, merci pour votre amitié et générosité, pour tous les bons moments passés au laboratoire, à la coloc, et pendant tous les activités intra- et extra-Lyon. Vous m'avez beaucoup aidé à supporter mieux la distance que me séparait de ma famille et copine. Un merci plus particulier à Rémi, Alex et Quentin pour m'avoir fait découvrir quelques petites merveilles de son beau pays !

Je termine avec un grand merci à mes parents et à mon frère. Malgré la distance, les efforts que vous avez faits, et la force que vous m'avez transmis, m'ont fait surmonter toutes les difficultés et combattre le fait que vous me manquiez beaucoup. Un remerciement plus particulier à Rossana pour m'avoir soutenu lors de ma décision de partir en France, pour m'avoir motivé dans les moments moins bons, et pour avoir fait le sacrifice de s'éloigner de sa famille pour me rejoindre.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS ..... **III**

LIST OF FIGURES ..... **VII**

LIST OF TABLES ..... **XI**

LIST OF PUBLICATIONS AND COMMUNICATIONS ..... **XIII**

RESUME FRANÇAIS **1**

GENERAL INTRODUCTION **9**

CHAPTER I STATE OF THE ART **13**

1. PROTEIN BIOMARKERS IN ECOTOXICOLOGY AND MULTI-BIOMARKER APPROACHES 15
   - *1.1. Biomarkers: definition, classification, examples and limitations* *15*
     - 1.1.1. Examples of protein biomarkers used in animal Ecotoxicology 16
     - 1.1.2. Limitations and perspectives 20
   - *1.2. Multibiomarker approaches as an indicator of the health of organisms* *26*
2. OMICS TOOLS AND NEW STRATEGIES FOR MOLECULAR BIOMARKER DEFINITION 30
   - *2.1. Transcriptomics* *32*
   - *2.2. Proteomics* *34*
     - 2.2.1. Protein identification by mass spectrometry 35
     - 2.2.2. Proteogenomics for protein discovery in non-model species 39
     - 2.2.3. Mass-spectrometry-based protein quantitation 41
   - *2.3. Proteomics applications in Aquatic Ecotoxicology* *47*
   - *2.4. Proposal of a proteomics biomarker pipeline for environmental applications* *50*
3. GAMMARIDS AS SENTINEL SPECIES FOR ENVIRONMENTAL MONITORING OF FRESHWATERS 53
   - *3.1. Classification, morphology and ecology of Gammarus fossarum* *54*
   - *3.2. Crustacean endocrine systems* *57*
   - *3.3. Gammarus in Ecotoxicology and currently used biomarkers* *62*
   - *3.4. Current status in "omics" approaches using Gammarus* *64*
4. OBJECTIVES OF THE THESIS 67

   Development and application of a multiplexed protein biomarker measurement by mass spectrometry *67*

   Endocrine disruption biomarkers *68*

CHAPTER II MATERIALS AND METHODS **71**

1. BIOTESTS IN *GAMMARUS FOSSARUM* 73
   - *1.1. Sampling and maintenance of organisms* *73*
   - *1.2. Biological procedures* *74*
     - 1.2.1. Reproductive toxicity test 75
     - 1.2.2. Gonad dissection 77
     - 1.2.3. Sampling of organisms at different physiological conditions 77
   - *1.3. Protocols of contamination exposures* *78*
     - 1.3.1. Laboratory exposures to model contaminants 78
     - 1.3.2. In situ exposures 81
2. PROTEOMIC ANALYSES 84
   - *2.1. Targeted proteomic analysis by Selected Reaction Monitoring* *84*
     - 2.1.1. Candidate selection based on proteogenomics-derived protein sequences 84
     - 2.1.2. Analytical validation of the methodology 89
     - 2.1.3. Protein extraction and digestion 91
     - 2.1.4. LC-MS/MS analysis 92
     - 2.1.5. Data exploitation and absolute quantification 93

2.2. Shotgun proteomics of male gonads *95*
2.2.1. Protein extraction and digestion 95
2.2.2. LC-MS/MS analysis 95
2.2.3. Protein identification and relative quantification by spectral counting 96
3. GENETIC ANALYSES 97
3.1. Sequence alignments and phylogenetic analyses *97*
3.2. Protocols for gene expression *99*
3.2.1. Total RNA extraction and cDNA synthesis 99
3.2.2. PCR amplification, cloning and sequencing 100
3.2.3. Quantitative real-time PCR 102
3.2.4. Data exploitation and relative quantification of gene expression 104

CHAPTER III DEVELOPMENT AND APPLICATION OF A MULTIPLEXED PROTEIN BIOMARKER MEASUREMENT BY MASS SPECTROMETRY IN *GAMMARUS FOSSARUM* 107

PUBLICATION Nº1 109
PUBLICATION Nº2 125
PUBLICATION Nº3 143

CHAPTER IV DEVELOPMENT OF SPECIFIC BIOMARKERS FOR ENDOCRINE DISRUPTION ASSESSMENT IN *GAMMARUS FOSSARUM* 163

PUBLICATION Nº 4 165
NOTE 1: IDENTIFICATION OF CANDIDATE GENES VIA LITERATURE SEARCH 177
NOTE 2: REPROTOXICITY ASSESSMENT OF MODEL MOLECULES IN FEMALE GAMMARIDS 183
PUBLICATION Nº5 185

CHAPTER V SYNTHESIS AND GENERAL DISCUSSION 203

1. DEVELOPMENT OF A NEW MS-BASED MULTIBIOMARKER ABSOLUTE QUANTIFICATION IN *GAMMARUS FOSSARUM* 205
2. ENDOCRINE DISRUPTION BIOMARKERS 216
2.1. Comparative shotgun proteomics for ED biomarker discovery *216*
2.2. Candidate gene approach *219*

CONCLUSIONS & PERSPECTIVES 229

REFERENCES 235

20-HE :	20-hydroxyecdysone
2D-PAGE:	two-dimensional polyacrylamide gel electrophoresis
AChE :	acetylcholinesterase
ACN :	acetonitrile
AGH :	androgenic gland hormone
AHH :	aryl hydrocarbon hydroxylase
AhR :	arylhydrocarbon receptor
ALAD :	delta-aminolevulinic acid dehydratase
AMBIC:	ammonium bicarbonate
AOP :	adverse outcome pathway
AQUA :	absolute quantification
AST :	allatostatin
ATP :	adenosine triphosphate
ATPase:	adenosine triphosphatase
BBP :	n-Butyl benzyl phthalate
BDE-47:	pentabromodiphenyl ether
BR :	broad-complex
CAT :	catalase
CE :	collision energy
CHH :	crustacean hyperglycemic hormone
CXP :	cell exit potential
CYP :	cytochrome P450
DBD :	DNA binding domain
DIA :	data independent acquisition
DIGE :	DIfferential Gel Electrophoresis
DNA :	Deoxyribonucleic acid
DP :	declustering potential
DTT :	dithiothreitol
E74 :	ecdysone induced protein 74EF
E75 :	ecdysone induced protein 75B
E78 :	ecdysone induced protein 78C
EcR :	ecdysone receptor
ED :	endocrine disruption
EDC :	endocrine disrupting compound
EDTA :	ethylenediaminetetraacetic acid
EF :	elongation factor
ELISA :	enzyme-linked immunosorbent assays
ERA :	environmental risk assessment
EROD :	ethoxyresofurin O-deethylase
ESI :	electrospray ionisation
EU :	European Union
FaMET:	farnesoic methyl transferase
FDA :	US Food and Drug Administration
FKBP39:	FK506-binding protein 39 Kda
FTIR :	Fourier-transform infrared spectroscopy

GAPDH: glyceraldehyde 3-phosphate dehydrogenase
GC : gtas chromatography
GIH : gonad inhibiting hormone
GO : gene ontology
GPX : glutathione peroxidase
GSH : gonad stimulating hormone
GST : glutathione-S-transferase
GWAS : genome-wide association study
HCD : higher-energy collisional dissociation
HF : high-field
HIS : health status index
HPLC : high-performance liquid chromatography
HR3 : hormone receptor 3
HSP : heat shock protein
IAM : iodoacetamide
IBP : integrated biomarker proteomic
IBR : integrated biomarker response
ICAT : isotope-coded affinity tags
ICR : ion cyclotron resonance
IGFL : institut de génomique fonctionnelle de Lyob
ISA : institut de sciences analytiques
iTRAQ : isobaric tags for relative and absolute quantitation
JH : ion cyclotron resonance
JHAMT: juvenile hormone acid methyltransferase
JHBP : juvenile hormone binding protein
JHE : juvenile hormone esterase
JHEBP : juvenile hormone esterase binding protein
JHEH : juvenile hormone epoxide hydrolase
LBD : ligand-binding domain
LC : liquid chromatography
LDS : lithium dodecyl sulfate
LOD : limit of detection
LOQ : limit of quantification
MALDI: matrix-assisted laser desorption/ionization
Met : methoprene
MF : methyl farnesoate
MIH : molt-inhibiting hormone
MOIH : mandibular organ-inhibiting hormone
MRM : multiple reaction monitoring
MS : mass spectrometry
MS/MS: tandem mass spectrometry
MT : metallothionein
MUSCLE: multiple sequence comparison by log-expectation
NaCl : sodium chloride
NCBI : national center for biotechnology information
NMR : nuclear magnetic resonance
NSAF : normalized spectral abundance factor
ORF : open reading frame
OSPAR: Oslo-Paris

PBO : piperonyl butoxide
PCB : polychlorinated biphenyl
PCR : polymerase chain reaction
PFOS : perfluorooctanesulfonic acid
PO : phenoloxidase
POD : peroxydase
PPO : prophenoloxidase
PSAQ : protein standard absolute quantification
PTM : post-translational modification
PYR : pyriproxyfen
QC : quality control
QQQ : triple quadrupole analyzer
QTRAP: triple quadrupole mass spectrometer
REACH: Registration, Evaluation, Authorization and Restriction of Chemicals
RMC : Rhone-Meditéranée-Corse
RNA : ribonucleic acid
RT : retention time
RXR : retinoid X receptor
SD : standard deviation
SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis
SID : stable isotope dilution
SILAC : stable isotope labeling by amino acids in cell culture
SNP : single-nucleotide polymorphism
SOD : superoxide dismutase
SRM : selected reaction monitoring
SWATH: Sequential Window Acquisition of all THeoretical mass spectra
TBT : Tributyltin
TEB : tebufenozide
TG : transglutaminase
TOF : time-of-flight
TPCK : L-1-Tosylamide-2-phenylethyl chloromethyl ketone
TTK : tramtrack
USA : United States of America
VIH : vitellogenesis inhibiting hormone
Vtg : vitellogenin
WES : whole exome sequencing
WFD : Water Framework Directive
WGS : whole genome sequencing
XIC : extracted ion chromatogram

LIST OF FIGURES

Figure I-1 – Principle of a molecular biomarker. In a contaminated situation, the levels of the molecule will be significantly modulated, by either inhibition or induction. 16

Figure I-2 – Schematic representation of xenobiotic (R) detoxification mechanism. 17

Figure I-3 – Free radicals, oxygen reactive species, and oxidative damage to biomolecules. Adapted from (Kumar et al., 2012) 20

Figure I-4 -When absorbed in the body, an endocrine disruptor can decrease or increase normal hormone levels (left), mimic the body's natural hormones (middle), or block the hormone (right) 22

Figure I-5 – General outlines for the construction of an AOP. Adapted from (Ankley et al., 2010) 25

Figure I-6 – The high-throughput omics technologies offer a multitude of options for unraveling components, mechanisms, and regulation of biological processes. GWAS - genome-wide association study; WGS – whole genome sequencing; WES – whole exome sequencing; SNP - single-nucleotide polymorphism; CNV - Copy number variation; PTM – Post-translational modification; miRNA – micro RNA; ncRNA – non-coding RNA; mRNA – messenger RNA. 30

Figure I-7 – General contribution of omics techniques to Ecotoxicology. 32

Figure I-8 – Complexity of the functional repertoire of gene expression products. 35

Figure I-9 – Three major components of a mass spectrometer: the ion source, mass analyzer, and detector. Mass spectrometry is often preceded by some form of chromatography (gas, liquid or thin layer chromatography) to separate analytes of interest before analysis. Adapted from (Wheelock et al., 2013). FTIR stands for "Fourier-transform infrared spectroscopy". 36

Figure I-10- General view of an experimental workflow used in shotgun proteomics analyses. Sample proteins are first digested into peptides. After chromatographic separation, peptides are ionized and fragmented to produce signature tandem mass spectrometry (MS/MS) spectra. Peptides are identified from MS/MS spectra using automated database search programs. Peptide assignments are then statistically validated and incorrect identifications filtered out (example: peptide STHICR). Sequences of the identified peptides are used to infer which proteins are present in the original sample. Some peptides are present in more than one protein (peptide HYFEDR), which can complicate the protein inference process. Adapted from (Nesvizhskii, 2007). 38

Figure I-11 – Proteogenomics for non-model organisms. Typically, a six reading frame translation of the nucleic acid information (either complete or draft DNA-seq or RNA-seq data) generates a first protein database. The MS/MS spectra recorded by tandem mass spectrometry are assigned to peptide sequences and the resulting mass spectrometry-certified proteins can be further structurally and functionally characterized, and thus can improve the protein database. Adapted from (Armengaud et al., 2014) 40

Figure I-12 – The two main approaches used in label-free proteomics. The "area under the curve" strategy is performed at the MS level and peptides that have differential expressions are posteriorly identified at the MS/MS level. Peptide quantitation and identification by "spectral counting" are performed simultaneously at the MS/MS level. (Neilson et al., 2011) 44

Figure I-13 – Protein analysis using the SRM mode with a triple quadrupole. The target protein (in red) is quantified via unique peptides filtered in the first quadrupole. 45

Figure I-14 – Workflow of peptide and transition validation for a SRM-based proteomic experiment. Adapted from (Gallien et al., 2011) 46

Figure I-15 – Workflow of the combination of proteogenomics and targeted proteomics for biomarker development and validation in an environmental scope. (Trapp et al., 2014a) 51

Figure I-16 – Mature male (left) and female (right) organisms of *Gammarus fossarum*. (Source: Hervé Quéau, Irstea Lyon) 53

Figure I-17 – Lateral view of a gammarid (adapted from Chevreux and Fage 1970) 54

Figure I-18- Distribution of *Gammarus fossarum* in Europe. (Barnard and Barnard, 1983) 55

Figure I-19 – Integumental changes of dactylopodite and protopodite from first and second periopod pairs in *Gammarus fossarum* during its molt cycle. (Geffard et al., 2010) 56

Figure I-20 - Synchronization between molt cycle, the embryonic development and vitellogenesis in female *Gammarus fossarum*. Scale bars: 100µm for AB and C1 molting stages and 200µm for C2, D1, and D2 molting stages. FC=follicle cells; EVO=early vitellogenic oocyte;LVO=late vitellogenic oocyte;YV=yolk vesicle;LG=lipid globule. (Geffard et al., 2010) 57

Figure I-21 – Two types of hormonal action (genomic and non-genomic signaling pathways). Hydrophobic hormones diffuse through the membrane and bind to an internal receptor; the hormone-receptor complex binds to the DNA and initiates protein synthesis (left panel). Hydrophylic hormones bind to a cell surface receptor; a seconder messenger is formed through enzymatic-catalyzed reactions within the cell; the second messenger will energize reactions to effect hormonal action (right panel). 58

Figure I-22 – Amphipod regulatory endocrine system. Stimulatory effects are indicated with solid arrows, and inhibitory effects with dashed arrows. The neuropeptides released by the complex [X-organ/sinus gland] will regulate the release of hormones from the target tissues. Although is still uncertain, it is possible that the MF released by the mandibular organ upregulates the Y-organ. Adapted from (Hyne, 2011) 59

Figure I-23 – Simplified pathway of terpenoid synthesis in insects and crustaceans 60

Figure I-24 – Schematic representation of the different functional domains of a nuclear receptor. Adapted from (Aranda and Pascual, 2001). 61

Figure I-25 – Main molecules involved in the ecdysone regulatory cascade in insects. Nuclear receptors are boxed. Adapted from (Bonneton et al., 2008) 61

Figure II-1 – Localization of the source population used in this thesis, in La Tour du Pin. (Xuereb, 2009) 73

Figure II-2 – Organism collection using the fishnet and the sieve. 74

Figure II-3 – Application of the test protocol developed for *G. fossarum* organisms, based on the female molting cycle. (Geffard et al., 2010) 75

Figure II-4 – Schematic representation of the procedures for determining the gammarid size, molt stage, embryo number and stages, oocyte number and apparent surfaces. (Xuereb, 2009) 76

Figure II-5 – Main steps in the male *G. fossarum* gonad dissection protocol. (Lacaze, 2011) 77

Figure II-6 – General experimental setup for the laboratory-controlled contaminant exposures. 79

Figure II-7 – Stages of the PYR exposure protocole during two cycles of spermatogenesis. (Trapp et al., 2015) 80

Figure II-8 – Illustration of gammarid *in situ* caging. 82

Figure II-9 – Workflow used for the protein/peptide selection, and the following analytical validation. In orange is represented the number of initial proteins/peptides selected, and the final number which can be measured simultaneously after analytical validation. 84

Figure II-10 – Main stages of sample preparation for SRM analysis. SPE – *solid phase extraction*. 91

Figure II-11 – Overview of skyline window for analysis of SRM MS/MS peaks..... 94

Figure II-12 – Example of the results obtained from a BLASTP search against the GFOSS theoretical protein database. 98

Figure II-13 – Outline of the RNA extraction procedure using the RNeasy® Fibrous Tissue Mini Kit from QIAGEN. (Source: RNeasy® Fibrous Tissue Mini Kit HANDBOOK)..... 99

Figure II-14 – Protocol used for cDNA synthesis using the SuperScript® III First-Strand Synthesis System from Thermo Fisher Scientific. In this work we followed the protocol with oligo(dT). 100

Figure II-15 – Schematic representation of the strategy used for gene sequencing..... 101

Figure II-16 – Gel image of the amplicons obtained from the temperature gradient PCR for the three target genes. L corresponds to a DNA ladder sizes, and gene name with a "-" means negative control (without cDNA template). 102

Figure II-17 – Overview of CFX Manager window for gene expression data analysis. 104

Figure II-18 - Application of the relative quantification method ΔΔCt for calculating the fold change expression of a reference and target genes, in a control (blue lines) and treated sample (red lines). 105

Figure V-1 – Example of chromatographic peaks obtained for two peptides in fifteen different samples (histograms). The different transitions are represented with different colors (violet, blue and brown). Histograms represent the percentage of contribution of each transition to the total area of the peak. 206

Figure V-2 – Example of the area correlations for three peptides from protein 17046. Data points were obtained from samples analyzed with the QTRAP 4000..... 207

Figure V-3 – Percentage increases of the several Vtg-like proteins at each stage of the female reproductive/molt cycle. 208

Figure V-4 – Molt-cycle-dependent modulation of proteins involved in several key functions in female gammarids. 209

Figure V-5 – Heatmap illustrating the yolk protein levels in 21-day exposed females. Asterisks represent females with significant lower levels of yolk proteins. 211

Figure V-6 – Down-regulated proteins in 21-day exposed male organisms. Red asterisks indicate statistical significant differences between the contaminated and control condition (non-parametric Mann-Whitney t-tests, $p<0.05$). 212

Figure V-7 – Heatmap illustrating the global biomarker responses in male (A) and female (B) organisms. Field sites are represented in columns, and individual biomarker peptides in lines. The color gradient ranges from green (lowest concentrations) to red (highest concentrations). 213

Figure V-8 - Schematic representation of the proportion of low-abundance proteins in human plasma. While a dozen proteins represent more than 90% of the protein content, less than 0.1% is accounted, comprising many proteins of low and very low-abundance. (Righetti and Boschetti, 2013)..... 220

Figure V-9 – Workflow for the discovery and functional validation of *in silico* deduced biomarker candidates. 221

Figure V-10 – Protein sequence alignment in SeaView, using ortholog (same gene for different arthropod species) and paralog sequences (different members of the multigenic family of the target gene) of the candidate sequence Gammarus_fossarum_83504_fr5 (potential RxR). 223

Figure V-11 – Phylogenetic tree of the homologous sequence dataset constructed for the functional validation of the GFOSS BR sequence 1452. 224

Figure V-12 – Expression profiles of E75, BR, and RXR during the female reproductive cycle. Expression values of target genes were normalized to the expression of the reference gene EF. Each bar corresponds to the mean ± SD obtained from the three experimental replicates. 225

Figure V-13 – Fold change gene expressions of E75, RXR, and BR regarding the control condition. Red lines represent the 1.5 fold change threshold. 227

LIST OF TABLES

Table I-1 – *In situ* applications of multibiomarker indices for toxicity assessment. 28

Table I-2 – Labelling strategies commonly used in MS-based protein quantification. Adapted from (DeSouza and Siu, 2013) 42

Table I-3 – Examples of proteomic studies in aquatic ecotoxicology. 49

Table I-4 – Examples of studies employing molecular biomarkers in *Gammarus* species. 63

Table II-1 – Proteins and proteotypic peptides selected for SRM method development. 86

Table II-2 – The ten protein IDs annotated as having chitinase function, and their respective spectral count values from previous shotgun experiments. 89

Table II-3 – Nucleotide sequences of the primers retained for gene expression studies. 103

Table V-1 – Fundamental characteristics of the validated SRM assay. 207

Table V-2 – Statistically modulated proteins after PYR exposure. 217

Table V-3 – Physiological relevant proteins presenting strong fold changes in contaminated samples but needing further tests for validation. 218

Table V-4 – Sequence similarity search results from blastp of RXR and EcR query sequences against GFOSS. 223

Publications

Gouveia D, Silva A, Vitorino R, Domingues M, and Domingues P, The efficiency of trypsin digestion for mass-spectrometry-based identification and quantification of oxidized proteins: evaluation of the digestion of oxidized bovine serum albumin. European Journal of Mass Spectrometry 2014, 20(3), 271–278. DOI: 10.1255/ejms.1279

Charnot A, **Gouveia D,** Armengaud J, Almunia C, Chaumot A, Lemoine J, Geffard O, Salvador A, Multiplexed assay for protein quantitation in the invertebrate Gammarus fossarum by liquid chromatography coupled to tandem mass spectrometry. Analytical and Bioanalytical Chemistry. 2017, 409(16), 3969-3991. doi: 10.1007/s00216-017-0348-0

Charnot A, **Gouveia D,** Ayciriex, S, Lemoine J, Armengaud J, Almunia C, Chaumot A, Geffard O, Salvador A, On-line solid phase extraction liquid chromatography mass spectrometry method for multiplexed proteins quantitation in an ecotoxicology test species: *Gammarus fossarum.* Submitted to Journal of chromatography A, under revision.

Gouveia, D.; Chaumot, A.; Charnot, A.; Queau, H.; Armengaud, J.; Almunia, C.; Salvador, A.; Geffard, O., Assessing the relevance of a multiplexed methodology for proteomic biomarker measurement in the invertebrate species Gammarus fossarum: a physiological and ecotoxicological study. Aquatic Toxicology 2017**,** 190, 199-209. DOI:10.1016/j.aquatox.2017.07.007

Gouveia, D.; Chaumot, A.; Charnot, A.; Almunia, C.; François, A.; Navarro, L. ; Armengaud, J.; Salvador, A.; Geffard, O., Ecotoxico-proteomics for aquatic environmental monitoring: first *in situ* application of a new proteomics-based multibiomarker assay using caged amphipods. Environmental Science and Technology, 2017, *51* (22), 13417-13426. DOI: 10.1021/acs.est.7b03736.

*Trapp, J.; ***Gouveia, D.;** Almunia, C.; Pible, O.; Degli-Esposti, D.; Gaillard, J-C; Chaumot, A.; Geffard, O.; Armengaud, J., Digging deeper the pyriproxyfen-response of the amphipod *Gammarus fossarum* with the next-generation ultra-high-field Orbitrap analyzer: new perspectives for environmental toxicoproteomics. Submitted to Frontiers in Environmental Science. *1st co-authors

Gouveia, D.; Bonneton, F. ; Almunia, C.; Armengaud, J.; Quéau, H. ; Degli-Esposti, D.; Geffard, O. ; Chaumot, A., Identification of three ecdysone-responsive genes (RXR, E75, BR) in the sentinel species *Gammarus fossarum*: expression during female reproductive cycle and modulation after exposure to endocrine-disrupting insecticides. Submitted to Scientific Reports, under revision.

Oral communications

National conferences

Gouveia, D., Charnot, A., Salvador, A., Almunia, C., Armengaud, J., Chaumot, A., Geffard, O. 2015. Approches moléculaires pour la découverte, le développement et l'application de biomarqueurs chez les gammaridés. *2ème Colloque d'Ecophysiologie Animale, La Rochelle, France, 4-6 November.*

Gouveia, D., Bonneton, F., A., Almunia, C., Armengaud, J., Geffard, O., Chaumot, A. 2016. Development of biomarker genes for endocrine disruption assessment in the non-model crustacean species *Gammarus fossarum.* *Colloque de la Société Française d'Ecotoxicologie Fondamentale et Appliquée, Reims, France, 29-30 June.*

Gouveia, D., Charnot, A., Salvador, A., Almunia, C., Armengaud, J., Chaumot, A., Geffard, O., 2017. Protéomique ciblée pour la quantification multiplexée de biomarqueurs protéiques chez le crustacé *Gammarus fossarum* : perspectives pour la surveillance environmentale. *Société Française d'Ecotoxicologie Fondamentale et Appliquée à Lille, 29-30 juin.*

International conferences

Gouveia, D., Charnot, A., Salvador, A., Almunia, C., Armengaud, J., Chaumot, A., Geffard, O. 2016. Application of a multiplexed quantitation of protein biomarkers in the invertebrate species *Gammarus fossarum :* interest to environmental monitoring. *26th Annual Meeting of SETAC-Europe, Nantes, France, 22-26 May.*

Gouveia, D., Charnot, A., Salvador, A., Almunia, C., Armengaud, J., Chaumot, A., Geffard, O. 2016. Multiplexed quantitation of protein biomarkers in the invertebrate species *Gammarus fossarum :* new perspectives for environmental monitoring. *30th Congress of the New European Society for Comparative Physiology and Biochemistry, Barcelona, Spain, 4-7 September.*

Gouveia, D., Charnot, A., Salvador, A., Almunia, C., Armengaud, J., Chaumot, A., Geffard, O. 2016. Shotgun proteogenomics & multiplexed proteomics define robust, specific protein biomarkers from non-model species for environmental monitoring. *SETAC/iEOS Joint Meeting - Environmental and (eco)toxicological Omics and Epigenetics: Science, Technology and Regulatory Applications, Ghent, Belgium, 12-15 September.*

Almunia, C., **Gouveia, D.,** Trapp, J., Cogne, Y., Pible, O., Salvador, A., Chaumot, A., Geffard, O., Armengaud, J., When proteogenomics helps ecotoxicology. *27th Annual Meeting of SETAC-Europe, Brussels, Belgium, 07-11 May*.

Poster presentations

National conferences

Gouveia, D., Bonneton, F., Almunia, C., Armengaud, J., Geffard, O., Chaumot, A., 2017. Expression des récepteurs nucléaires (Rxr, E75) pendant le cycle reproductif du crustacé *Gammarus fossarum* et modulation lors de l'exposition a 3 insecticides perturbateurs endocriniens. *Société Française d'Ecotoxicologie Fondamentale et Appliquée à Lille, 29-30 juin.*

International conferences

Charnot, A., **Gouveia, D.,** Chaumot, A., Lemoine, J., Geffard, O., Salvador, A., 2016. Development of a SRM-based multiplexed quantitation of protein biomarkers in *Gammarus fossarum.* *26th Annual Meeting of SETAC-Europe, Nantes, France, 22-26 May.*

Gouveia, D., Bonneton, F., A., Almunia, C., Armengaud, J., Geffard, O., Chaumot, A., 2017. Development of biomarker genes for endocrine disruption assessment in the non-model crustacean species *Gammarus fossarum.* *30th Congress of the New European Society for Comparative Physiology and Biochemistry, Barcelona, Spain, 4-7 September.*

Gouveia, D., Trapp, J., Geffard, O., Chaumot, A., Almunia, C., Armengaud, J., 2017. Proteome analysis of male gonads in Gammarus fossarum exposed to Pyriproxyfen: mining for endocrine disruption biomarkers. *27th Annual Meeting of SETAC-Europe, Brussels, Belgium, 07-11 May*.

Gouveia, D., Bonneton, F., Almunia, C., Armengaud, J., Geffard, O., Chaumot, A., 2017. Development of biomarker genes for endocrine disruption assessment in the non-model crustacean species *Gammarus fossarum.* *27th Annual Meeting of SETAC-Europe, Brussels, Belgium, 07-11 May*.

Présentation du sujet de thèse

Ce projet de thèse s'est centré sur le développement et la proposition de nouveaux biomarqueurs moléculaires spécifiques du dysfonctionnement de grandes fonctions physiologiques chez l'amphipode *Gammarus fossarum*. Cette espèce est utilisée depuis plusieurs années par le laboratoire d'écotoxicologie à Irstea Lyon. Cette thèse s'inscrit dans les travaux engagés il y a déjà plusieurs années à Irstea sur le diagnostic de la contamination et de la toxicité des milieux, notamment via l'utilisation du gammare (biomonitoring actif par encagement).

De nombreuses cibles biologiques telles que des hormones, des enzymes et autres protéines, ont déjà été validés comme outils d'évaluation des effets des contaminants dans les milieux aquatiques. Cependant, ces outils ont principalement été développés chez les vertébrés, pour qui l'information moléculaire est aujourd'hui facilement accessible. Le développement de biomarqueurs moléculaires chez les invertébrés d'intérêt en écotoxicologie n'a connu que peu d'avancées ces dernières années, du fait de l'existence de plusieurs verrous scientifiques et techniques. Parmi les verrous scientifiques, un des plus limitants est l'absence des méthodes spécifiques pour la quantification spécifique des biomarqueurs, notamment liés à des perturbations endocriniennes (PE). Grâce à sa capacité d'interférence avec le fonctionnement normal du système endocrinien, responsable pour le contrôle des fonctions biologiques clés comme la reproduction et développement des organismes, les perturbateurs endocriniens sont une préoccupation majeure au niveau mondiale. Cependant, le manque de connaissance sur l'endocrinologie des invertébrés a bloqué le développement de biomarqueurs spécifiques de PE pour ces espèces. De plus, les effets des perturbateurs endocriniens sont souvent difficiles à expliquer parce qu'ils n'agissent pas de la même façon entre différents taxons.

Pour surmonter ces difficultés, une nouvelle approche dite de « protéogénomique », couplant séquençage du transcriptome (RNAseq) et caractérisation du protéome par spectrométrie de masse, a été mise en œuvre chez *Gammarus fossarum* en collaboration avec le laboratoire Innovations technologiques

pour la Détection et le Diagnostic du CEA (Commissariat pour l'Energie Atomique de Marcoule). Ces travaux ont permis d'obtenir une base de données de séquences de gènes codants chez notre espèce et un catalogue de protéines pour différents tissus (1873 protéines), dont les gonades. Ces travaux sont d'une grande importance, puisqu'ils permettent aujourd'hui l'analyse directe des gènes et ses produits, comme les ARNm et les protéines. Alliés aux capacités haut-débit des nouvelles technologies et instrumentation, ces outils permettent l'identification des modes d'actions des contaminants, fournissant des signatures moléculaires d'effets spécifiques comme la PE. Ces méthodes mécanistiques ouvrent de nombreuses opportunités pour générer des connaissances inestimables sur les effets et les réponses des systèmes biologiques au stress environnemental. Ils permettent aussi d'envisager, chez un arthropode (élément majeur de la biodiversité aquatique, mais peu représenté parmi les espèces modèles pour l'étude des perturbations endocriniennes environnementales), le développement de méthodes spécifiques pour quantifier ces protéines et évaluer leur intérêt comme biomarqueurs de la perturbation, notamment de la fonction de reproduction, chez *G. fossarum.*

Dans mon projet de thèse, j'ai utilisé les catalogues de protéines et de gènes disponibles pour développer et valider des biomarqueurs moléculaires pour le diagnostic de perturbations toxiques chez *G. fossarum.* J'ai organisé ma thèse en deux parties :

1) Méthode de quantification multi-biomarqueurs - cette partie de ma thèse résulte d'un partenariat avec l'Institut de Sciences Analytiques (ISA) de Lyon, et a eu pour objectif le développement d'une stratégie analytique pour quantifier simultanément plusieurs protéines (entre 20 et 40 protéines) par spectrométrie de masse (MS). Cette méthodologie innovante dans le domaine de l'écotoxicologie permet, en une seule analyse, de suivre les concentrations de différents biomarqueurs et de construire des signatures moléculaires (1) en réponse à un type de contaminants et (2) en lien avec de grandes fonctions physiologiques devant nous permettre de proposer à terme un indicateur de l'état de santé.

2) Biomarqueurs de perturbations endocriniennes – cette partie constitue la suite du partenariat avec le CEA de Marcoule et a visé le développement de biomarqueurs spécifiques d'une PE. Deux approches ont été suivies pour répondre à cet objectif. La première approche a consisté à utiliser un spectromètre de masse dernière génération dans une étude de protéomique comparative. Cette étude a eu pour objectif de découvrir les molécules modulées lors d'une exposition des organismes à un perturbateur endocrinien modèle, le pyriproxyfen, et de proposer des biomarqueurs. La deuxième approche, via l'exploitation du transcriptome et de la littérature, a consisté à déterminer des gènes d'intérêt (non retrouvés encore dans nos catalogues de protéines) liés aux processus hormonaux et à les séquencer chez *Gammarus fossarum*. Cette approche a eu pour objectif d'identifier des acteurs clés impliqués dans les processus reproductifs de la femelle, et à tester leur pertinence comme biomarqueurs de PE chez *G. fossarum*.

Méthodologie développée et principaux résultats

Pour la première partie du travail (méthode multibiomarqueurs), j'ai fait une sélection de protéines candidates (177 peptides appartenant à 55 protéines) qui jouent un rôle clé dans les différents processus physiologiques (mue, régulation hormonale, reproduction, immunité), ainsi que des protéines déjà largement utilisées comme biomarqueur en écotoxicologie (ex catalase, SOD, GST,…). Dans une première étape, la méthode analytique par chromatographie liquide couplée à la spectrométrie de masse « selected reaction monitoring » (SRM) a été développée pour une quantification multiplex de protéines ciblées. Parmi la liste initiale, nous avons été capables d'identifier 71 peptides rapporteurs au total, appartenant à 40 protéines. La méthodologie développée a ensuite été utilisée pour valider la spécificité et la fonction physiologique des peptides choisis, à travers l'établissement de profils protéiques au cours de processus physiologiques connus et décrits chez cette espèce (spermatogenèse, ovogenèse, privation alimentaire). Ultérieurement, une étape de validation des biomarqueurs via des contaminations au labo et des expositions en

milieu naturel (encagement) a été faite. Nous avons fait des contaminations au laboratoire avec des concentrations environnementales pour deux contaminants modèles (cadmium et plomb). Après exposition, l'impact de ces deux métaux a été évalué à l'aide de marqueurs individuels (trait de vie) disponibles, comme la reproduction et par l'analyse en MS d'un ensemble de 20 protéines identifiées précédemment. Des retards de mue ont été observés chez 20% des individus, et le niveau de 7 des protéines recherchées a été modulé. Ces travaux ont permis de valider la preuve de concept et l'intérêt d'une approche multiplexée chez un invertébré pour la mesure de plusieurs marqueurs protéiques.

Pour la validation sur le terrain, des organismes ont été exposés sur des rivières soumises à différentes pressions polluantes dans la cadre du projet Agence de l'eau RMC. Des modulations ont été observées pour plusieurs protéines, en accord avec la pression chimique définie *a priori* par les agences pour ces sites, montrant la pertinence et la sensibilité d'une telle démarche pour le développement de nouveaux biomarqueurs et leur utilisation dans la surveillance. Des modulations des niveaux de vitéllogenines ont été observées chez les mâles (inductions) et les femelles (inhibitions), re-ouvrant la question de l'utilisation de ces protéines comme biomarqueurs chez les invertébrés. De plus, deux protéines liées à la mue ont présenté des inhibitions très fortes sur des sites contaminés par les pesticides. Finalement, d'autres protéines liées à des processus comme la détoxification et immunité ont présenté des modulations significatives. Comme pour les marqueurs « classiques » ou « ancienne génération », les perspectives du travail doivent maintenant s'attacher à la proposition de valeurs seuils, intégrant la variabilité naturelle de ces réponses, pour conduire à une interprétation fiable de leur modulation en termes de contamination et/ou de toxicité, dans des milieux très contrastés

Dans le deuxième axe de la thèse, l'étude de protéomique comparative a permis de mettre en évidence les avantages d'utiliser des spectromètres de masse récents (donc plus sensibles) pour l'obtention des listes de candidats biomarqueurs d'un stress PE. En analysant le protéome des gonades des mâles exposés au pyriproxyfen, nous avons détecté 53 protéines avec une modulation significative par rapport à la condition témoin. Ces résultats ont mis en évidence les voies métaboliques impactées par une

exposition au pyriproxyfen (liées à l'homéostasie cellulaire et formation du cytosquelette), et ont permis d'identifier et cibler quelques candidats biomarqueurs pour une validation postérieure via des approches de protéomique ciblée.

Pour la deuxième approche du travail, j'ai fait une liste d'environ 30 candidats biomarqueurs d'une PE. Cette liste comprend des protéines ayant un rôle essentiel dans la régulation hormonale des processus reproductifs chez le mâle et la femelle, ceci à partir de la littérature et de l'exploitation de la base de données transcriptomique chez *G. fossarum*. Pour cela, j'ai recherché dans la littérature des gènes impliqués dans les voies hormonales (récepteurs hormonaux, enzymes du métabolisme hormonale, hormones, gènes de détermination du sexe, et gènes qui régulent ou sont régulés par des hormones) chez les arthropodes (insectes, crustacés,…). J'ai réalisé ensuite une analyse par blast pour retrouver ces gènes chez notre espèce (blasts chez ncbi et sur notre base de données du transcriptome de *G. fossarum*), puis effectué une étape de validation de l'annotation du gène/protéine par des approches phylogénétiques.

Par cette approche, trois gènes clefs, dont l'annotation fonctionnelle probable a été validée, ont été identifiés (RXR, Broad-complex et E75). Pour ces trois récepteurs, nous sommes partis sur une approche de clonage de gènes pour avoir une séquence propre. Aujourd'hui nous avons les séquences spécifiques pour les trois gènes. Ces travaux ont été menés en collaboration avec l'Institut de Génomique Fonctionnelle de Lyon (IGFL) et le CEA à Marcoule. La validation fonctionnelle et la pertinence de ces gènes comme biomarqueurs ont été évaluées au niveau transcrit. Pour cela, nous avons étudié leur expression au cours du cycle reproductif femelle et suite à l'exposition d'organismes à des molécules modèles de type PE (Pyriproxyfène – analogue de l'hormone juvénile ; Piperonyl butoxide – synergiste de plusieurs pesticides ; Tebufénozide – agoniste du récepteur d'ecdysone). Les transcrits de RxR et E75 ont été modulés pendant le cycle de la femelle. RxR présente un pic d'expression en début de cycle, tandis que E75 présente un pic très intense au stade C2/D1 (correspondant à la mise en place de la vitellogénèse secondaire) et un second pic au dernier stade D2 juste avant la mue. Lors de la contamination, une forte inhibition du pic d'E75 chez les femelles en satde de mue C2/D1 s'est produite avec le tebufénozide. Les résultats obtenus dans cette étude ont ainsi confirmé l'implication de ces gènes dans le cycle de reproduction de la femelle *G. fossarum* et montré leur sensibilité à des contaminants

chimiques de type perturbateurs endocriniens. Ces travaux ouvrent la possibilité d'utiliser l'expression de RxR et E75 comme marqueurs spécifiques d'une exposition de cette espèce à des PE.

Principaux Conclusions

En conclusion, dans ces travaux nous avons adressé deux problématiques importantes dans le développement de biomarqueurs chez des espèces non-modèles : le manque de biomarqueurs spécifiques à l'espèce d'intérêt et de méthodes de quantification directe de (multi)biomarqueurs. En se basant sur les études protéogénomiques développées précédemment, on a ciblé l'identification et validation de nouveaux biomarqueurs de toxicité pour une utilisation en programmes de biomonitoring.

L'approche multi-biomarqueurs innovante développée dans la première partie de la thèse est la première méthodologie proposée pour la quantification simultanée de plusieurs biomarqueurs spécifiques en écotoxicologie chez des invertébrés. L'essai permet la détection simultanée de 20 à 40 biomarqueurs protéiques, en fonction du spectromètre de masse utilisé. L'application de cet essai dans le cadre d'une étude sur le terrain a montré des résultats très satisfaisants. Grâce à sa combinaison avec la stratégie d'encagement développée chez *G. fossarum*, les sources de variabilité liées aux facteurs de confusion biotiques ont été réduites et le réalisme environnemental des conditions d'exposition a été amélioré. Malgré quelque variabilité interindividuelle, nous avons observé des réponses claires de biomarqueurs en réponse à une contamination, et les capacités haut-débit ont permis de mesurer ces réponses en un temps très court. D'un point de vue biosurveillance, il s'agit d'une avancée majeure car ces travaux montrent qu'il est possible de déterminer rapidement l'état de santé d'organismes. Ces travaux ont démontré que l'approche multibiomarqueur SRM est une stratégie prometteuse pour une application en tant qu'outil pour la surveillance réglementaire.

Avec des biomarqueurs pertinents, rapporteurs de fonctions physiologiques majeures, ce type d'approches à haut débit peut fournir des informations sur le type de contaminants présents dans les milieux d'exposition à travers de ces signatures. Dans ce

contexte, le deuxième axe de cette thèse portait spécifiquement sur la perturbation endocrinienne, une des préoccupations majeures pour l'évaluation des risques environnementaux. Les travaux réalisés ici ont permis d'acquérir des connaissances importantes sur la régulation endocrinienne et sa perturbation de *Gammarus fossarum*. L'étude en protéomique shotgun effectuée sur les gonades mâles a démontré la grande évolution dans la technologie des spectromètres de masse. Avec les 4031 protéines détectées chez une espèce non-modèle, cette étude nous a permis d'élargir le catalogue de protéines pour notre espèce de référence *G. fossarum*, qui contenait précédemment 1873 protéines. Le fait que nous puissions suivre simultanément des milliers de protéines ouvre également la voie à de futures études visant à déterminer les modes d'action des contaminants, les voies affectées, ainsi qu'à étudier la diversité inter-populations et inter-espèces pouvant expliquer, par exemple, les mécanismes d'adaptation à la contamination. Les protéines mâles modulées après l'exposition au pyriproxyfen ont également permis de proposer des biomarqueurs potentiels de dysfonctionnement de la reproduction.

Au cours de cette thèse, nous avons également développé une stratégie alternative pour la découverte et la détection de molécules peu abondantes, mais d'intérêt comme biomarqueurs de PE. La mise en œuvre de cette stratégie nous a permis de découvrir et de valider, pour la première fois chez des gammares, la fonction de gènes clés impliqués dans la voie ecdysone. Ces travaux ont souligné l'importance d'utiliser des espèces sentinelles dont la physiologie est largement connue, ceci afin de comprendre / valider la fonction des biomarqueurs candidats. En raison du manque de connaissances du système endocrinien chez la majorité des crustacés, nous avons également observé des annotations fonctionnelles erronées dans les bases de données publiques, liées à la limite de l'annotation par homologie de séquences. Nous proposons donc d'utiliser la phylogénie moléculaire comme outil de validation de l'annotation fonctionnelle automatique. Afin de bien comprendre la base moléculaire des processus hormonaux chez les gammares, des études mécanistiques et fonctionnelles plus approfondies doivent encore être réalisées.

Aquatic systems are continuously loaded with anthropogenic contaminants released from several sources such as industry, agriculture, urbanization, transport, tourism, and everyday life. Around 100.000 molecules are rejected in aquatic systems, the majority being very persistent and accumulating in organisms and in food chains. Consequently, aquatic organisms are increasingly being exposed to this multiple and complex contamination. Contaminants can induce short- and long-term biological effects, constituting a permanent risk from an ecological and sanitary point of view.

Seen the importance of aquatic systems to human well-being, several countries have enforced specific regulations to the protection, restoration, and management of ecosystems. In Europe, legislations have been adopted to improve the status of aquatic ecosystems mainly through the "Registration, Evaluation and Authorization of Chemicals" (REACH) and the European Comission "Water Framework Directive" (WFD). These regulations aim at developing tools for predicting the impact of contaminants and diagnosing chemical contamination/toxicity of aquatic systems. The WFD (n°2000/60/CE) was implemented in 2000 in order to homogenize the management of watercourses among the European Union (EU) member countries. The Directive aims for good chemical and ecological status for all ground and surface waters (rivers, lakes, transitional waters, and coastal waters) in the EU. This "good status" is achieved through the compliance with environmental quality standards. These standards comprise maximum concentration values for prioritized pollutants in water or biota, and biotic indices built on the presence and abundance of bioindicator species (diatoms, macrophytes, macroinvertebrates, and fishes) that must fall into good quality levels. Chemical analysis allows to measure pollutants present in several matrices such as waters and effluents, sediment and leachates, organismal tissue and fluids. Bioindicators give information about the global evaluation of the ecosystem, but do not allow identifying the toxicity processes or discerning between physical, biological, or chemical pressures. Moreover, these tools do not allow for precocious evaluations of the impact of a pollution in the ecosystem.

For these reasons, biological responses that can be assessed at the individual and sub-individual levels are of high importance to identify, understand, and discriminate

the effects of pollutants from other stressors. Biomarkers are recognized as relevant tools for diagnostic and hazard assessment of aquatic systems (Amiard-Triquet and Berthet, 2015). Ideal biomarkers represent early warning indicators of relevant ecological impairments due to an exposure of organisms to toxic chemicals (Forbes et al., 2006). Biomarker responses at molecular levels are increasingly being used in the last decades to evaluate exposure of sentinel organisms to several types of contaminants (Hook et al., 2014a; Sarkar et al., 2006). Molecular biomarkers can provide early assessments of pollutant effects in the health of organisms or environmental disturbances, and classic molecular biomarkers comprise mainly the measurement of the activities of several key physiological enzymes (Ahn et al., 2009; Jemec et al., 2010; Picado et al., 2007). One single biomarker is not representative of the entire set of modes of action and specific effects of contaminants present in ecosystems that can disturb the health status of organisms. In order to take into account contaminant diversity and the multiplicity of effects they can exert, it is necessary to use multibiomarker approaches based on the measurement of several complementary biomarkers covering a broad range of effects (Galloway et al., 2004; Minier et al., 2000; Serafim et al., 2011). Due to the major developments in molecular biology, nowadays there are tools that allow direct analysis of gene and/or gene products such as mRNA and proteins. Allied with the high-throughput capacities of new technologies and instrumentation, these tools allow determining contaminants modes of action, thus providing molecular signatures of specific effects such as endocrine disruption. Such mechanistic methods are opening many opportunities for generating invaluable knowledge on the effects and responses to environmental stressors upon biological systems (Connon et al., 2012).

Due to their capacity of interfering with the normal functioning of the endocrine system, which controls key biological functions of an organism such as reproduction and development, endocrine disruptors (ED) are a major scientific concern worldwide. These contaminants are known to have deleterious effects in several species, including humans. Some examples include the masculinization of female gasteropods (imposex) due to an exposure to tributyltin (Horiguchi et al., 1995), or the feminization of male fishes due to the action of oestrogenic compounds (Sumpter and Jobling, 1995). Endocrine disruption modes of action and associated effects were extensively studied in

aquatic vertebrate species, leading to a solid body of knowledge and the development of robust, specific, and reliable biomarkers for risk assessment. Vitellogenin (Vtg) induction in male fishes is the most commonly employed biomarker of ED. However, despite the obvious importance of invertebrates and crustaceans in aquatic systems, the effects and/or hazard assessment of ED remains difficult to establish. To date, no specific ED biomarkers are available in crustacean species currently used in ecotoxicology. This can be attributed mostly to the lack of fundamental comprehensive knowledge of their endocrine systems, but also to the fact that relevant proteins used as specific biomarkers in vertebrates are not conserved in invertebrates. In order to progress in understanding the impacts of ED in ecosystems, one must account for species diversity and gain mechanistic insights of these impacts in each species of interest. Nowadays, the advances in genome and proteome sequencing allow for rapid and large-scale discovery of new proteins in non-model but key sentinel species in environmental sciences.

In this context, the works presented in this manuscript approached two major limitations in biomarker application for biomonitoring. First, the lack of high-throughput multibiomarker methodologies that can be applied in biomonitoring; and second, the lack of specific biomarkers in non-model but relevant sentinel species. The objective was to use new omics approaches in order to develop and apply new general and ED-specific biomarkers of toxicity in the sentinel non-model amphipod *Gammarus fossarum*. The thesis is structured in five major chapters. Chapter I consists in a literature review to expose the current state of the art in the use of protein biomarkers in ecotoxicology, as well as the advantages of using the new technological advances for biomarker development. Chapter II details the experimental procedures implemented for answering the objectives of the thesis, notably in terms of strategy design, analytical techniques, and laboratory/field experiments. The results obtained from these experimentations are presented under the form of articles in chapters III and IV. Chapter V synthesizes the main results obtained during these studies, followed by a presentation of main conclusions and future perspectives.

This first chapter is a literature review of the current use of molecular biomarkers in ecotoxicology and environmental assessment of water pollution, with a strong emphasis on protein biomarkers and on the use of non-model sentinel species. Section 1 gives the definition of biomarkers, and exemplifies the most commonly employed protein biomarkers for evaluating the impact of pollutants. Limitations for the application of the biomarkers in routine biomonitoring are also exposed. Section 2 explores the advantages presented by the new "omic" approaches for biomarker development in ecotoxicology, highlighting the strong potential of proteomic methodologies for multi-biomarker measurement. Section 3 briefly characterizes the sentinel species used in our study, its putative neuroendocrine regulation processes, and currently employed molecular biomarkers. Finally, the last section of this chapter lists the several objectives proposed for this thesis.

1. Protein biomarkers in Ecotoxicology and multi-biomarker approaches

1.1. Biomarkers: definition, classification, examples and limitations

The detection of pollutants in aquatic environments does not indicate a deleterious effect in organisms and/or ecosystems. To evaluate the impact of pollutants, assessments of biological responses in organisms must be done. Biomarkers in response to stress have become an important tool in environmental quality evaluation and risk assessment. They were originated in human toxicology (Timbrell, 1998), where they are routinely used to diagnose exposures to specific chemicals and/or diseases. Tools developed in medicine are being borrowed by ecotoxicologists and applied for predicting ecosystem health. One of the most important advantages of biomarkers is that they can provide information on the biological effects of a pollutant (Ghisi, 2012). Due to the multitude of works developed around the last thirty years, several definitions of biomarker emerged. In the literature, the most commonly used definition is the one by Depledge (1994) (Depledge, 1994), in which a biomarker is defined as being a "biochemical, cellular, physiological or behavioral variation that can be measured in tissue or body fluid samples or at the level of whole organisms that provides evidence of exposure to and/or effects of, one or more chemical pollutants".

Biomarkers can reflect either an exposure to environmental stressors, and/or the adverse health effects resulting from pollutant exposure. Some biomarkers can fit into these two categories, the most known example being the acetylcholinesterase (AChE). Biomarkers of exposure are normally specific to a particular contaminant or class of contaminants, as for example the (ALAD) enzyme, inhibited only by lead, the AchE inhibition by organophosphorus and carbamate pesticides, or vitellogenin induction in response to estrogens in vertebrate species. Biomarkers of effect indicate the adverse effects on organisms provoked by exposure to pollutants, and tend to be less specific than exposure biomarkers. Both specific and nonspecific biomarkers are important in hazard assessment, providing complementary information. For example, ALAD provides information about lead poisoning but does not provide information on the presence of other pollutants. Monooxygenase inductions, provoked by a wide range of pollutants, are an excellent indicator of organism exposure to pollutants, but without revealing a specific cause.

In ecotoxicology, sublethal biomarkers that can be measured at low levels of biological organization (molecular biomarkers) represent early and sensitive indicators of chemical stress (Smit et al., 2009). Pollutants can trigger a cascade of biological responses that can be used as biomarkers (Figure I-1). This represents a big advantage over responses at higher levels because the latter are detectable only when significant damage has occurred. Furthermore, biomarkers can also provide relevant insights into the modes of action of contaminants, attribute that can be very useful when analyzing complex mixtures of pollutants and/or when extrapolating effects across different species (Miller et al., 2007). Biomarkers can have different natures, such as damages to genetic material, lipids, and proteins, or formation/modulation of metabolic compounds, increased oxidation status, enzymatic responses, and membrane integrity. Enzyme activities and other sub-cellular components are among the most commonly used biochemical and molecular biomarkers.

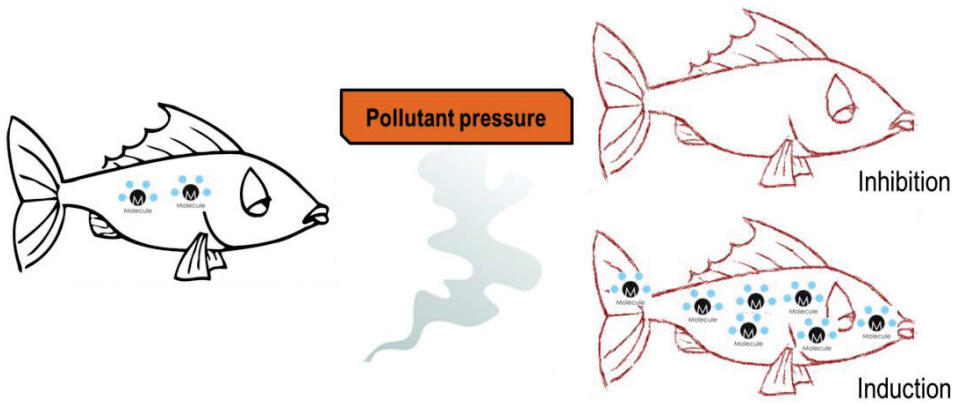


Figure I-1 – Principle of a molecular biomarker. In a contaminated situation, the levels of the molecule will be significantly modulated, by either inhibition or induction.

1.1.1. Examples of protein biomarkers used in animal Ecotoxicology

CYP1A induction

Some biochemical processes transform xenobiotics present in the cells, facilitating its removal from the organism. By reducing the concentrations of xenobiotics in cells, these processes prevent further interactions that could cause cellular damage. The detoxification process is often divided in two main phases as represented in Figure

I-2. A group of xenobiotic metabolizing enzymes that have a primary role in protecting the organism characterizes each phase. Phase I reactions involve hydrolysis, reduction/oxidation of xenobiotics, and normally result in an increase in reactivity and hydrophilicity of the metabolized product (Parkinson, 2008). The monooxygenases containing cytochrome P450 (CYP) as their catalytic center are the major Phase I group of enzymes. The CYP system comprises a family of structurally and functionally related heme proteins, recognized as being important in the oxidative metabolism of diverse drugs and xenobiotics, but also several endogenous compounds (Cajaraville et al., 2000).

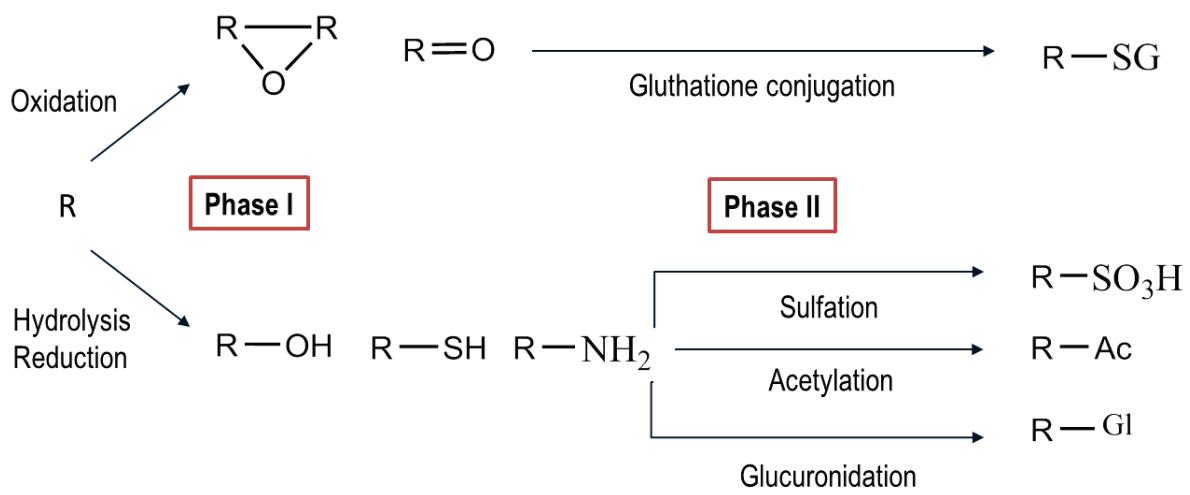


Figure I-2 – Schematic representation of xenobiotic (R) detoxification mechanism.

Biochemical shifts linked with the induction of CYP monooxygenases are associated to a wide range of organic pollutants such as dioxins, furans, polychlorinated biphenyls, and polycyclic aromatic hydrocarbons. The most common used biomarker is the CYP1A subfamily, whose induction is measured at either the mRNA or protein level (Courtenay et al., 1993; Goksøyr and Förlin, 1992), or catalytic activity through the enzymes EROD (ethoxyresofurin O-deethylase) and AHH (aryl hydrocarbon hydroxylase) (Whyte et al., 2000). Xenobiotics induce CYP1A through binding to the arylhydrocarbon receptor (AhR) in vertebrates. It has been shown that EROD activity provides good discriminating abilities in trout fishes exposed to different pollution-pressured streams (Behrens and Segner, 2005).

Metallothionein induction

Another mechanism through which the organism reduces the bioavailability of xenobiotics is **binding** to another molecule to excrete or storage the product. For example, elevated levels of both essential and nonessential metals induce the metallothionein (MT) proteins. These are small heavy metal binding proteins (around 6kDa), rich in cysteine residues, that participate in the normal homeostasis of essential metals by binding to elements such as copper and zinc. They have two globular subunits, each with several cysteine residues that do not form disulfide bonds between them, and whose capacity for sequestering metals comes from their sulfhydryl groups (Shariati and Shariati, 2011). MTs also have the capacity of sequestrating toxic metal ions such as cadmium, mercury, and silver, playing an important role in detoxification processes by lowering the cellular concentrations of these metals (Cobbett and Goldsbrough, 2002; Shariati and Shariati, 2011). MTs are therefore commonly proposed as biomarkers of exposure to non-essential and essential metals (Amiard et al., 2006). Due to its capacity to detoxify metals, MTs are promptly induced when there is an increase in metal concentrations in several organisms (Benedicto et al., 2005; Geret and Cosson, 2000; Ladhar-Chaabouni et al., 2012; Martinez et al., 1996; Pedersen et al., 1997). Due to their wide use as biomarkers, there are several analytical methods for MT determination, such as electrochemical methods, metal saturation, spectrophotometry, chromatography (sometimes coupled to mass spectrometry), immunological methods, electrophoresis and mRNA measurements (detailed in (Shariati and Shariati, 2011)). Moreover, MTs are highly conserved between different animal groups, and were already reported in several aquatic invertebrates, especially molluscs, crabs, and crustaceans. The fact that we find these proteins along different taxa facilitates its transfer as a biomarker.

Acetylcholinesterase

Contaminants also exert strong effects on a wide range of proteins that participate in other molecular mechanisms. Through modification of protein structures and stability, the associated protein-mediated processes will be affected. Enzymes, being the catalysts of all metabolic processes, are particularly susceptible to be influenced by contaminants. Among all the affected enzymes, one of the most known is the specific

effect of organophosphorus and carbamate insecticides in the activity of AChE (Fulton and Key, 2001). AChE catalyzes the breakdown of acetylcholine into choline and acetic acid. This enzyme has a very important role in nerve conduction processes at the neuromuscular junction. The mechanism of action of organophosphorus and carbamate pesticides is through inhibition of the enzyme and accumulation of acetylcholine at the nerve synapses, which will cause deleterious effects leading eventually to death. Therefore, AChE activity is one of the most used biomarkers of both exposure and effect to these kinds of compounds (Fulton and Key, 2001), and allows detecting sub-lethal toxicological effects before the appearance of clinical symptoms (Ghisi, 2012).

Biomarkers of oxidative stress

The organisms response mechanisms to chemical pollutants are important sources of free radicals in biological species, whose abnormal levels in the cell can lead to an **oxidative stress**. Free radicals are often non-specific regarding its biochemical targets, and can damage all components of the cell, including proteins (protein dysfunction through oxidation), lipids (membrane dysfunction through lipid peroxidation), and DNA (oxidative damage can lead to mutations) (Figure I-3). Proteins, being the major component of most biological systems and having high rate constants for reactions, are a major target for oxidants (Davies, 2005). Xenobiotics can increase oxidative damage through many different processes, either by interfering with the normal antioxidant mechanisms, or by directly participating in reactions leading to oxidative stress (such as reactions from phase I and II detoxification system) (Di Giulio, 1995; Lushchak, 2011; Newman, 2015). Cellular defense against these radicals is provided by the production of antioxidants (ex. vitamin E, vitamin C, glutathione) or antioxidant enzymes (ex. superoxide dismutase SOD, catalase CAT, glutathione peroxidase GPX) that reduce the amount of circulating free radicals.

The application of biomarkers of oxidative stress increased considerably after verifying the importance of free radical damage in the mechanisms of toxicity of chemical pollutants. Molecular biomarkers are often used to test an oxidative stress in sentinel organisms (Enis Yonar et al., 2011; Nunes et al., 2006; Radwan et al., 2010; Sroda and Cossu-Leguille, 2011), mainly by analyzing the activities of enzymatic

antioxidants like SOD, CAT and GPX, or other non-enzymatic biomarkers such as glutathione, vitamin E, and ascorbate.

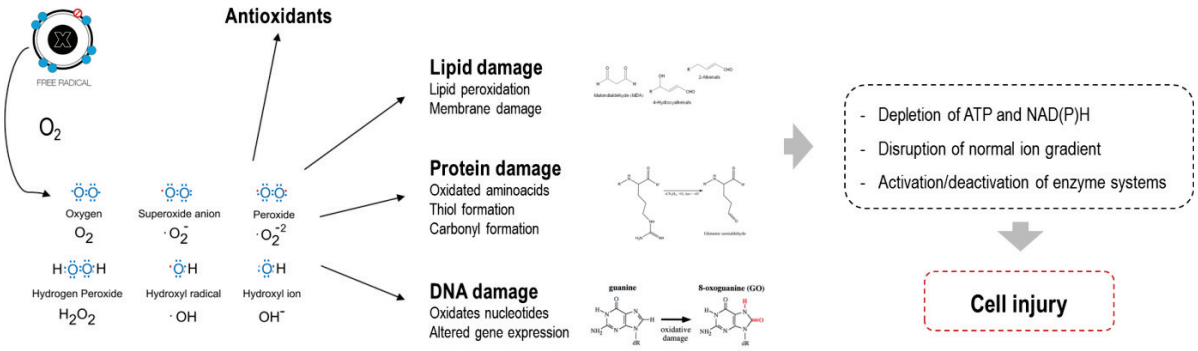


Figure I-3 – Free radicals, oxygen reactive species, and oxidative damage to biomolecules. Adapted from (Kumar et al., 2012)

Vitellogenin induction

Vitellogenin is the precursor protein of the egg yolk proteins, and is expressed in females of almost all oviparous animal species. Due to the action of endocrine disrupting compounds (EDCs), male fishes can express the Vtg gene in a dose dependent manner in contaminated conditions. Vtg induction is therefore a widely used biomarker of exposure to environmental estrogen mimetics, especially in fish species (Garcia-Reyero et al., 2004; Sumpter and Jobling, 1995; Tyler et al., 1996). This induction can be measured both at the protein (Carnevali et al., 2003; Lomax et al., 1998) and mRNA level (Garcia-Reyero et al., 2004).

1.1.2. Limitations and perspectives

Upon entry into an organism, all toxicants will firstly interact with biomolecules, potentially changing their structural and/or functional properties. The understanding of these interactions can provide invaluable insights into the origin of the cause of the effects observed at the individual level. Molecular interactions between pollutants and sites of action are varied and can be specific to a type of chemical or to certain phylogenetic groups of organisms. Species specificity occurs when one species possesses

a certain site of action that is not present in other species. Pesticides that act in the nervous systems of their target species will not exert any effect in plants, or pesticides that disturb chitin exoskeleton formation will only act on chitin-dependent arthropods. On the other hand, non-specific effects may occur in nearly all organisms and/or in response to a wide range of chemicals (Escher and Hermens, 2002).

Cross-species transferability: illustration with endocrine disruption biomarkers

These specific and non-specific effects of pollutants account for several limitations in biomarker development, especially for their transferability across different taxa. The best example is the one of EDCs, which have a well-known mode of action, but in a restricted phylogenetic group (*e.g.* synthetic estrogens in vertebrates, juvenoid insecticides in insects, …). EDCs are exogenous substances that interfere with hormone-regulated physiological processes and provoke adverse health effects in exposed organisms and/or in its progeny (Birnbaum, 2013). EDCs will typically interfere with hormone signaling, acting as agonist or antagonist, or with hormone synthesis, through anti-hormonal effects for example (LeBlanc, 2007; Rodriguez et al., 2007) (Figure I-4). These compounds constitute a worldwide concern for potential human health implications (Cravedi et al., 2007), especially due to the multiple developmental and reproductive disorders observed in wildlife (Crisp et al., 1998; Ford et al., 2004; Höss and Weltje, 2007; Jobling and Tyler, 2003; Kloas et al., 2009; Oehlmann et al., 2007; Oetken et al., 2004; Rodríguez et al., 2007; Soin and Smagghe, 2007). One of the particularities of these compounds is that they can exert toxic effects at much lower concentrations that other types of toxicants. Due to the large amount of discharges in aquatic systems (from pulp and paper mills, agricultural practices, pharmaceutical use etc.), these compounds are prevalent in these environments. Therefore, some of the best-documented examples of the effects of these compounds in wildlife are for aquatic vertebrates, especially fishes (Jobling and Tyler, 2003; Tyler et al., 1998). However, regardless of the species, the reproductive processes are always connected to the coordination of the hormonal system: in fish through estradiol and testosterone; in arthropods mainly through ecdysone, juvenile hormone, and peptide hormones. Therefore, interactions between EDCs and hormone-related molecules of living organisms will lead to consequences such as, for example, male feminization events

(Ford et al., 2004; Sumpter and Jobling, 1995; Yuan et al., 2013), reproductive impairments (Jobling et al., 1996), and intersex incidence (Depiereux et al., 2014; Ford et al., 2004).

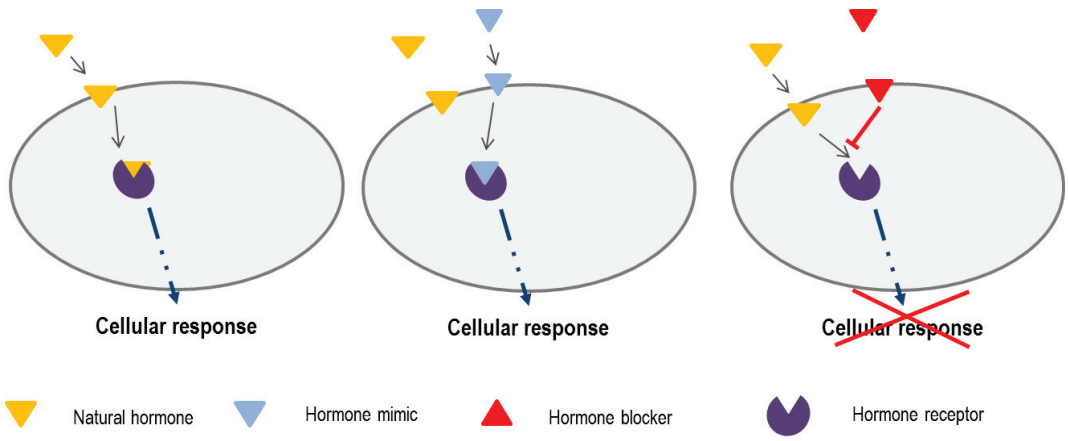


Figure I-4 -When absorbed in the body, an endocrine disruptor can decrease or increase normal hormone levels (left), mimic the body's natural hormones (middle), or block the hormone (right).

The presence of EDCs in aquatic environments is a top priority in environmental risk assessment. However, the taxonomic extent to which the toxicity of such compounds can be generalized is often problematic, even more when abusive data extrapolation from mammals to animal diversity is encouraged. Several robust, specific, and reliable biomarkers of EDC exposure are available for vertebrates, such as the Vtg induction in male fishes. However, despite their obvious ecological importance, the issue of ED has received much less attention in aquatic invertebrates. This is likely due to the scarce knowledge about basic endocrinology in species routinely used in ecotoxicology, for instance crustaceans. The potential endocrine effects of EDCs are often difficult to explain or inexistent because they seem to do not act through the expected modes of action. Numerous efforts were done to transpose fish molecular biomarkers to invertebrate species, the majority without success. As demonstrated recently on Vtg biomarker in *Gammarus* (Jubeaux et al., 2012b), and for the sexual steroid pathway in molluscs (Scott, 2012, 2013), ED knowledge available in fish should not be straightforwardly applied to invertebrates without performing a thorough functional "re-validation" in the targeted sentinel species. This task is difficult for EDC biomarkers because the diversifying biological evolution gave rise to divergent endocrine systems

between the main metazoan lineages, with specific own modes of hormonal regulation (Campbell et al., 2004). Indeed, current animal endocrine systems result from a complex diversification of molecular players such as hormone receptors or enzymes implicated in hormone synthesis (Markov et al., 2009). The relevance of cross species extrapolation of biomarkers is thus strongly dependent on the understanding of the mechanisms of action of pollutants and the conservation of biomarker sequences throughout animal evolution.

Specificity and ecological relevance

Despite being among the recommended operational procedures for assessment and management of chemical contamination in aquatic systems in the United States and in Europe, along with chemical analyses and bioindicators (Amiard-Triquet and Berthet, 2015; Chapman et al., 2013; Dagnino and Viarengo, 2014), biomarker-based approaches have been scarcely considered in the context of an integrated monitoring system for risk assessment protocols (Forbes et al., 2006; Hook et al., 2014a). This is mainly due to criticisms regarding their lack of specificity and ecological relevance. These criticisms revolve around the high natural variability of biomarkers that can hide stress-induced responses (Cairns, 1992), and the lack of ability from single biomarkers to accurately reflect an ecologically meaningful effect or response (Forbes et al., 2006).

Several confounding factors can exert an effect on biomarker responses. Among the most studied confounding factors are temperature, salinity, sex, parasitism, and weight/age/size/reproductive status of the organism. Many studies addressed the effects of these biotic and abiotic confounding factors on biomarker responses. For example, seasonal variabilities and species-specific differences were observed for MTs, AChE, and peroxisomal enzymes in two mollsucs (Bocchetti et al., 2008). Biotic factors like weight/age and size of the organism will also influence the responses due to different pollutant uptake rates for different organisms, duration of exposure, or the age of the organisms, but also in relation to methodologies used for sample preparation before measurement. One study highlighted the fact that organism body weight and female reproductive status strongly influenced AChE activity in crustacean gammarids (Xuereb et al., 2009b).

All of this comprises important complications when working in field studies, seen that organisms are exposed to different physico-chemical parameters and to complex mixtures of pollutants. Therefore, it is very difficult to exclude this "background noise" and link biomarker responses exclusively to the effects of a particular contaminant or class of contaminants. To surpass these difficulties, new strategies have been developed aiming at the control of confounding factors, such as caging calibrated organisms from a reference population (Cappello et al., 2013; Oikari, 2006; Xuereb et al., 2009b). The transplantation method allows for controlling the intrinsic biological parameters like age, sex, source, and reproductive stage of the organisms. However, one cannot control the physico-chemical characteristics of the study sites. To address this, it is imperative to establish reference values for each biomarker, taking into account their natural variability during the organism physiological processes, and with external stimuli like temperature, pH, etc. The control of biological and physico-chemical factors influencing the outcome of the assay in biomonitoring studies will yield reliable results for environmental hazard and risk assessment (Jha, 2008). In gammarids, these parameters were taken into account in order to establish reference and threshold values for several biomarkers and thus improve their reliability: feeding rate (Coulaud et al., 2011), digestive enzymes (Charron et al., 2014; Charron et al., 2013), DNA damage (Lacaze et al., 2011a), or AChE (Xuereb et al., 2009b). These studies applied strategies for determining biomarker variability in both laboratory (concerning intrinsic factors) and field deployments (long-term *in situ* spatio and temporal changes).

The second main limitation of the use of biomarkers as a tool in routine biomonitoring is the absence of works demonstrating its predictive power from an ecological quality perspective. In general, the more specific is the biomarker, the less ecologically relevant the response is going to be. Ideally, biomarkers need to be linked to adverse effects at higher levels of biological organization so they can be interpreted as early warning indicators of the presence of chemical stressors. However, studies demonstrating the relation between biochemical responses and long-term effects in organisms and populations are scarce. Nevertheless, some works managed to provide ways of doing it with some biomarkers. Miller et al. 2007 (Miller et al., 2007), for example, managed to link the inhibition of Vtg with fecundity success and population growth in the fathead minnow *Pimephales promelas*. In the crustacean amphipod

Gammarus fossarum, Xuereb et al. (Xuereb et al., 2009a) found quantitative relationships between AChE inhibition, and feeding/locomotor behavior of the organisms.

Recent studies used adverse outcome pathways (AOP) to perform the linkage between a molecular key event and its putative adverse outcome in the organism/population, through a structured representation of biological events across scales of biological organization. This framework was proposed (Ankley et al., 2010) as a tool for incorporating data and knowledge collected at many levels of biological organization and synthetize them in a useful structure for risk assessment (Figure I-5). AOPs consists in *a priori* risk evaluations that can be used to verify the linkage between sub-individual biomarker modulations and effects at higher levels of organization, thus reinforcing the usefulness of the biomarker approach. For example, AOPs explaining ecdysone receptor agonism leading to lethal molting in arthropods (Song et al., 2017), and AChE inhibition leading to acute mortality (Russom et al., 2014) were already developed. AOPs have been proposed as a tool for link directly molecular events with adverse outcomes into the framework of Environmental Risk Assessment (ERA) (Ankley et al., 2010). The utility of AOPs in biomarker-based ERA was discussed in a recent review using three case studies (Lee et al., 2015). The author concluded that AOPs, through combination with chemical modes of action and molecular initiating events (sub-individual biomarkers), could facilitate the development of qualitative and quantitative predictive toxicity models, thus aiding the choices of suitable bioassays for emerging pollutants.

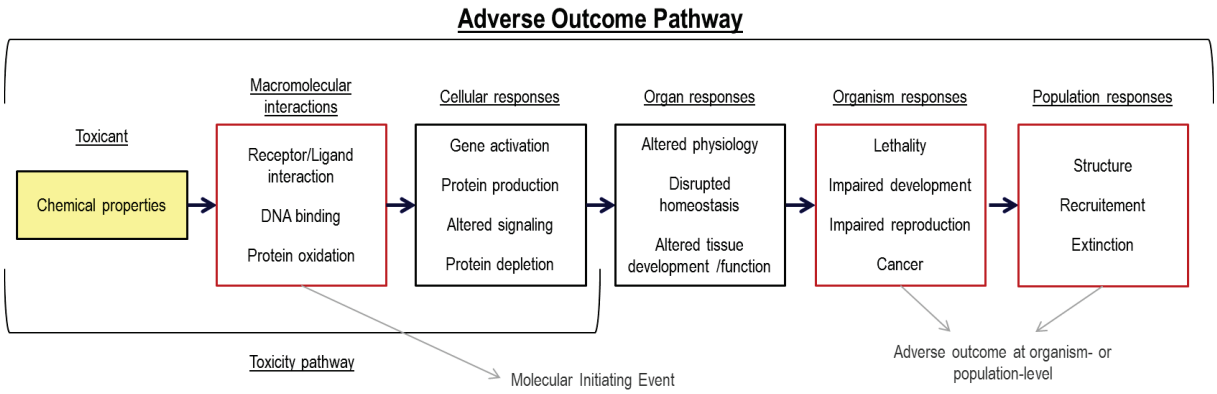


Figure I-5 – General outlines for the construction of an AOP. Adapted from (Ankley et al., 2010)

Technical restrictions

From a technical point of view, there is a lack of direct quantitation methods for biomarker measurements. Direct quantification of a biomarker ensures the repeatability of the methodology over time, and comparison between studies, providing a single unit of measurement that will no longer be arbitrary or protocol-dependent. As discussed in a recent review (Trapp et al., 2014a), most of the protein biomarkers used in Ecotoxicology are indirect assays for measuring enzymatic activities. Each assay is specific for one biomarker of interest, multiplying the laboratory work, cost, and time necessary for the analysis of a large number of samples. From a biomonitoring point of view, this hinders the routine application of biomarkers. Due to a lack of genomic data for invertebrate species, few species-specific methods are available for quantifying molecular biomarkers in these species.

The rapid growing "omics" techniques provide answers to these questions. The technologies available today allow for acquiring large-scale molecular information even in non-sequenced organisms, elucidating molecular modes of actions of contaminants, and/or developing sensitive methodologies for protein quantification. High-throughput transcriptional, proteomic, and metabolomic analysis allows for digging deeper into the molecular changes that occurs in sentinel species upon chemical contamination, analyzing the impacted pathways, and identifying new biomarkers that can potentially reflect effects at higher levels better than the currently used biomarkers. The opportunities provided by these approaches to develop direct quantification methodologies will be discussed further in Chapter 2.

1.2. Multibiomarker approaches as an indicator of the health of organisms

Seeking an exhaustive assessment, multibiomarker approaches try to associate a wide range of biological responses to reveal potential stress due to different classes of contaminants. The interpretation of these responses in an integrated manner has been proposed through several indices such as the Integrated Biomarker Response (IBR) (Beliaeff and Burgeot, 2002), the expert system Health Status Index (HIS) (Dagnino et al., 2007), the aquatic ecosystem health index (Yeom and Adams, 2007) and the Biomarker

Response Index (Hagger et al., 2008). These approaches were designed to simplify the interpretation of individual biomarkers and create an index capable of detecting and monitoring the biological effects of pollutants in organisms. Marigomez et al. 2013 (Marigómez et al., 2013) compared five different integrative indices, and showed that despite having different sensitivities, all of them provide coherent information for ecosystem health assessment.

The most commonly used multibiomarker index is the IBR, initially developed by Beliaeff and Burgeot (Beliaeff and Burgeot, 2002). This method uses computation of star plots to allow a visual integration of a set of early warning responses (biochemical biomarkers). Since most of the validated individual biomarkers concern marine environment, the use of biomarker index mainly focused in marine fish and mussel species. The biomarkers used for integration were the enzymatic activities from glutathione-S-transferase (GST), AChE, CAT, EROD, and genotoxicity from DNA adducts. When using integrative approaches, biomarkers are carefully selected *a priori* so that each one can provide different information associated with their specificity. This choice is done depending on the specific objectives of the study and the characteristics of the targeted sites. The IBR normally correlates well with the contamination data obtained from the sites, and provides a simple interpretation of biological effects of pollution in biomonitoring. Improvements to this approach were proposed by several authors, associated with new calculations (Devin et al., 2014), or addition of other types of biomarkers into the index (Fossi Tankoua et al., 2013; Richardson et al., 2011; Yeom and Adams, 2007).

Despite the several proposals and appeals to the use of these multibiomarker indices in ecotoxicology, such methods have been rarely applied since the first proposal of the IBR in 2002. This can be related to some limitations connected to the use of a single value to discriminate pollution-related changes from natural variations or from biological factors like inter-individual variability (Roland et al., 2016). However, several *in situ* applications demonstrated the usefulness and coherence of these approaches for either determining organism health status and/or discriminating between contaminated sites, especially using mussel species (Table I-1). Multibiomarker indices also present some constraints at the methodological level.

Table I-1 – *In situ* applications of multibiomarker indices for toxicity assessment.

Organism	Biomarkers	Exposure	Integration method	Main conclusion	Reference
Mytilus galloprovincialis	LMS, NL, LF, lysosomal content, DNA damage, and micronuclei frequency, CAT, MT, AChE, GST activities, Stress on stress response (SOS), Lysosome/cytoplasm volume ratio (L/C)	Caged mussels at a polluted site along the Ligurian coast (Italy)	DiSAV Expert System (DES)	The expert system provides a clear indication of the degree of stress syndrome induced by pollutants in mussels	Dagnino 2007
Mytilus spp.	Protein carbonyls, LF, neutral lipids, lysosomal stability and volume, ferric reducing antioxidant power [FRAP] and malonaldehyde [MDA]	Caged mussels in 5 sites from the Tamar River and Estuary	Mussel Expert System (MES)	MES determined higher levels of stress on all contaminated sites	Shaw 2011
Mytilus galloprovincialis	MT, ALAD, mixed function oxidase system (MFO), GST, AChE, SOD, CAT, GPx, LPO	Caged mussels in two sites in the South coast of Portugal	DiSAV Expert System (DES)	Transplanted mussels had significant alterations in some biomarkers that reflect the type of contaminants present in each site	Serafim 2011
Mytilus edulis	Reproductive condition, heart rate, feeding rate, lysosomal stability	Mussels collected for 1-year from the Exe Estuary	Integrative biomarker index (IBR)	By integrating biomarker responses into the BRI and creating an index of health, the natural variability of individual responses was limited	Hagger 2010
Crassostrea gigas	SOS, SOD, CAT, LPO, GST, LMS and DNA damage (COMET), AChE and MT	Organisms collected from seven sites along the north coast of the Shandong Peninsula	Integrative biomarker index (IBR)	The IBR index is a powerful tool to assess the sensitivity of organisms to contaminants, which may be used to distinguish between polluted and less polluted areas	Xie 2016
Zacco platypus	**molecular** [mRNA expression of CAT, SOD, GST, and MT], **biochemical** (enzyme activities of CAT, SOD, GST, and concentration of MT), and **physiological** [condition factor (CF) and liver somatic index (LSI)] levels	Pale chub (*Zacco platypus*) collected from various habitats in the Miho Stream	Multi-level Integrated Biomarker Response IBR	This study suggests that the multi-level IBR approach is very useful for quantifying *in situ* adverse effects of WWTP effluents	Kim 2015
Eisenia andrei	H_2O_2, Catalase (CAT) activity, Genotoxic damage (Comet assay), Glutathione S-Transferase (GST) activity, Malondialdehyde (MDA), Phenoloxidase (PO), Metallothioneins (MTs), Cd bioaccumulation (CdB)	Earthworms of *E. andrei* were breed in the lab	Rank-Based Index (RBI)	Rank-based biomarker index showed that both different contaminated soils had an effect on the earthworms	Panzarino 2016

Despite simplifying the analysis by integrating the several biomarkers in one single index, there is still a need of performing a specific analytical procedure for each biomarker, leading to extremely resource-consuming protocols (time, cost, organisms). Direct protein quantification methods are scarce, with rare exceptions such as the vitellogenin measurements through immunoassays in fishes (Meucci and Arukwe, 2005) and mass-spectrometry in crustacean gammarids (Simon et al., 2010). Enzyme-linked immunosorbent assays (ELISA) are the standard protocol for routine measurement of protein biomarkers, but still there are few assays available for use in ecotoxicology. Seen the large number of sentinel species possible in ecotoxicology (both vertebrate and invertebrate species), developing species-specific ELISA-antibodies is not feasible in environmental sciences. For routine biomonitoring, a multibiomarker approach allowing measurement a panel of biomarkers simultaneously in one biological sample would be advantageous and allow high-throughput analysis. Mass spectrometry (MS) based proteomics, for example, offers a valid alternative for high-throughput analysis of several proteins simultaneously. A recent study already combined the IBR and proteomic analysis for creating a new "Integrated Biomarker Proteomic" (IBP) index (Roland et al., 2016). In this work, 34 biomarker proteins were analyzed and integrated through IBR application to proteomic data. Authors showed the relevance of this new index in risk assessment by discriminating, in eels, the type of pollutant inducing a proteomic response. Proteomic approaches in ecotoxicology will be further discussed in section 2.

2. Omics tools and new strategies for molecular biomarker definition

In the last decades, with the continuous technological and computational advances, a large number of methodologies emerged for robust molecular analyses (Figure I-6). Next-generation RNAseq and microarrays allow analyzing entire transcriptomes, while nuclear magnetic resonance (NMR) and separation techniques coupled to mass spectrometry allow analyzing the entire metabolome, lipidome, and proteome (Simmons et al., 2015). Such technologies yield therefore large amounts of molecular data, which could benefit the development of molecular biomarkers.

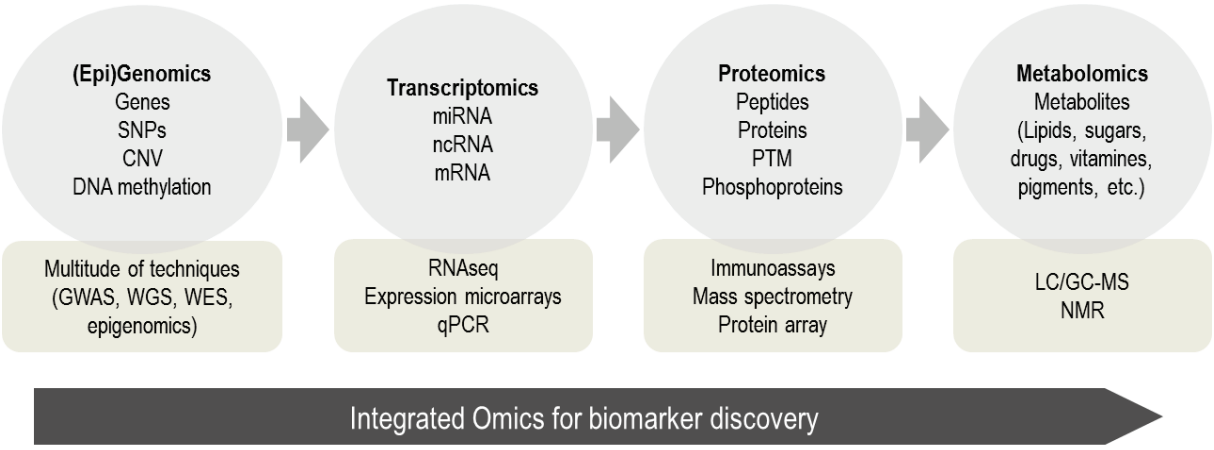


Figure I-6 – The high-throughput omics technologies offer a multitude of options for unraveling components, mechanisms, and regulation of biological processes. GWAS - genome-wide association study; WGS – whole genome sequencing; WES – whole exome sequencing; SNP - single-nucleotide polymorphism; CNV - Copy number variation; PTM – Post-translational modification; miRNA – micro RNA; ncRNA – non-coding RNA; mRNA – messenger RNA.

In this chapter, we will focus mainly in transcriptomics and proteomics. Transcriptomics, the global analysis of gene expression, can be used to constitute sensitive indicators of the interaction of organisms with their environment. It provides the physiological responses initiated at the molecular level through network analysis of the triggered pathways. However, it is known that transcriptional and proteomic responses do not often correlate (Greenbaum et al., 2003), mainly due to factors such as the rate of degradation of proteins, translational efficiency differences, and relativity to the status of the cell at the assessed time (Connon et al., 2012). In that sense, proteomics offers a robust approach in terms of assessing the functional biochemical pathways that respond to environmental stimuli. These high-throughput technologies brought

important innovations to the field of ecotoxicology, mainly through the transposition of strategies and methodologies developed in human medical research. The completion of several genomes in the beginning of the century also opens many opportunities into transcriptome and proteome characterization of different species. This allowed the establishment of new functional, evolutionary, and physiological correlations between taxa, individuals, organs, and cell types. In medical research, "omics" tools are now routinely employed in identification of the molecular pathways and discovery of biomarkers for many diseases (Abu-Asab et al., 2008; Erickson et al., 2012; Kussmann et al., 2006; Wheelock et al., 2013). However, the incorporation of these tools in ecotoxicology is challenging, due to: 1/ the lack of sequenced genomes for a big part of the sentinel species used (especially for invertebrates); and 2/ the need of encompassing populations, species communities, and ecosystems into the results. Nevertheless, "omics" approaches still offer the possibility to analyze simultaneously the ensemble of gene and protein expression changes due to environmental and/or anthropogenic stresses.

Recent reviews (Bahamonde et al., 2016; Martyniuk and Simmons, 2016) highlighted that "omics" approaches in Ecotoxicology should provide answers to 1/ identifying modes of actions of contaminants, 2/ identifying molecular initiating events to construct AOPs of environmental contaminants and understand how chemicals are associated with individual and population responses, 3/ investigate mechanisms of organismal adaptations, 4/ monitor adverse effects in organisms, and 5/ identifying candidate biomarkers of exposure (Figure I-7). The application of "omics" technologies in ecotoxicology lead to the generation of unprecedented volumes of data, increasing the molecular knowledge of model and non-model species used in ecotoxicological assays, as well as the nucleotide/protein information available in databases (Martyniuk and Simmons, 2016). However, it is still very challenging to find novel statistical analyses for interpreting the large amounts of data obtained. Similarly to biomarkers, "omics" indicators are also limited by the establishment of reference values, with seasonal, temporal and organismal variability, adaptive and/or toxic responses, and to link them with effects at higher levels of biological organization. Therefore, it is critical to identify biomarkers that eventually lead to adverse outcomes in population survival, reproduction, and susceptibility to disease. This is why the high-throughput, hypothesis-

free "omics" techniques are a must for biomarker discovery. Many works have already shown the usefulness of "omics" in generating valuable information on the effects of pollutants in organisms.

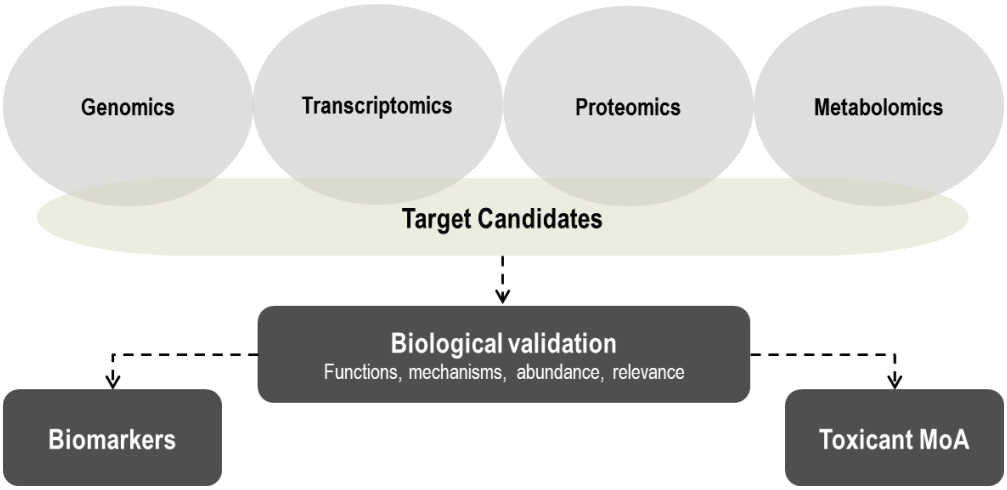


Figure I-7 – General contribution of omics techniques to Ecotoxicology.

2.1. Transcriptomics

Transcriptional expression levels are commonly used to characterize biological processes of interest through the understanding of the associated functional pathways. DNA microarrays, *i.e.* chips containing a collection of synthetized DNA sequences that hybridize with thousands of complementary targets for determining their expression, have been the gold standard for studying changes in mRNA expression. The implementation of this technique has been a major advance in the field of ecotoxicology, providing invaluable insights into toxicants modes of action. Poynton et al. (Poynton et al., 2007) for example, identified contaminant-specific expression profiles in *Daphnia magna* exposed to sublethal doses of cadmium, copper and zinc. Through this assessment, they proposed several biomarkers of exposure to metal contamination in that species, such as metallothionein, ferritin, and an iron-response element mRNAs. Another study about the effects of different model toxicants in the rainbow trout *Oncorhynchus mykiss*, observed that most modulated genes responded to only one of the toxicants and relatively few were co-expressed, providing therefore specific molecular modes of action for the contaminants studied (Hook et al., 2006). Despite the advantages

of this high-throughput analysis, these chips are restricted to model species that have considerable sequence information available such as the zebrafish *Danio rerio*, the frog *Xenopus larvis*, or the fathead minnow *Pimphales promelas* (Kim et al., 2015a). For non-model species, it is considerably harder to build a microarray containing pertinent targets, due to the lack of species-specific sequence information in databases. Fortunately, new sequence-based approaches (RNAseq) are emerging as preferred methods for transcriptome analysis. This approach allows identifying transcripts present in a sample without any *a priori* knowledge of their sequence, boosting the discovery of molecular information in sentinel species for which the molecular information was scarce. Thus, RNAseq constitutes a valuable tool for studying the molecular effects of pollutants in a wider range of species (Garcia-Reyero et al., 2008). RNAseq was already applied to several non-model species of interest in ecotoxicology such as the gastropod *Lymnea stagnalis* (Bouétard et al., 2012) challenged with the pesticide diquat, or the amphipod *Melita plumulosa* (Hook et al., 2014c) exposed to contaminated sediments. These studies showed the great promise of RNAseq technology to assess gene expression signatures of toxic contamination in non-model species, and to identify exposure biomarkers. Moreover, the establishment of a reference transcriptome for the concerned species can be subsequently used in other studies, namely for biomarker discovery, allowing for a better understanding of the physiology of the organisms. Microarrays and RNAseq are also often combined, with transcriptomic sequencing serving as a screening method for providing unique sequences, building gene libraries, and then constructing the microarrays using annotated genes of interest for the analysis specific functions of interest (Bahamonde et al., 2015; Pascoal et al., 2013).

While the majority of the transcriptomic studies in ecotoxicology are laboratory-based, an increasing number of field works already applied these techniques with some success. Bahamonde et al. (Bahamonde et al., 2015), for example, analyzed rainbow trouts exposed to municipal effluents to investigate the intersex occurrence often found in this species (Bahamonde et al., 2015). The authors managed to connect the mechanistic information obtained from the transcriptome (eg. higher expression of genes associated with oogenesis) with higher-level endpoints in order to better describe phenotypic profiles after exposure (eg. presence of eggs within the testis). Another

vertebrate study (Maes et al., 2013) with eels *Anguilla anguilla L.* collected from three Belgian rivers suggested an important impact of pollution in the health of the adult organisms. They documented a significant down-regulation of hepatic and gill gene transcription in populations from highly polluted sites, correlated with low energy reserves and condition.

Studies of these kind show the potential applicability of transcriptomics in field studies, and the importance of bridging mechanistic effects observed with the effects that the contaminants may have at higher levels of organization (Bahamonde et al., 2016; Connon et al., 2012). Regardless of being conducted in the field or in the laboratory, this mechanistic approach always relies on genome functional annotations available in databases (eg. Gene Ontology) to identify transcripts modulated by exposure. Due to a lack of common pathways between taxa, automated pathway analysis is sometimes limited and can generate numerous and sometimes disparate responses that need to be carefully analyzed. Therefore, major advances still need to be made, notably to 1/ identify and characterize the natural/adaptive and experimental variation of the large amounts of data obtained, 2/developing reproducible techniques, 3/ developing more pathways specific of relevant species that can be applied to other phylogenetically close species, and 4/ obtaining new molecular information from non-model but ecologically relevant species.

2.2. Proteomics

Proteins are crucial molecules for a vast array of functions within organisms, bridging the functional responses of gene expression and the physiology of the organism. As highlighted in Chapter 1, these molecules participate and play a major role in mediating the biochemical processes involving the action of contaminants in organisms. Moreover, proteins represent the actual functional molecules, since predicting the behavior of a gene product is impossible to determine due to the post-translational modifications (Figure I-8). According to (Anderson and Anderson, 1998), proteomics is the field that "uses quantitative protein-level measurements of gene expression to characterize biological processes (e.g., disease processes and drug effects) and decipher the mechanisms of gene expression control". The primary objective of

proteomics is to identify all the proteins in a cell, tissue, or organism at a given time. It can also be used to study protein sequences, protein-protein interactions, protein post-translational modifications, and subcellular location of proteins (Patterson and Aebersold, 2003). Proteomics collects various technical disciplines for protein analysis, normally based on immunoassays or mass spectrometry. Here the focus is on mass spectrometry based proteomics (MS-based proteomics). In the next section, a brief description of protein analysis by mass spectrometry is given before exemplifying proteomics applications in ecotoxicology.

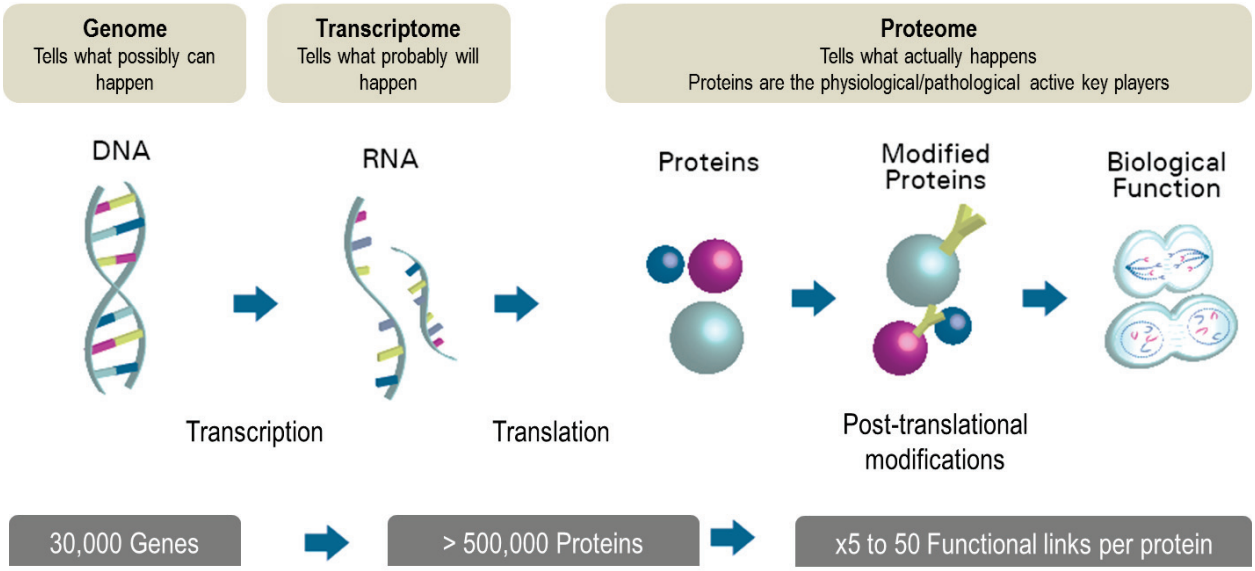


Figure I-8 – Complexity of the functional repertoire of gene expression products.

2.2.1. Protein identification by mass spectrometry

Mass spectrometry is a versatile and crucial tool in the field of proteomics. Its high sensitivity allows for the analysis of proteins at the fentomolar level with relatively small mass errors (less than 10 ppm) (Jensen, 2006; Vidova and Spacil, 2017). A mass spectrometer determines the mass/charge (m/z) ratios of gas-phase ions. Firstly, molecules are ionized in the ion source, originating gas-phase ions with different masses and charges (sometimes multiple charges). Secondly, gas ions are separated according to their m/z in the mass analyzer, and thirdly, the number of ions at each m/z is detected

by the ion detector (Han et al., 2008) (Figure I-9). These three components can be combined in several fashions in order to create different platforms with different capabilities.

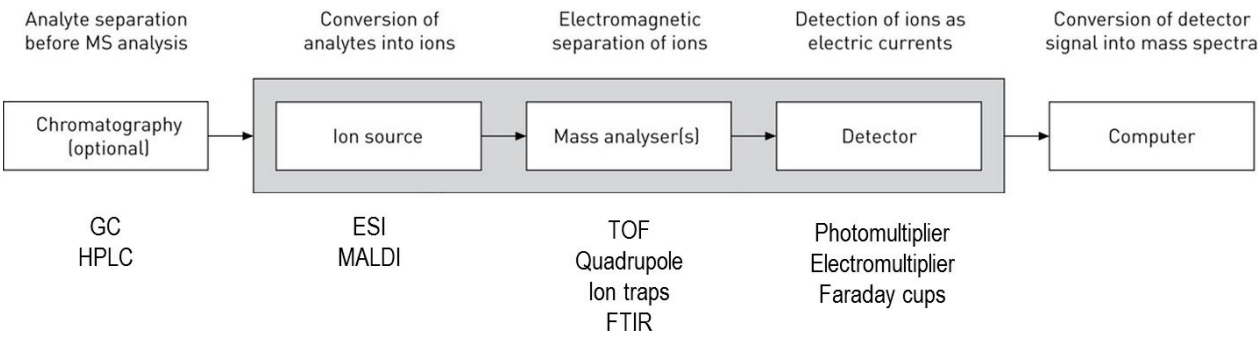


Figure I-9 – Three major components of a mass spectrometer: the ion source, mass analyzer, and detector. Mass spectrometry is often preceded by some form of chromatography (gas, liquid or thin layer chromatography) to separate analytes of interest before analysis. Adapted from (Wheelock et al., 2013). FTIR stands for "Fourier-transform infrared spectroscopy".

Mass spectrometry applications in proteomic analyses originated after the development of soft ionization techniques such as Electrospray ionization (ESI) (Loo et al., 1989), and Matrix-assisted laser desorption/ionization (MALDI) (Tanaka et al., 1988), that allowed a less destructive ionization of peptides and proteins. However, the mass analyzer is the key component for the analysis, since it allows ion separation based on their m/z. Iontraps, Orbitraps, and quadrupoles, separate ions based on their stability, the ICR (Ion Cyclotron Resonance) based on their cyclotronic resonance, and time-of-flight (TOF) analyzers based on their time of flight. Nowadays, hybrid analyzers are the most commonly used, as they allow tandem mass spectrometry analyses, i.e., the analysis of previously MS separated ions. The targeted ions (precursor ions) are isolated and fragmented, and the product ions are detected in a second stage of mass spectrometry (Khalsa-Moyers and McDonald, 2006).

The efficiency of a MS proteomic analysis is highly dependent on the separation technology employed for protein/peptide separation in complex mixtures. For a long time, the most common approach was protein separation by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). In 2D-PAGE, proteins are resolved according to their isoelectric point and molecular weight, being posteriorly stained and

digested in the gel. This allows a good separation of abundant proteins based on two distinct properties, and to visually identify the protein spots after gel staining. Protein spots of interest can then be isolated and analyzed by mass spectrometry in order to identify the protein through peptide mass fingerprinting. By using reference gels with annotated protein spots, it is possible to identify the protein without performing the MS analysis, just by superposing the gels. However, gel-based approaches have some limitations linked to their low-sensitivity, detection linearity, and gel reproducibility (Monteoliva and Albar, 2004). Liquid chromatography (LC) separation emerged as an excellent alternative to gel-based separation, because it allows separation at the peptide level (instead of the whole protein), and based on other properties depending on the column used, such as charge or hydrophobicity (America and Cordewener, 2008). This increases substantially the sensitivity and selectivity of the separation (Chen and Pramanik, 2009).

LC-MS is the method of choice for the analysis of complex protein samples, and opened the door for toxicologists addressing their biological problems in a more broad manner, through large-scale protein analysis, rather than being limited by the conventional focus on the role of single genes or proteins (Wetmore and Merrick, 2004). The "shotgun" proteomics approach, illustrated in Figure I-10, is the most commonly used LC-MS based technique, a global protein analysis where the aim is to identify the maximum number of proteins in a sample. In shotgun proteomics, proteins are enzymatically digested prior to the MS analysis with specific peptidases (e.g., trypsin, pepsin, proteinase K, endopeptidase Asp-N), followed by separation and sequencing of peptides by LC-MS/MS. Protein digestion can be performed *in gel* or in solution. Hundreds of proteins are identified through the comparison of the masses of MS/MS spectra obtained for the proteolytic peptides in the sample, with those predicted *in silico* from proteins present in public databases available in NCBI.

The number of proteins identified is dependent on the analytical method employed, and is proportional to the number of exact protein sequences available in databases for the concerned species. This constitutes one major drawback for proteomic analyses in non-model species, whose genomes are not sequenced and/or poorly annotated. The majority of species used in ecotoxicology do not have a sequenced

genome, so for these species the proteomics experiments are based on sequence similarities, using protein cross-species matching.

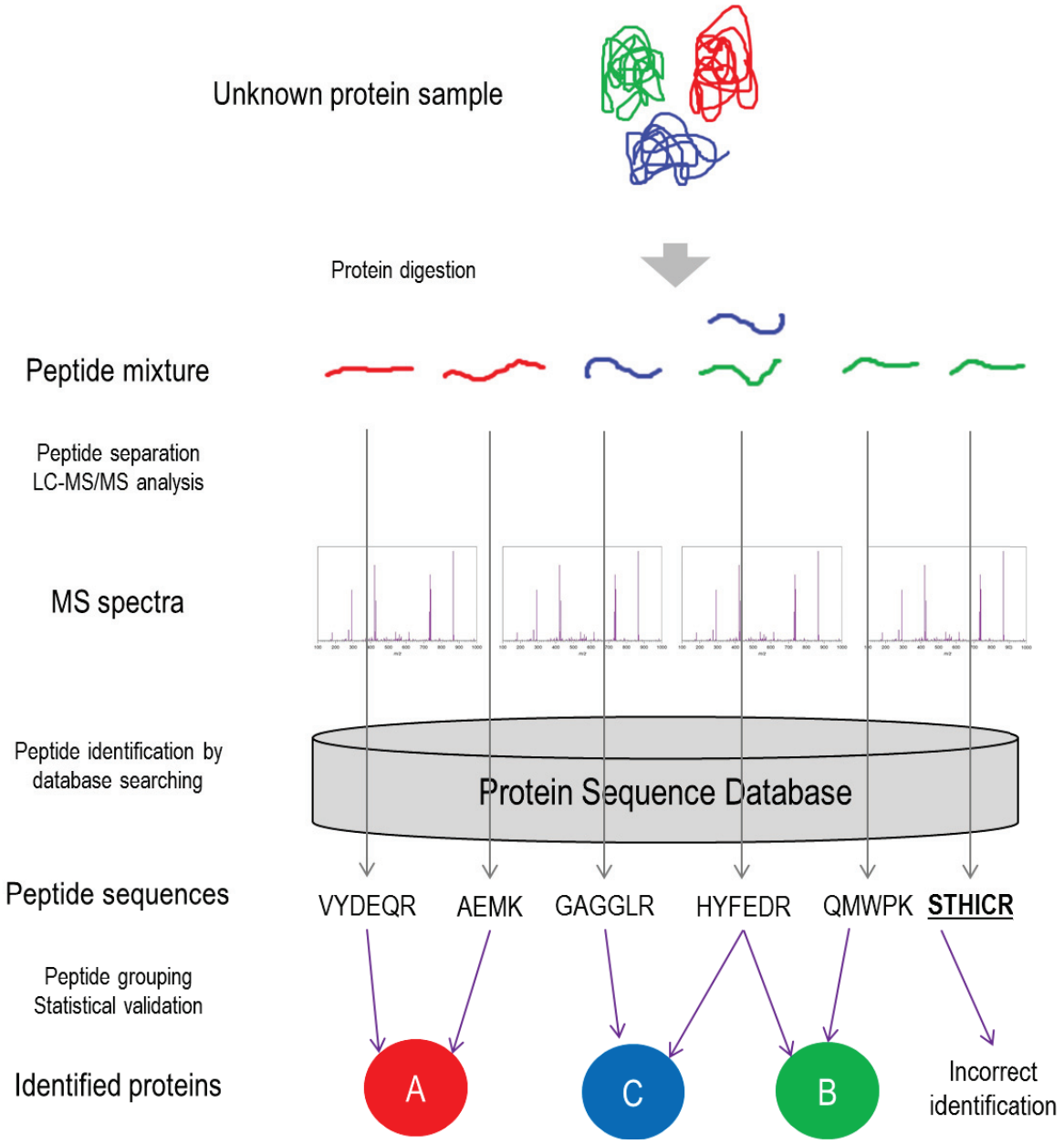


Figure I-10- General view of an experimental workflow used in shotgun proteomics analyses. Sample proteins are first digested into peptides. After chromatographic separation, peptides are ionized and fragmented to produce signature tandem mass spectrometry (MS/MS) spectra. Peptides are identified from MS/MS spectra using automated database search programs. Peptide assignments are then statistically validated and incorrect identifications filtered out (example: peptide STHICR). Sequences of the identified peptides are used to infer which proteins are present in the original sample. Some peptides are present in more than one protein (peptide HYFEDR), which can complicate the protein inference process. Adapted from (Nesvizhskii, 2007).

Many studies conducted in the last years have run into problems matching peptide masses because of low scores or false-positive matches (Armengaud et al., 2014). The most frequently identified proteins in invertebrates are involved in ATP (adenosine triphosphate) supply and cytoskeleton maintenance, i.e. the so-called "déjà-vu proteins" involved in housekeeping functions with low relevance from an ecotoxicological point of view (Petrak et al., 2008). Moreover, the function is predicted based on sequence conservation and functional information from various taxonomic groups, mainly vertebrates. Due to high diversification and long evolutionary times, homologue sequences in distant organisms will often present different functions. The solution for this problem would be sequencing the genome of the species of interest for creating a complete protein sequence database. However, genome annotation is a fastidious process that takes a long time to conclude. For crustacean species, for example, only four genomes were sequenced until now (*Daphnia pulex, Neocaridina denticulata, Parhyale hawaiensis, Hyallela Azteca*). With the exception of *Daphnia*, the other genomes only became available in the last three years, so there is still a big road ahead concerning gene annotation and functional studies before the protein sequences and annotations are available for interpreting proteomic data.

2.2.2. Proteogenomics for protein discovery in non-model species

Proteogenomics is a rapid alternative to genome sequencing, and it consists in sequencing only the mature RNAs (coding part of the genome) for having information of species-specific protein sequences. Nowadays, RNAseq technology allows a cost-effective deep sequencing of the transcriptome for quickly identifying protein-encoding genes. Besides allowing comparative transcriptomics experiments (highlighted in section 2.1.1.), a protein database can also be generated by six-reading or three-reading frame translation of the RNAseq nucleic acid sequences. These breakthrough advances allowed multiomics alliances between proteomics and genomics, termed "**proteogenomics**". First directed at better annotating genomes using proteomic information, proteogenomics was applied later for discovering proteins in non-model organisms through the six-reading frame translation of transcripts (Armengaud et al., 2014) (Figure I-11). Through sequencing the entire transcriptome, a species-specific

transcriptome database can be established and used for interpreting MS spectra for protein identification.

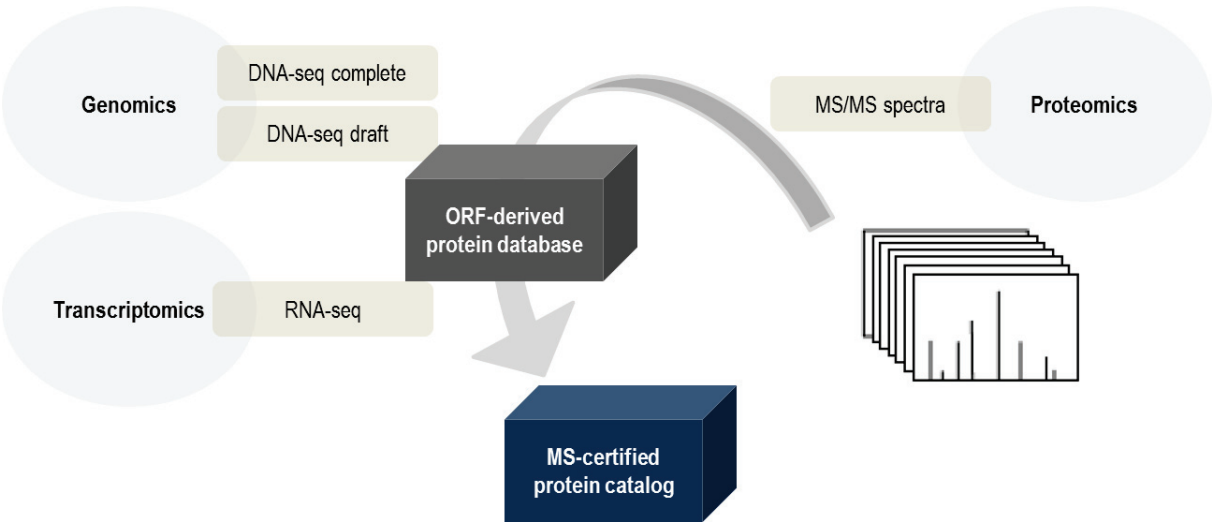


Figure I-11 – Proteogenomics for non-model organisms. Typically, a six reading frame translation of the nucleic acid information (either complete or draft DNA-seq or RNA-seq data) generates a first protein database. The MS/MS spectra recorded by tandem mass spectrometry are assigned to peptide sequences and the resulting mass spectrometry-certified proteins can be further structurally and functionally characterized, and thus can improve the protein database. Adapted from (Armengaud et al., 2014).

This *de novo* approach was applied, for example, to *Gammarus fossarum*, leading to the identification of 1873 reproduction-related proteins, to the nematode *Heligosomoides polygyrus* (209 identified proteins) (Moreno et al., 2011), or even the domesticated tomato *Solanum lycopersicum* (more than 1200 identified proteins) (Lopez-Casado et al., 2012). More recently, Grossmann et al. (Grossmann et al., 2017) also applied a transcriptome-assisted proteomics analysis that led to the identification of 1397 proteins in the apogamous fern *Dryopteirs affinis*, four times more than the proteins identified from a public viridiplantae available database. RNAseq technologies are an inexpensive and effective alternative to genome sequencing for performing proteomic studies in non-model eukaryotes. However, despite increasing greatly the number of protein identifications, proteogenomics is still limited by possible errors arising from the *de novo* assembly of the transcriptome, and by the inference of protein functions. Nevertheless, its high-throughput capacities allows having invaluable molecular information that can be used as a first screening method for identifying key

players in the organisms physiology and to develop biomarkers. Even without functional annotations, associating physiology knowledge in the sentinel species used, one can infer experimentally the putative functions of proteins and target those proteins as being relevant for key functions. This way, invertebrate ecotoxicologists can direct their researches to species-specific proteins, and developing customized biomarkers instead of directly transposing those from vertebrate species (Trapp et al., 2014a).

2.2.3. Mass-spectrometry-based protein quantitation

In large-scale proteome analysis, and especially when performing comparative proteomics between different experimental conditions, one must determine the abundance of the proteins present in the sample. MS protein quantification is a valid alternative to the classic ELISA immunoassays, with the advantage that it is not antibody-dependent. From a biomarker development point of view, the development of a MS quantification method is easier, quicker, and cheaper than ELISA. Moreover, the high-throughput capabilities of a mass spectrometer allow analyzing much more proteins simultaneously. Quantitative proteomics by MS can be divided in two major approaches: labelled and label-free approaches. Stable isotope labelling is the most used approach and consists in spiking the sample with a labelled version of the peptide or protein of interest, i.e. a "mass-tagged" variant of the peptide/protein. Labelled peptides/proteins are obtained through the introduction of one or more non-radioactive isotopes (^{2}H, ^{13}C, ^{15}N, ^{18}O), the mass tags, that can be inserted chemically after expression, metabolically during *in vivo* expression, by synthesizing the isotopically-labeled references or enzymatically by inserting the labels during protein digestion (Table I-2) (DeSouza and Siu, 2013).

Label-based approaches

Isobaric tags for relative and absolute quantitation (iTRAQ), Isotope-coded affinity tags (ICAT) and Stable isotope labeling by amino acids in cell culture (SILAC) techniques are among the most used labeling methods applied to MS protein quantification. **iTRAQ** (Aggarwal et al., 2006) has the advantage of allowing the

simultaneous analysis of several samples, by using isobaric tags that label the N-terminus and side chain amines of digested peptides in a maximum of eight different samples. The labeled peptide samples are then pooled and analyzed by LC-MS/MS. The iTRAQ reporter ions with different masses will be released from the peptides during the fragmentation, and their peak intensities used for calculating the abundance of their corresponding peptides and proteins.

Table I-2 – Labelling strategies commonly used in MS-based protein quantification. Adapted from (DeSouza and Siu, 2013)

Category	Method and principle	Labelling site	Quantification basis	Nature of samples
Chemical	ICAT	Cysteine residues	Relative area of MS peak	cell lysate or tissue lysate (post-extraction)
	iTRAQ	primary amines at peptide N-terminii and Lys-side chains	Relative area of signature ion peaks in each MS/MS spectrum	cell lysate or tissue lysate (post-extraction)
	ICPL	primary amines at protein N-terminii and Lys-side chains	Relative area of MS peak	cell lysate or tissue lysate (post-extraction)
	dimethylation	primary amines at peptide N-terminii and Lys-side chains	Relative area of MS peak	cell lysate or tissue lysate (post-extraction)
	mTRAQ	primary amines at peptide N-terminii and Lys-side chains	Relative area of MRM traces	cell lysate or tissue lysate (post-extraction)
Metabolic	^{15}N	All nitrogen atoms in a peptide are theoretically labeled.	Area of MS peaks	plants and actively growing autotrophs
	SILAC	depending on the labeled amino acids used, peptides can be labeled at Leu, Lys and Arg	Area of MS peaks	actively dividing cells auxotrophic for the labeled amino acid
Synthetic labeling	AQUA	Labeled residues used	Relative area of MRM traces	Can be spiked into any sample
	QconCAT	Depends on the choice of labeled residues, typically Arg or Lys	Relative area of MRM traces	Can be spiked into any sample
Proteolytic labeling	^{18}O labeling	Carboxyl termini of peptides	Area of MS peaks	Sample type is unrestricted

ICAT labeling (Gygi et al., 1999) comprises the use of a chemical labeling reagent with three elements: a protein reactive group (iodoacetamide), an isotope-coded linker, and an affinity tag (biotin). The protein-reactive group reacts with the sulfhydryl group of cysteine residues in the protein forming a covalent bond to the residue (DeSouza and Siu, 2013). Labeled peptides are isolated by avidine-affinity chromatography, and then the purified sample is submitted to a 2D-LC-MS/MS analysis. Quantification is done by

integrating the peak areas of extracted ion chromatograms (XIC) for each peptide (at the MS level). Since only the thiol groups are labeled, ICAT has the advantage of simplifying the sample, facilitating the analysis, and allowing for the quantification of low-abundant proteins. However, only Cys-containing proteins can be analyzed.

SILAC metabolic labeling is done *in vivo* (Ong et al., 2002), by introducing isotopically labeled essential aminoacids in the culture media for cell growth. Once labeled, the proteomes for the different conditions are mixed, digested, and analyzed by LC-MS/MS. As in ICAT, quantification is done on the MS scan, and MS/MS is used for obtaining sequence information. These techniques, although very useful and widely applied in the last years, are still limited by the cost of the isotopes, expert data analysis, and technical complications such as high-dependency of labeling efficiency, and the limited capacity for quantifying several samples.

Label-free approaches

Due to the rapid development in MS instrumentation and the shift from gel-based approaches to LC-MS/MS shotgun methods, label free approaches started gaining more relevance in quantitative proteomics. These approaches are a simpler, cheaper alternative to labelling methods, being automated and allowing for large scale monitoring of the proteome (Wang et al., 2008). However, quantification precision and accuracy are much lower than isotopic labelling methodologies. Label-free approaches are mainly based in the high correlation between protein abundance and chromatographic peak areas (Bondarenko et al., 2002) or number of MS/MS spectra (Liu et al., 2004) produced by peptide fragmentation (Figure I-12). Despite its simplicity, label-free shotgun proteomics produces large amounts of data that requires rigorous statistical assessment. These large scale quantification methods of entire proteomes is now the most used technique for biomarker discovery, since it allows the identification of unique expression patterns of a protein or a group of proteins in response to a particular stimulus and the approach is rather robust.

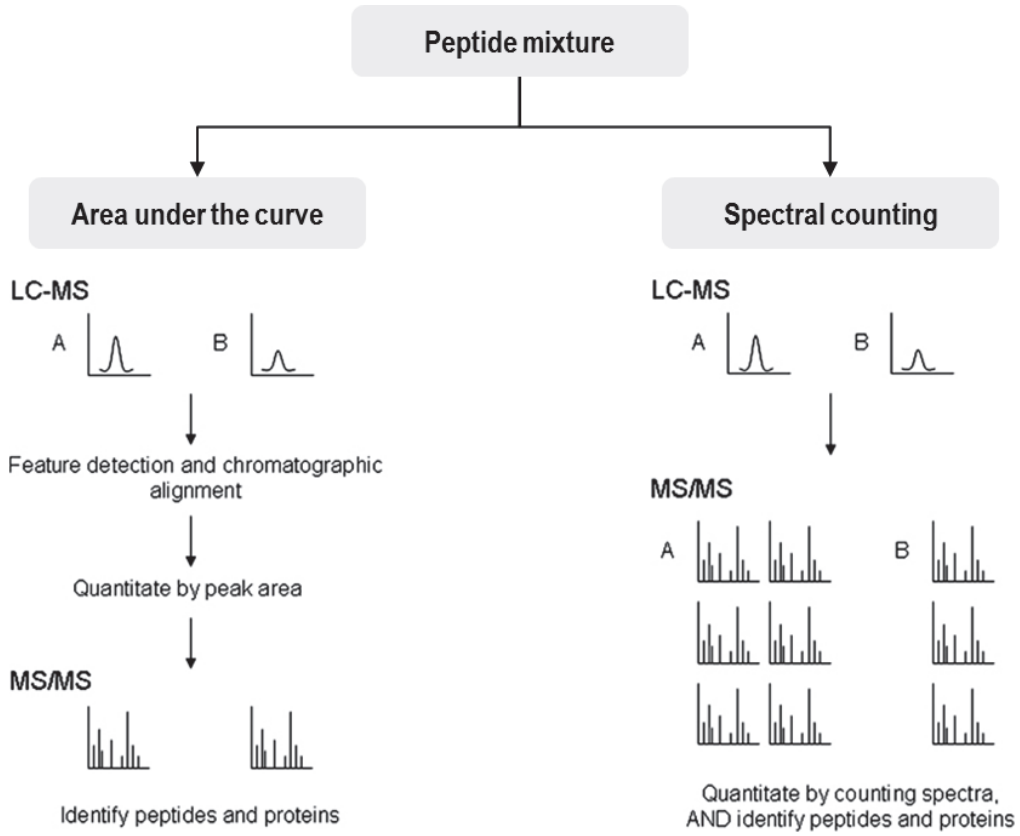


Figure I-12 – The two main approaches used in label-free proteomics. The "area under the curve" strategy is performed at the MS level and peptides that have differential expressions are posteriorly identified at the MS/MS level. Peptide quantitation and identification by "spectral counting" are performed simultaneously at the MS/MS level. (Neilson et al., 2011)

Targeted approaches for relative and absolute quantification

Targeted proteomics is commonly used to quantify peptides/proteins from complex mixtures. In these approaches, instead of analyzing the whole proteome, only a number or target peptides and corresponding transition ions are selected manually for quantification. **Selected Reaction Monitoring (SRM)** (sometimes referred to as Multiple Reaction Monitoring, MRM) has emerged recently as the technique of choice for precise quantification of targeted proteins by mass spectrometry (Picotti and Aebersold, 2012). SRM has been used for a long time in small molecule analysis in complex sample backgrounds, being able to quantify proteins with fentomolar concentrations (Vidova and Spacil, 2017) across multiple samples in a consistent and reproducible manner (Neilson et al., 2011). One study already reported SRM protein quantifications in the attomolar scale (Onisko et al., 2007). In contrast with shotgun proteomics, aimed at analyzing the largest possible number of proteins in a sample, SRM methods use the

mass spectrometer as a filter in order to target the proteins of interest. Combining chromatographic retention values and m/z values for specific ions, the unequivocal identification/quantification of targeted proteins is possible even in highly complex mixtures. One important particularity is that SRM measures peptides originated from enzymatic digestion of a proteome. These peptides are then used as surrogates for the concentration of the corresponding proteins. SRM can be used either for relative-quantification, through comparison between two samples, or absolute-quantification, by spiking the sample with heavy labelled synthetic peptides of known concentration (QconQAT and AQUA peptides, Table I-2). In absolute quantification, heavy-labelled peptides will have the same retention times as the endogenous peptides, and quantification is easily achieved by comparing the chromatographic peak areas of both peptides. SRM has therefore reduced technical variability, higher sensitivity, and lower detection limits, greatly enhancing the quantification accuracy and dynamic range when compared to shotgun label-free methods (Simmons et al., 2015).

SRM generally involves the use of a LC-MS system with ESI as the ion source, and exploits the unique capability of the triple quadrupole analyzer (QQQ) to act as mass filters, as illustrated in Figure I-13.

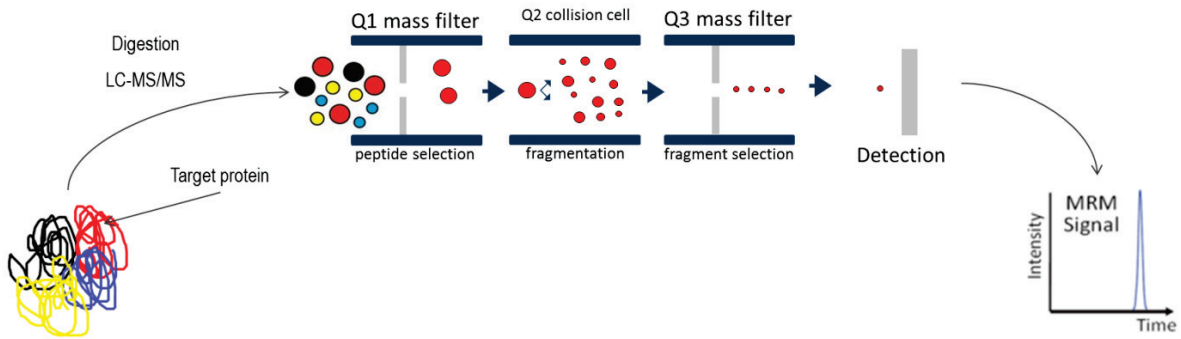


Figure I-13 – Protein analysis using the SRM mode with a triple quadrupole. The target protein (in red) is quantified via unique peptides filtered in the first quadrupole.

When using a QQQ, one can selectively monitor a specific analyte molecular ion and one or several fragment ions generated from the fragmentation of the analyte (Picotti and Aebersold, 2012). In Q1, a precursor ion from a pre-selected peptide is filtered, fragmented by collision-activated dissociation in the Q2, and the resulting

product ions uniquely derived from the targeted peptide are measured in Q3. Each m/z signal for a precursor-product ion pair is termed "transition". Several transitions can be sequentially and repeatedly measured, originating a chromatographic peak for each transition, thus allowing the simultaneous quantification of multiple peptides in the same sample. The fact that each transition is specific to a particular peptide offers a high multiplexing capability to proteomic SRM quantification. The quantification of the targeted peptides, and by inference, the protein, is achieved via integration of the chromatographic peaks for each transition. SRM robustness is highly dependent on the selection and design of suitable transitions for the target proteins. The higher is the number of proteotypic peptides and transitions for a protein, the higher is the specificity of the analysis. A general workflow for an SRM experiment is given in Figure I-14.

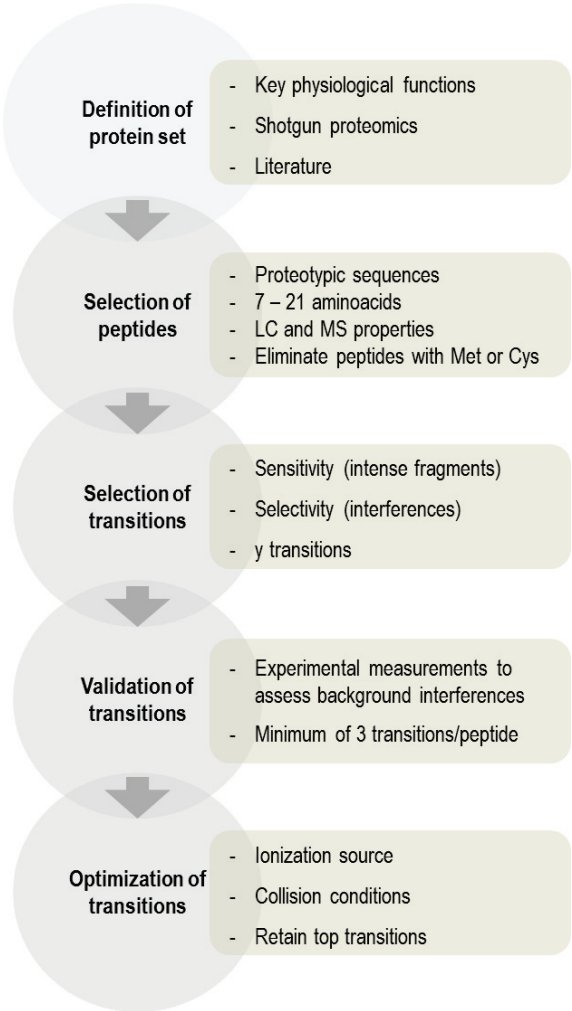


Figure I-14 – Workflow of peptide and transition validation for a SRM-based proteomic experiment. Adapted from (Gallien et al., 2011)

Finally, **Sequential Window Acquisition of all THeoretical mass spectra (SWATH)** is a recent developed methodology that overcomes some limitations of the SRM approach. By operating in the "data-independent acquisition" (DIA) mode using sequential isolation window acquisition principle, this strategy generates fragment ion spectra of all precursor ions within a defined precursor retention time and m/z (Gillet et al., 2012). The use of DIA mode increases greatly the throughput of the analysis, combining the sensitivity of a SRM analysis with the screening capabilities of global proteomics. Transition ions are then matched to a previously defined spectral library, generating protein identification and quantification. Although having a high-throughput potential, since SWATH methods rely on spectral libraries obtained by shotgun proteomics, it will not allow obtaining deeper proteomes than that from shotgun proteomics. However, the combination of this high-throughput targeted methodology holds great promise for label-free quantification (Huang et al., 2015).

2.3. Proteomics applications in Aquatic Ecotoxicology

Proteomics-based ecotoxicology is mainly directed to the identification of exposure fingerprints that can be used as a tool for determining modes of action, and for biomarker discovery (Benninghoff, 2007; Wetmore and Merrick, 2004). High-throughput analysis of the whole proteome of an organism was first applied in ecotoxicology in the year 2000 (Shepard et al., 2000), in a study that analyzed the proteomic response of the mussel *Mytillus edulis* exposed to polychlorinated biphenyls (PCBs), Cu, and salinity stress, in order to identify chemical-specific protein expression signatures. At the time, the majority of works employed the conventional approach based on 2D-PAGE to compare the profiles of control and environmentally challenged organisms. These works showed some promise by showing that controlled exposure to different contaminants yielded distinct spotting patterns on the gels. However, limitations involving gel to gel variations, overlapping spots, low abundant proteins, and the laborious sample handling hindered more vast applications of gel-based proteomics. Gradually, due to the major technological advancements made in instrumentation and software, the classic 2D-PAGE approach was replaced by the use of LC-MS, which allows minor sampling handling, and has a greater dynamic range and efficiency (Martyniuk

and Simmons, 2016). Along with the growth of available genome sequences, LC-MS based approaches facilitated protein separation and identification, and increased substantially the information obtained from the analyses. The appearance of the proteogenomics approach also allowed a big boost for proteomic analysis in non-model species. In its first application for an aquatic invertebrate sentinel species (*Gammarus fossarum*), it led to the discovery of 1873 proteins from the reproductive tissues. The functional annotation of these proteins was performed both via sequence similarity (bioinformatics annotations), and through the analysis of proteome dynamics through several physiological conditions. These studies allowed for the identification of male and female specific proteins, and proteins implicated in key reproductive processes such as oogenesis, embryogenesis, and spermatogenesis (Trapp et al., 2016; Trapp et al., 2015; Trapp et al., 2014b).

Similarly to transcriptomics, most of the proteomic studies in ecotoxicology were performed in fishes (Biales et al., 2011; Malécot et al., 2009; Zhu et al., 2006) and mussels (Hu et al., 2014; Liu et al., 2012; Nzoughet et al., 2009). The number of studies in other invertebrate species (especially freshwater) is still much lower. Some important studies that used non-model invertebrate species in order to study the proteomic responses of stressed organisms appeared in the last years (Jubeaux et al., 2012a; Silvestre et al., 2006; Son et al., 2011; Trapp et al., 2015). Table I-3 summarizes several examples of both laboratory and field-based proteomic applications in ecotoxicological studies using vertebrate and invertebrate species. The majority of them were performed under laboratory-controlled conditions. This is mainly due to the higher difficulty in controlling the conditions of exposure and to the fact that the comparisons between biological responses are not as straightforward as in laboratory-controlled experiments. Exposures in which biological (genetic background, age, sex, life stage, reproductive status, parasites) and physico-chemical factors (temperature, pH, conductivity) are not controlled, introduce a lot of variation in proteomic biomarker responses in the field, especially when comparing different populations (Bahamonde et al., 2016; Hidalgo-Galiana et al., 2014).

Table I-3 – Examples of proteomic studies in aquatic ecotoxicology.

Exposure	Organism	Objective	Reference
Laboratory	Chinese mitten (*Eriocheir sinensis)*	Characterize the proteomic expression profiles in anterior gills after cadmium exposure in a non-model species	(Silvestre et al., 2006)
Laboratory	Zebrafish (*Danio rerio*)	Investigate the alterations of membrane protein profiles in zebrafish liver cells exposed to methyl parathion	(Huang and Huang, 2012)
Laboratory	Fathead minnow (*Pimephales promelas*)	Investigate the effects of 17α-ethinylestradiol in the cephalon proteome	(Martyniuk et al., 2010)
Laboratory	Atlantic cod (*Gadus morhua*)	Define an estrogen-responsive protein expression signature in plasma of Atlantic cod	(Mæland Nilsen et al., 2011)
Laboratory	Zebrafish (*Danio rerio*)	Investigate the occurrence of triclosan acclimation and the biological mechanisms underlying the stress response triggered in early-life stage of zebrafish	(Falisse et al., 2017)
Laboratory	Fathead minnow (*Pimephales promelas*)	Identify and quantify differentially expressed hepatic proteins from female fathead minnows exposed to fadrozole	(Ralston-Hooper et al., 2013)
Laboratory and field study	European eel (*Anguilla Anguilla*)	Evaluating the toxicological effects of PFOS in European eel peripheral blood mononuclear cells at the protein expression level	(Roland et al., 2014)
Field	Flatfish (*Limanda limanda*)	Produce proteomic profiles of plasma from 213 dab collected in a UK National Marine Monitoring Programme	(Ward et al., 2006)
Field	White Sucker (*Catostomus commersonii*)	Generate protein profiles of mature male and female White Sucker that were collected from various sites along the main stem of the Athabasca River	(Simmons and Sherry, 2016)
Laboratory	Honeybees (Apis mellifera)	Assess the effects of fipronil in the brain of Africanized *Apis mellifera* workers	(Roat et al., 2014)
Laboratory	Abalone (*Haliotis diversicolor supertexta*)	Understand the molecular mechanisms of EDCs-toxicity to aquatic organisms	(Zhou et al., 2010)
Laboratory	Collembolan (*Paronychiurus kimi)*	Identify promising new biomarkers of cadmium by identifying differentially expressed proteins in *Paronychiurus kimi* after exposure to cadmium	(Son et al., 2011)
Field Laboratory	Beetle (*Agabus ramblae*)	Compare the response to temperature extremes of two geographically distant populations of a diving using 2-D DIGE (DIfferential Gel Electrophoresis)	(Hidalgo-Galiana et al., 2014)
Laboratory	Amphipod (*Gammarus pulex)*	Investigate protein expression differences in caeca of *Gammarus pulex* exposed to pentabromodiphenyl ether BDE-47	(Gismondi et al., 2015)
Laboratory	Amphipod (*Gammarus fossarum)*	Assess the integrated response and possible acclimation mechanisms in *Gammarus fossarum* following chronic exposures to Cd, Cu or Pb	(Gismondi et al., 2017)
Laboratory	Amphipod (*Gammarus fossarum)*	Semiquantitative proteomics based on spectral counting procedure was carried out on male gonads to observe the biological impact of exposure to EDCs	(Trapp et al., 2015)
Laboratory	Decapod (*Macrobrachium rosenbergii)*	Impact of chlordecone on the proteome of *Macrobrachium rosenbergii*, by gel-free proteomic analysis	(Lafontaine et al., 2017)
Laboratory	European eel (*Anguilla anguilla*)	Identify specific protein expression signatures on *Anguilla anguilla* peripheral blood mononuclear cells after exposure to p,p′-DDT and Cd	(Roland et al., 2016)
Laboratory	White shrimp, *(Litopenaeus vannamei)*	Understand molecular responses of crustacean hepatopancreas to cold stress	(Fan et al., 2016)
Laboratory	Marine mussels *(Mytilus spp.)*	Evaluate sublethal effects of diclofenac and gemfibrozil on the protein profiles of marine mussels	(Schmidt et al., 2014)
Field	Sydney Rock oysters (*Saccostrea glomerata*)	Assess the impacts of metal contamination in the field on Sydney Rock oysters protein profiles	(Thompson et al., 2012)
Field	Amphipod (*Gammarus fossarum)*	Field application of a vitellogenin (Vg) mass spectrometry-based assay to assess the exposure of the amphipod *Gammarus fossarum* to endocrine-disrupting chemicals in freshwater hydrosystems	(Jubeaux et al., 2012b)
Field	Blue mussel (Mytilus sp.)	Analyze the proteome response of the blue mussel in relation to the concentration of different trace metal contaminants	(Helmholz et al., 2015)

Nevertheless, the data obtained from the field proteomic studies performed until now generated valuable data on the biological responses from a wide range of organisms from various polluted sites, and on the advantages/limitations of the techniques, approaches, instrumentation, and biological material used for analysis. All of this data should be used for improving future applications of proteomic studies in the field.

2.4. Proposal of a proteomics biomarker pipeline for environmental applications

For many years, the use of molecular biomarkers in ecotoxicology holds the promise of revolutionizing diagnosis and prognosis of pollution impacts in aquatic ecosystems. These expectations continued to grow along with the rapid developments in the high-throughput omics technologies, which offer great perspectives for biomarker research. This led to a steady increasing number of studies that search for biomarkers using these techniques. As highlighted by (Harlan and Zhang, 2014), in the field of clinical proteomics, thousands of publications arose with propositions of new potential disease biomarkers obtained from discovery experiments. However, there are very few biomarker assays available for clinical use, mainly due to the lack of high-throughput quantitative assays that allows for biomarker verification and validation. The same applies to the field of ecotoxicology. Proteomics studies are producing large amounts of data and new candidate biomarkers, but it is still challenging to refine the true positives from these long candidate lists. Refined candidates also need to be validated before their use in biomonitoring. Therefore, the systematic development of protein biomarkers should follow a rational design that defines the several steps needed for the discovery and validation phases. Obviously, this biomarker development pipeline is strongly dependent on the type of biomarkers to search, as well as on the species into consideration.

Based on the general biomarker pipeline used in clinical proteomics (Drabovich et al., 2015), Trapp et al. proposed a multi-omics workflow for developing and validating biomarkers in an environmental scope (Trapp et al., 2014a), schematized in Figure I-15.

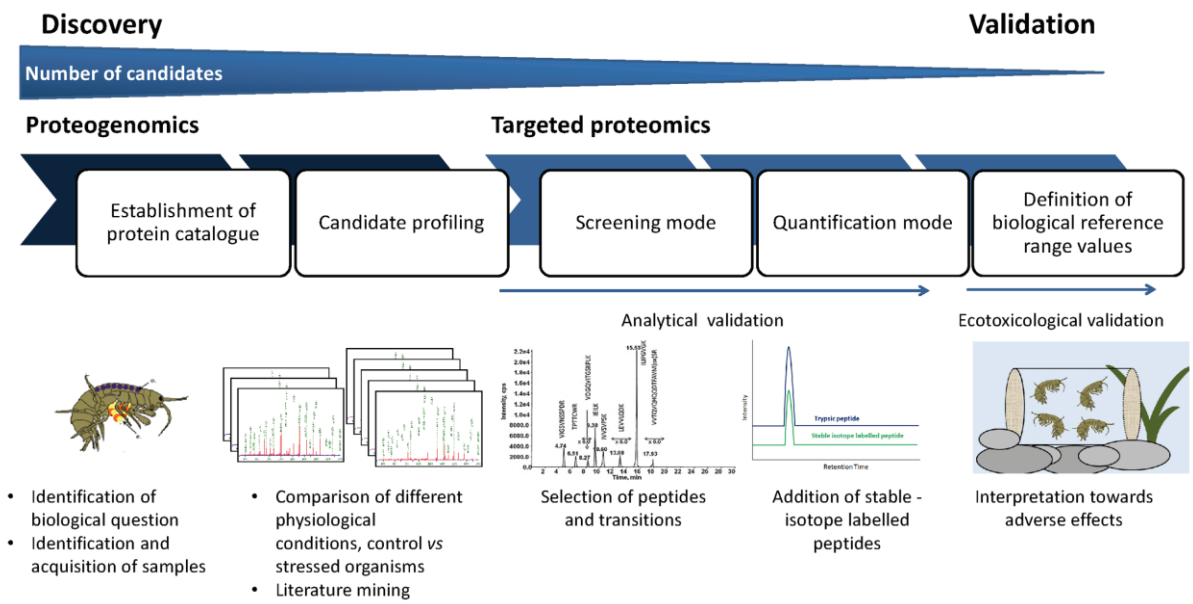


Figure I-15 – Workflow of the combination of proteogenomics and targeted proteomics for biomarker development and validation in an environmental scope. (Trapp et al., 2014a)

This conceptual framework was developed for non-model sentinel species, with the strategy comprising the use of the proteogenomics approach in the discovery phase. In this phase, the purpose is to assess a maximum of candidates and to exclude non-pertinent proteins as early as possible. Briefly, this will comprise identifying the proteins in relevant biological samples, and studying their modulations in different key physiological conditions and/or stress conditions through label-free shotgun proteomics. This will allow profiling the candidates according to their implication in the studied physiological processes and stress responses. Alternatively, candidate biomarkers may be identified through studies employing other omics approaches related to genome sequencing, RNAseq, or DNA/protein arrays (Figure I-6).

Despite being considered a reliable tool for biomarker discovery and relative quantitation of peptides and proteins, the dimensionality of the results from label-free shotgun proteomics lead to multiple bioinformatics analyses that yields semi-quantitative information. Therefore, it is imperative to develop downstream quantitative proteomic assays for biomarker validation. The assay must allow targeting the most promising candidates, confirm their ability to enable differentiation from control and stressed samples, and exclude the possible false candidates. The targeted proteomics approach SRM allows a high-sensitivity and high-throughput detection and

quantification of the biomarker candidates. During the screening mode, relevant peptides should be selected according to analytical criteria (specificity, sensitivity, repeatability). Later, when the best peptides have been selected, stable isotope-labeled peptides can be synthesized and used for absolute quantification. Proteomics SRM technology has been developed along the years, and the new scheduling capabilities enabled larger screening capacities, and consequently detection and quantification of large sets of analytes (several hundreds). By allowing the quantification of a large number of specific proteins across a variety of samples, SRM is being commonly used as a biomarker verification tool in clinical applications (Chen et al., 2017; Cho et al., 2011; Kim et al., 2015b; Mermelekas et al., 2015).

As illustrated in Figure I-15, once a robust targeted approach is optimized and available, the final candidates are subjected to an ecotoxicological validation. This phase should be conducted in the laboratory, in terms of the interpretation of adverse effects and the characterization of confounding factors and, finally, *in situ* for the determination of intrinsic variability, biomarker discrimination potential, and reference value ranges.

3. Gammarids as sentinel species for environmental monitoring of freshwaters

Among Amphipods, the *Gammarus* species (i.e. gammarids) are one of the most represented and widespread across several European inland and coastal habitats. Around 100 species of gammarids are known to exist in freshwater environments, distributed in the Northern Hemisphere (Barnard and Barnard, 1983). Their distribution is dependent on some abiotic factors such as temperature, salinity, oxygen, acidity, and pollution. Gammarids represent key species in aquatic systems because they are present in high densities, playing a major role in leaf-litter breakdown processes and constituting a food reserve for macroinvertebrates and fishes. Gammarids are known to successfully colonize and invade disturbance-prone ecosystems, probably due to their wide trophic repertoire, since they feed as herbivores, detritivores, predators, and even through cannibalism (Kunz et al., 2010). They are also sensitive to a wide range of pollutants representing therefore suitable species for use in laboratory and field ecotoxicological studies. Their use in Ecotoxicology started in the 1970s, using mainly two species - *Gammarus pulex* and *Gammarus fossarum* -, for which many sub-individual/individual markers have been developed (Chaumot et al., 2015). In this thesis, the test organisms used belonged to the species *Gammarus fossarum* (Figure I-16). In this section, a focus is done in the characterization of the species, their endocrine systems, and their use in Ecotoxicology, focusing in the molecular biomarkers developed for toxicity testing.

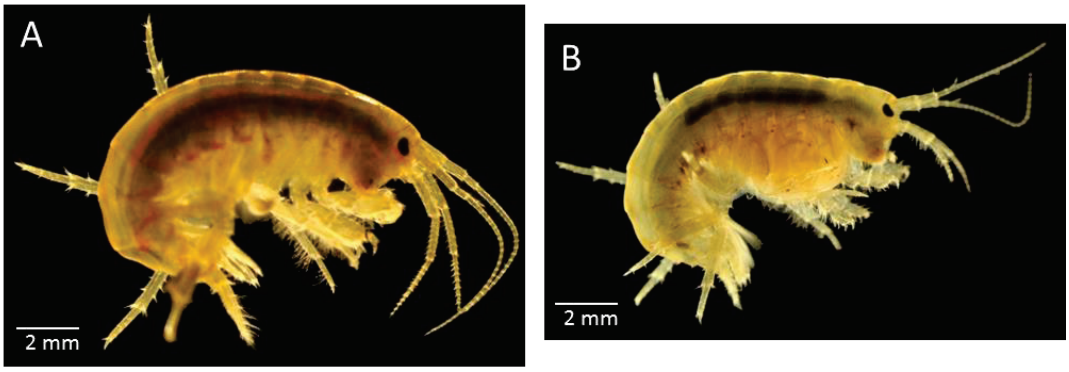


Figure I-16 – Mature male (left) and female (right) organisms of *Gammarus fossarum*. (Source: Hervé Quéau, Irstea Lyon)

3.1. Classification, morphology and ecology of *Gammarus fossarum*

The systematic classification of *Gammarus fossarum* is the following:

- Kingdom: Animalia
- Phylum: Arthropoda
- Subphylum: Crustacea
- Class: Malacostraca
- Order: Amphipoda
- Suborder: Senticaudata
- Family: Gammaridae
- Genus: ***Gammarus***

As represented in Figure I-17, amphipods are branchial-type crustaceans, with two pairs of antennas. The calcified chitin tegument forms the exoskeleton, a rigid and resistant structure that supports and protects the body of the organisms. Gammarids are aquatic amphipods that are distinguished by a bilateral, laterally flattened body. The body is divided in 4 distinct regions: the prosome (head), the mesosome (thorax), the metasome, and the urosome (the latter two form the abdomen).

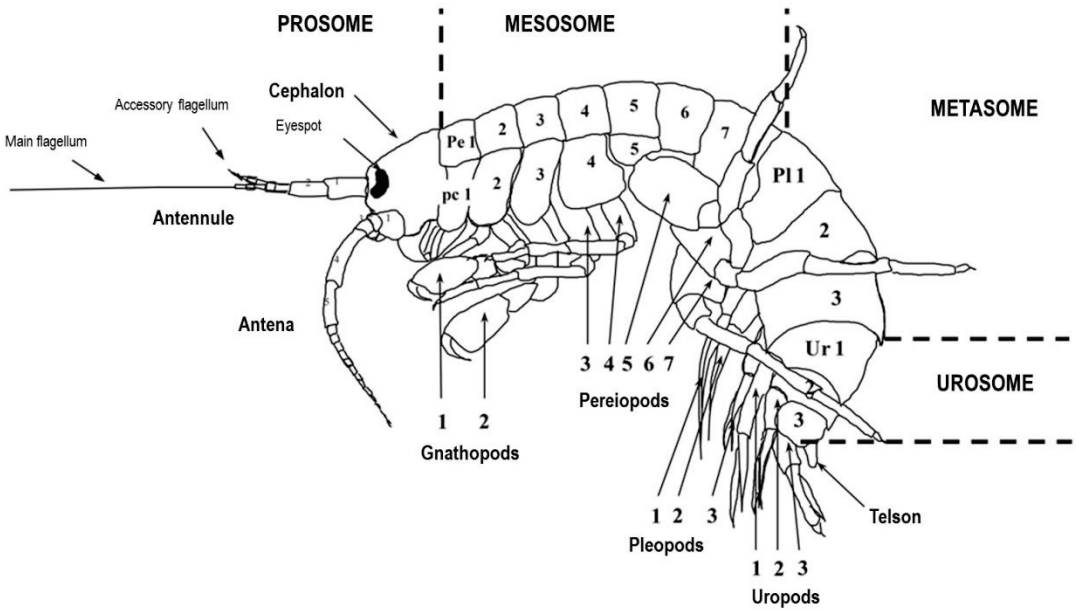


Figure I-17 – Lateral view of a gammarid (adapted from Chevreux and Fage 1970)

Their adult size revolves around 20 mm, and their morphology is characterized by a regular dorsal convexity. In the mesosome, two pair of gnathopods allows the animal to fixate to a substrate, and the 4 pair of pereiopods have a locomotion function. In the metasome, the pleopods are continuously agitated to ventilate the branchial cavity, and play an important role in locomotion. Their distribution in Europe is represented in Figure I-18. Gammarids are largely distributed in aquatic systems, from the Pyrenees to the Balkans. They are found mostly in sites that serve as shelter from predators and as supply of food: under rocks, in gravel, in coarse substrates, and among living and dead vegetation (Kunz et al., 2010). Since they are present in high densities, they represent important food sources for several species of macroinvertebrates, fishes, amphibians, and birds.

Gammarids have a complex life cycle that can last up to 2 years. They have a discontinuous growth with successive molts, and can reproduce during the whole year (with peaks in spring and end of summer) (Coulaud et al., 2014; Felten, 2003). The molt cycle is temperature and sex dependent. The reproductive cycle of gammarids is concomitant to the molt cycle, and in the case of female gammarids, perfectly synchronized.

Figure I-18- Distribution of *Gammarus fossarum* in Europe. (Barnard and Barnard, 1983)

Female molt and reproductive cycles were finely characterized by Geffard et al. (Geffard et al., 2010). The molt cycle lasts roughly 30 days (at 12°C) in an adult female, and five molt stages were defined: AB for postmolt, C1 and C2 for intermolt, and D1 and D2 for premolt. Each stage is determined by the microscopic observation of the development of the new cuticule and the integumental changes of dactylpodite and protopodite from the first and second periopod pairs (Figure I-19).

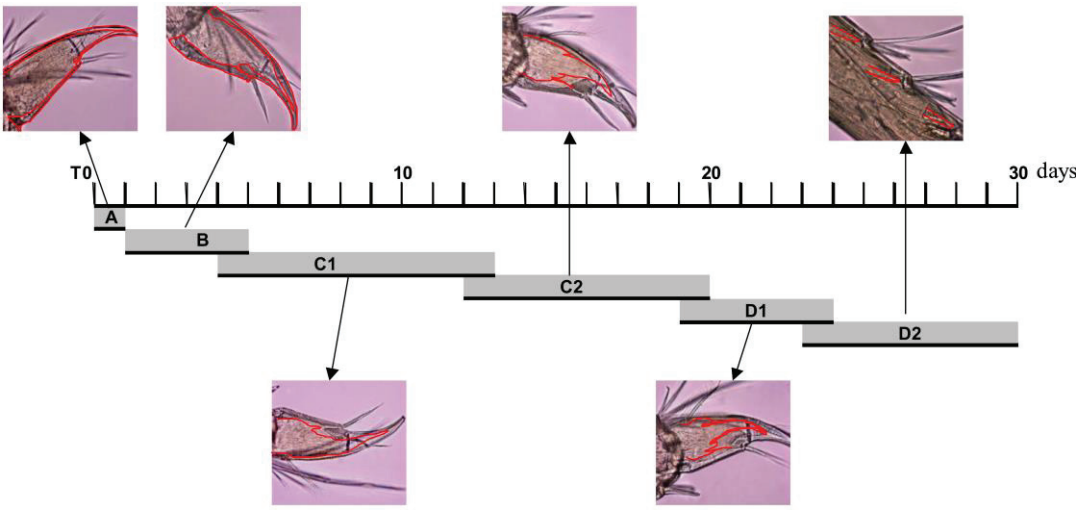


Figure I-19 – Integumental changes of dactylopodite and protopodite from first and second periopod pairs in *Gammarus fossarum* during its molt cycle. (Geffard et al., 2010)

When the female is ready for coupling, the male grabs and holds the female under and parallel to his body using the first pair of gnathopods, and performs all the swimming movements while carrying her (precopula pair). When the female sheds of the cuticule (ecdysis), the male can mate with her. After copulation, the precopula pair is separated and the female carries the eggs in the brood pouch (Kunz et al., 2010). The juvenile gammarids swim out of the brood pouch just before female ecdysis. Right after ecdysis, the female lays new mature oocytes that will be soon fertilized by a male. In parallel, in the gonads, a new lot of primary oocytes initiate vitellogenesis. Ecdysis is therefore both the starting point and the end-result of gonad maturation, as well as embryonic development in the marsupium (Figure I-20).

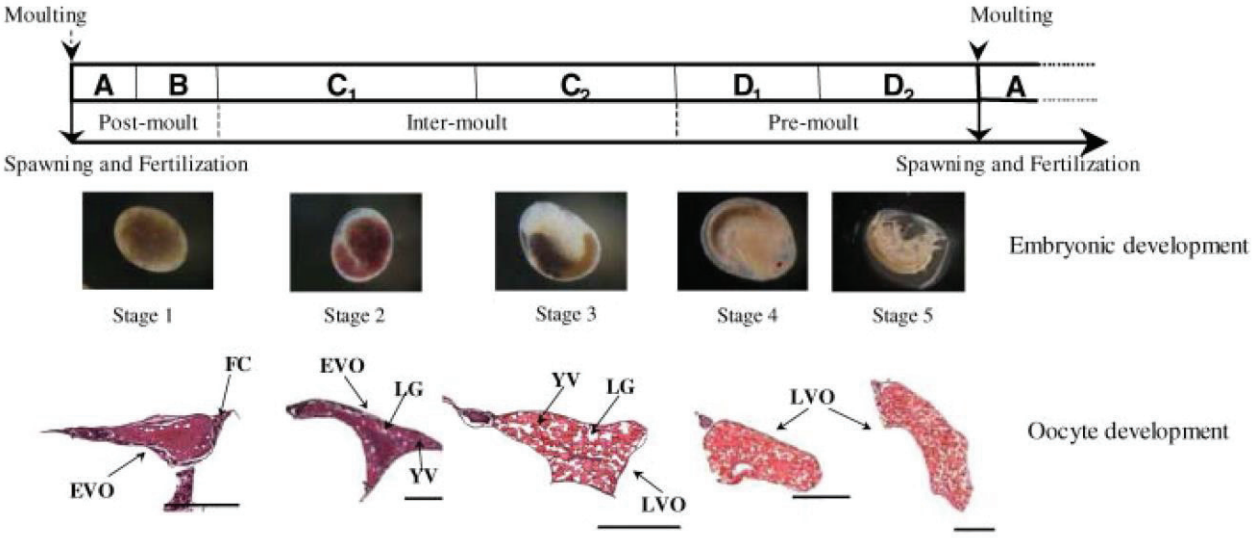


Figure I-20 - Synchronization between molt cycle, the embryonic development and vitellogenesis in female *Gammarus fossarum*. Scale bars: 100µm for AB and C1 molting stages and 200µm for C2, D1, and D2 molting stages. FC=follicle cells; EVO=early vitellogenic oocyte;LVO=late vitellogenic oocyte;YV=yolk vesicle;LG=lipid globule. (Geffard et al., 2010)

3.2. Crustacean endocrine systems

Generally, endocrine regulation is a consequence of interaction between the nervous system and the endocrine system of organisms. The interactions between them are mediated through a biochemical cascade initiated by neuroendocrine peptides. Environmental stimuli are perceived by the central nervous system, which in turn releases the neuropeptides through the action of neurotransmitters and neuromodulators. Neuropeptides either regulate directly the physiological processes, or modulate the production and secretion of hormones in the endocrine glands. The subsequent action of hormones is notably based on the presence of highly specific protein nuclear receptors in the target cells, and the formation of a complex hormone-receptor that mediate the biochemical response (Figure I-21 left panel). The physiological response is dependent of the hormonal balance between the different chemical mediators released. While the endocrine system of insects has been extensively studied, the crustacean endocrine system is still poorly understood (Breitholtz, 2013). However, the little that is known indicates that crustaceans have similar endocrine systems to insects (Covi et al., 2012; LeBlanc, 2007), due to their phylogenetic vicinity (Hui et al., 2010; Wilson et al., 2000).

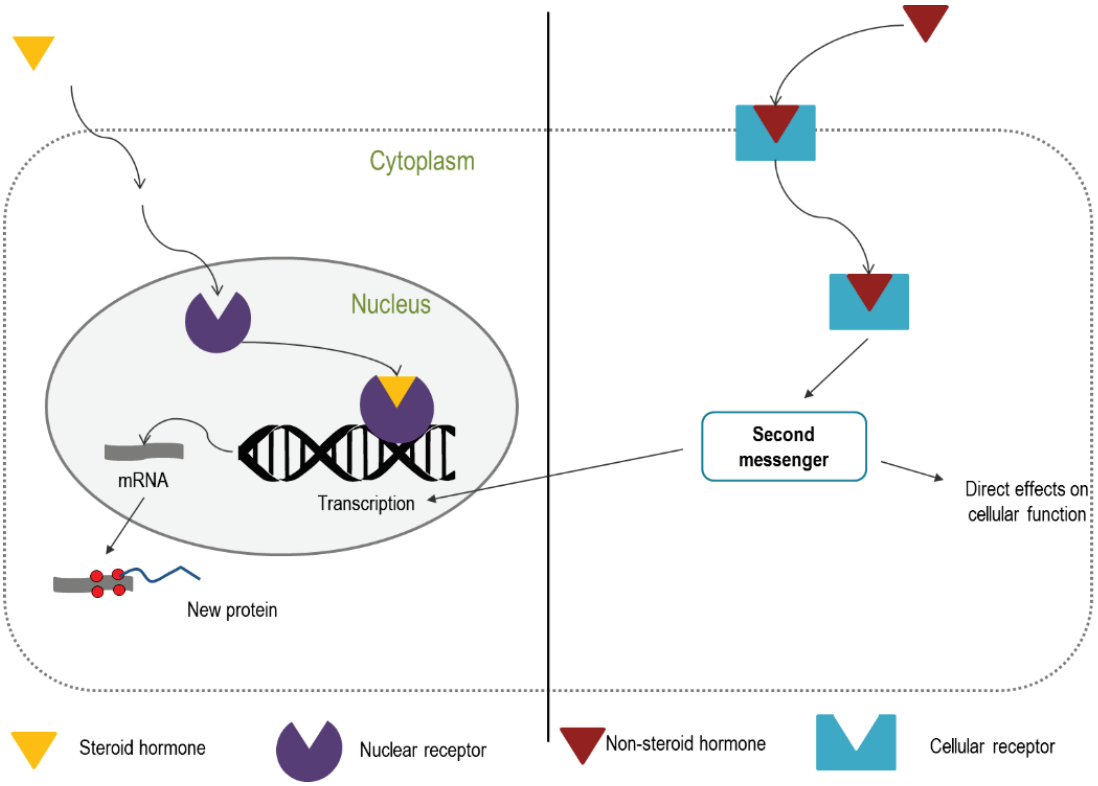


Figure I-21 – Two types of hormonal action (genomic and non-genomic signaling pathways). Hydrophobic hormones diffuse through the membrane and bind to an internal receptor; the hormone-receptor complex binds to the DNA and initiates protein synthesis (left panel). Hydrophylic hormones bind to a cell surface receptor; a seconder messenger is formed through enzymatic-catalyzed reactions within the cell; the second messenger will energize reactions to effect hormonal action (right panel).

There are four hormone/neurohormone classes identified in crustaceans – neuropeptides, steroid hormones, terpenoid hormones, and glycopeptides -, secreted by one of the three main endocrine tissues: 1/ the Y organ, 2/ the androgenic gland or ovary, and 3/ the mandibular organ (Figure I-22). The complex formed by the **X organ** and the **sinus gland** is the main neuroendocrine organ in crustaceans (Hyne, 2011). Neuropeptides are synthetized in the X organ, and transported and stored in the sinus gland before being released into the hemolymph. The peptide hormones released from the X organ regulate important physiological functions such as growth, molting, and reproduction. Some of the released neuropeptides are hormones: crustacean hyperglycemic hormone (CHH), the molt-inhibiting hormone (MIH), gonad or vitellogenesis-inhibiting hormone (GIH/VIH), mandibular organ-inhibiting hormone (MOIH), (Chan et al., 2003; De Kleijn and Van Herp, 1995).

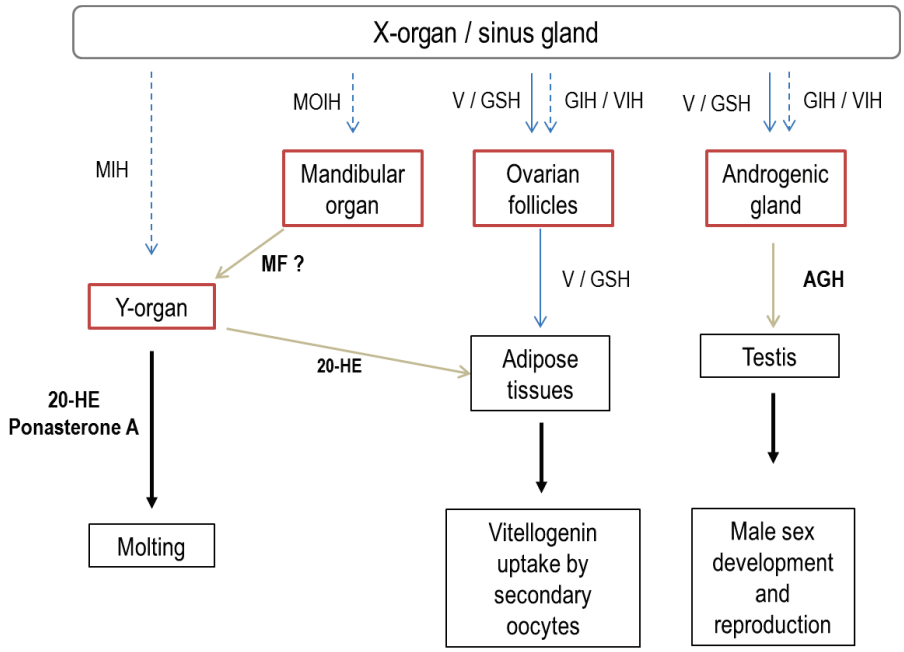


Figure I-22 – Amphipod regulatory endocrine system. Stimulatory effects are indicated with solid arrows, and inhibitory effects with dashed arrows. The neuropeptides released by the complex [X-organ/sinus gland] will regulate the release of hormones from the target tissues. Although is still uncertain, it is possible that the MF released by the mandibular organ upregulates the Y-organ. Adapted from (Hyne, 2011)

The **androgenic gland**, until now only found in malacostracan crustaceans (Isopoda, Decapoda), produces the androgenic gland hormone (AGH) that regulates sexual differentiation of the male reproductive system, its functioning, and the development of male secondary sexual characteristics (LeBlanc, 2007). This is only the case in males, since female organisms do not develop the androgenic gland - sexual differentiation in female is induced spontaneously. The production of AGH is influenced by the neuropeptides GIH/GSH (GSH -> gonad stimulating hormone) (Hyne, 2011).

The **mandibular organ** produces the terpenoid hormone methyl farnesoate (MF), an analog of the insect juvenile hormone III (JH). In insects, MF is a precursor of JH (Figure I-23). Along the evolution, insects acquired one epoxidase responsible for adding the epoxide group to MF and thus converting it to JH. The neuropeptide MOIH regulates the production of MF, whose functions are still not clear. Nevertheless, MF is thought to play important roles in gonad development and molting (Borst et al., 1987; Nagaraju; Tamone and Chang, 1993) in several crustaceans, notably daphnids and decapods (Nagaraju and Borst, 2008; Olmstead and Leblanc, 2002; Qu et al., 2015; Tiu et al., 2012).

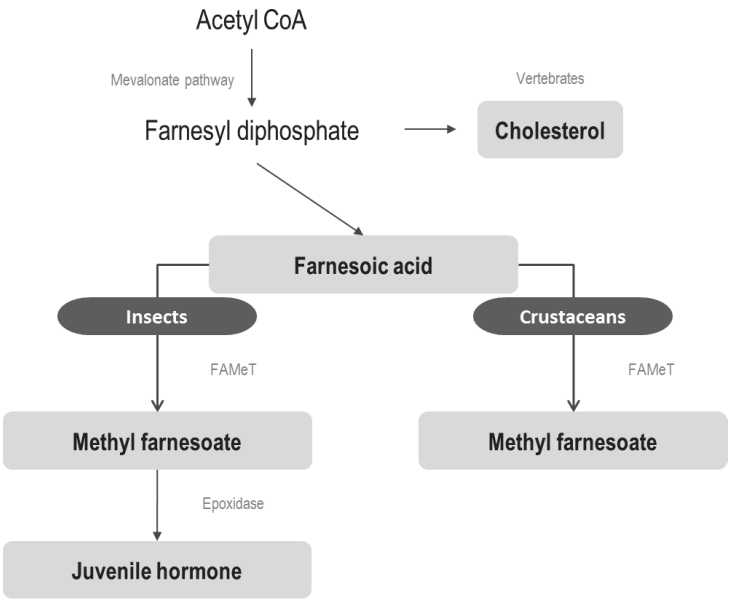


Figure I-23 – Simplified pathway of terpenoid synthesis in insects and crustaceans.

The **Y organ** synthesizes ecdysteroids from cholesterol. Ecdysone hormone is essential for the molting processes. In crustaceans, the Y organ releases ecdysone into the hemolymph, where it is converted to its active form 20-hydroxyecdysone (20-HE) and/or ponasterone A (25-deoxy-20-hydroxyecdysone). Both the X organ/sinus gland complex, and the Y organ regulate molting by controlling the ecdysone titers during the cycle (Figure I-22). Ecdysone production is negatively regulated by the neuropeptide MIH, and possibly positively regulated by MF (LeBlanc, 2007; Subramoniam, 2000). When stimulated, the X organ decreases the MIH titers in order to allow the production of ecdysone in the Y organ. In general, the ecdysone titer is low after ecdysis and during intermolt. A peak occurs at the beginning of premolt, followed by a strong decrease just before ecdysis. These ecdysone titers can vary in larger crustaceans (Chang, 1993; Mykles, 2011). The action of ecdysone is intermediated by the heterodimer protein nuclear receptor EcR/RXR (Ecdysone Receptor / Retinoid X Receptor). Nuclear receptors are paralog genes that have a highly evolutionary conserved domain called DNA-binding domain (DBD), and other conserved amino-acyl sequences in functional areas in the ligand-binding domain (LBD). Figure I-24 shows the typical structure of a nuclear receptor. The variable NH2-terminal region (A/B) contains the ligand-independent AF-1 transactivation domain. The conserved DNA-binding domain (C) is responsible for the recognition of specific DNA sequences. A variable linker region (D)

connected to the conserved (E/F) region that contains the ligand-binding domain, the dimerization surface, and the ligand-dependent AF-2 transactivation domain.

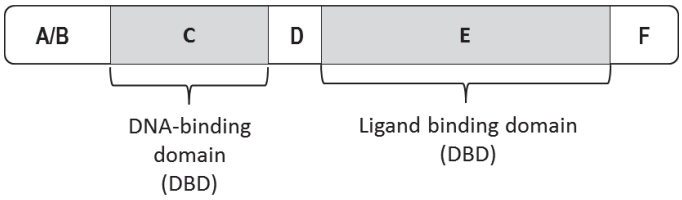


Figure I-24 – Schematic representation of the different functional domains of a nuclear receptor. Adapted from (Aranda and Pascual, 2001).

20-HE binds to and activates the EcR/RXR complex, initiating a cascade of gene-regulatory events that will mediate molting and reproduction (Figure I-25). The ecdysone pathway is well characterized in insects, however much less is understood in crustacean species. However, early response genes (transcription factors) are thought to be similar, and include the nuclear receptors ecdysone induced protein 75B (E75), ecdysone induced protein 78C (E78), and hormone receptor 3 (HR3), and the transcription factors broad-complex (BR), ecdysone induced protein 74EF (E74). In insects, the products of early genes represses the early gene transcription, and trigger the expression of late or effector genes, which perform metamorphic changes like programmed cell death, growth, and cell differentiation (Jiang et al., 1997).

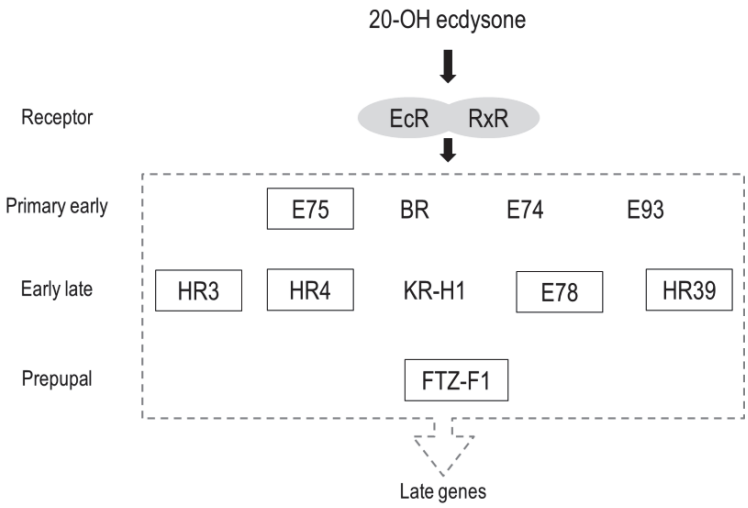


Figure I-25 – Main molecules involved in the ecdysone regulatory cascade in insects. Nuclear receptors are boxed. Adapted from (Bonneton et al., 2008)

Ecdysteroid receptor is a successful target for insecticides that act as ecdysone agonists. These chemicals are not easily metabolized, and upon binding to the receptor, they disrupt the molting process, leading to the death of the insect. Several studies have also shown the susceptibility of the ecdysone pathway to different types of endocrine disruptors in other arthropod species. Modulations of RXR were observed after tributyltin (TBT) exposure in the decapod *Crangon crangon* (Verhaegen et al., 2011), and in embryos from *Daphnia magna* females exposed to PYR (Wang et al., 2007). EcR was found induced in *Chironomus riparius* after exposure to a wide range of contaminants: n-butyl benzyl phthalate (BBP) (Herrero et al., 2015), triclosan (Martínez-Paz et al., 2017), TBT (Morales et al., 2013), PCP (Morales et al., 2014), bisphenol A (Planelló et al., 2008). Inductions were also observed after bisphenol A and tebufenozide (TEB) injection in the butterfly *Sesamia nonagrioides* (Kontogiannatos et al., 2014), and after 30h exposure to fipronil in the copepod *Amphiascus tenuiremis* (Gaertner et al., 2012).

3.3. *Gammarus* in Ecotoxicology and currently used biomarkers

Amphipod crustaceans are commonly used species in aquatic ecotoxicology. In Europe, the most used genre is *Gammarus*, due to its large distribution in European rivers (Figure I-18). *Parhyale hawaiensis* is found along tropical coastlines, and is a commonly used model organism for developmental studies and evolution of morphological diversity within arthropods (Liubicich et al., 2009; Pavlopoulos et al., 2009). In Australia, the species *Melita plumulosa* is largely employed for toxicity assessment of sediments (Hook et al., 2014b; Hyne et al., 2005). In North and Central America, *Hyalella azteca* is the most used species (Major et al., 2013). This species gained an even bigger relevance for ecotoxicological molecular studies seen that the genome was recently sequenced and assembled. The growing use of amphipod species for toxicity studies is explained by their status of key species in the respective ecosystems, but also due to their widespread representation, easy identification, sampling, sensitivity/resistance to toxicants, and maintenance in the laboratory.

In gammarids, the impact of toxicants has been studied in the last decades throughout different levels of biological organization. The impact of pollutants can be assessed by measuring life-history traits such as reproductive features (Geffard et al.,

2010), locomotor behavior and feeding rate (Coulaud et al., 2011), as well as using molecular biomarkers related to diverse modes of action such as neurotoxicity (Xuereb et al., 2009a), genotoxicity (Lacaze et al., 2011b), and digestive impairments (Dedourge-Geffard et al., 2013). In Table I-4 are listed examples of works that employed molecular biomarkers to study toxic impacts in different species of gammarids.

Table I-4 – Examples of studies employing molecular biomarkers in *Gammarus* species.

Experimental condition	Stress	Biomarkers	Species	Reference
Laboratory	Thermal stress	CAT, HSP70, sHSP, POD and GST	*G. oceanicus*	Bedulina 2010
Laboratory	Natural organic matter	CAT, POD, HSP70 et sHSP	*G. pulex*	Bedulina 2010
Laboratory	Salinity	Na^+/K^+ATPase	*G. duebeni*	Brooks and Lloyd Mills 2006
Active biomonitoring	Large scale study *in situ*	Amylase, cellulase, trypsin	*G. fossarum*	Charron 2013
Laboratory	Copper-rich sediments	MTs	*G. locusta*	Correia 2013
Active biomonitoring	*In situ* study in metal-polluted rivers	Amylase, cellulase, trypsin	*G. fossarum*	Dedourge-Geffard 2013
Laboratory	Cadmium	Na^+/K^+ATPase	*G. pulex*	Felten 2008
Laboratory	Cadmium, parasitism	HSP70 and MT	*G. fossarum*	Frank 2013
Laboratory	Cadmium, parasitism	MTs	*G. roeseli*	Gismondi 2012
Laboratory	Cadmium	Na^+/K^+ATPase	*G. fossarum*	Issartel 2010
Laboratory	Arthropod hormones, pesticides, *in situ* caging	Vtg	*G. fossarum*	Simon 2010, Jubeaux
Active biomonitoring	Organic pesticides	GST and AChE	*G. pulex*	Maltby and Hills 2008
Laboratory	Chlorpyrifos	AChE	*G. fossarum*	Xuereb 2009

In a recent review, Trapp et al. (Trapp et al., 2014a) summarized molecular biomarkers developed in *Gammarus* genus. In that review, the authors highlighted several biomarkers already developed and validated, such as the MT-like proteins (Stuhlbacher and Maltby, 1992; Correia et al., 2001; Geffard et al., 2007; Gismondi et al., 2012), AChE activity (McLoughlin et al., 2000; Xuereb et al., 2007, 2009a), phase II xenobiotic transformation enzymes (GST, GPX, CAT, and SOD) (Maltby and Hills, 2008; Bedulina et al., 2010a; Turja et al., 2014), heat shock proteins from different families (Schirling et al., 2006; Scheil et al., 2008; Bedulina et al., 2010a), changes in digestive enzymes including amylase, cellulase, endoglucanase, esterase, trypsin, β-galactosidase, and β-glucosidase (Dedourge-Geffard et al., 2009; Charron et al., 2013); and, finally, activity of the sodium pump Na^+/K^+ ATPase (Felten et al., 2008a,b; Issartel et al., 2010).

The large majority of biomarkers are protein biomarkers, but mainly enzymes using indirect methods relying upon determining their activity.

As highlighted in section 1.2.2, the use of biomarkers in field surveys is still scarce, mainly due to the difficulty of predicting/controlling confounding biological and environmental factors. This leads to difficulties in establishing direct relationships between biomarker changes and field contamination levels. The application of caging strategies with laboratory-calibrated organisms can reduce variability sources related to biotic factors. Thus, a robust active biomonitoring strategy based on caged organisms from the same reference *Gammarus fossarum* population was developed (Besse et al., 2013). This strategy, combined with mathematical and modeling approaches, was used in several *in situ* studies in order to propose benchmark values for specific biomarkers taking into account the natural variability of biomarkers in relation to environmental factors (Charron et al., 2014; Charron et al., 2013; Coulaud et al., 2011; Lacaze et al., 2011a; Xuereb et al., 2009b). This improved drastically the accurate interpretation of biomarker responses in terms of toxicity.

3.4. Current status in "omics" approaches using Gammarus

Despite being currently used as a sentinel species for toxicity assessment in aquatic environments, gammarids are still non-model species due to the lack of genomic/proteomic information available for amphipods. Consequently, few "omics" studies are available using these species. Few transcriptomic studies were found using gammarids: 1/ a reference hepatopancreas transcriptome of *Gammarus pulex* (Gismondi and Thomé, 2016), 2/ reference embryonic transcriptome for *Gammarus chevreuxi* (Truebano et al., 2016), 3/ transcriptome of *Gammarus chevreuxi* exposed to chronic hypoxia (Collins et al., 2017), 4/ transcriptomic screening approach to study the molecular basis of crustacean intersexuality, using *Echinogammarus marinus* (Short et al., 2014a). These studies were generally performed in order to explore the transcriptome in future ecotoxicological studies, especially to determine contaminant modes of action. One targeted metabolomic study (Gómez-Canela et al., 2016) allowed the quantitation of 29 metabolites of *Gammarus pulex* by LC-MS/MS. Pathway

alterations related to protein synthesis, oxidative stress and signaling cascades were observed in exposed samples.

In recent years, proteomics has been increasingly used in ecotoxicological studies. The oldest proteomic work found using gammarids dates back to 2006. In this work (Ponton et al., 2006), the authors aimed at studying the brains of parasite-infected *Gammarus insensibilis* and *Gammarus pulex*, in order to draw conclusions about the proximate causes involved in the parasite-induced alterations of host behavior. By applying 2D-PAGE, 556 and 838 proteins spots were obtained in *G. insensibilis* and *G. pulex*, respectively. Around 30 proteins were identified by mass spectrometry using peptide mass fingerprinting. In 2010, another study aimed at characterizing the effects of PCB contaminants in the protein profile of *G. pulex* (Leroy et al., 2010). Using again the 2D-PAGE as separation tool, 560 protein spots were detected, with 21 proteins exhibiting significant expression differences in PCB exposed identified by MS. This was the first ecotoxicological study using proteomics in gammarids, and allowed gaining insights into some biochemical mechanisms disturbed by the presence of the coplanar cogenders CB77 and CB169 in *G. pulex*. Gismondi et al. (Gismondi et al., 2015) performed a gel-free proteomic approach to study the gender differences in responses from *G. pulex* exposed to BDE-47 (Gismondi et al., 2015), and the response of *G. fossarum* to chronic exposures to three heavy metals (Gismondi et al., 2017). These works identified 45 and 35 proteins, respectively, of which 25 and 23 were significantly differently expressed between conditions. While in the first study, differences in responses of males and females to a specific contaminant were highlighted, a mechanism of action for BDE-47 was difficult to establish due to the limited number of identified proteins. In chronically heavy metal exposed gammarids, major effects were observed in reproduction. Another work also analyzed the potential of comparative proteomics as a multi-marker approach of metal contamination in *G. pulex* (Vellinger et al., 2016). Despite having identified 264 proteins (in all studied conditions), the majority were involved in housekeeping functions or energy metabolism, and consequently not useful to report for relevant pathways from an ecotoxicological point of view.

Undeniably, the common factor in all of these works is the limited number of identified proteins by MS. This limitation was surpassed with the application of the proteogenomics approach (described in section 2.2.) in 2014 (Trapp et al., 2014b). This

study used the potentialities of combining transcriptomics and proteomics in order to unravel a considerably higher number of specific protein sequences from *Gammarus fossarum*. Based on deep RNA sequencing, a comprehensive open reading frame (ORF) database was created. By means of a shotgun tandem mass spectrometry analysis of reproductive tissues, 1873 mass-spectrometry-certified proteins (of which 218 were lineage specific) were identified. This work was the beginning of a series of studies based on this novel resource obtained from *G. fossarum*. A proteomic investigation of the response of male gammarids to endocrine disruptors (Trapp et al., 2015) lead to strong modulations of 14 male-specific proteins in testis, thus not "déjà-vu" proteins, in a total of 871 proteins analyzed. Another functional study was also performed in order to identify the yolk proteins present in the female *G. fossarum* proteome (Trapp et al., 2016). Through a shotgun proteomic temporal analysis during oogenesis and embryogenesis, eight proteins originating from different families of the large lipid transfer protein superfamily (which includes vitellogenin) were identified as potential yolk proteins. All of these studies allowed the identification of candidate biomarkers for toxicity assessment in this species. As highlighted in Figure I-15, one can use the molecular and analytical information obtained from the shotgun experiments to develop quantitation methods for either 1/validating the biomarkers or 2/ develop targeted analytical tools allowing to quantify several proteins simultaneously. One MS targeted approach by SRM was already developed and validated for identifying and quantifying Vtg in *Gammarus fossarum* (Jubeaux et al., 2012c; Simon et al., 2010). This methodology has been used to investigate whether Vtg measurement could be a specific endocrine disruption biomarker in males and used as an indicator of feminization. The protein sequence of the Vtg protein for using in the SRM approach was obtained from polymerase chain reaction (PCR) experiments. However, its proposal as endocrine disruption biomarker for male feminization was discarded after laboratory and field experiments that yield very high variability and very low Vtg inductions in exposed organisms (Jubeaux et al., 2012b). Similar conclusions were reached by Short et al. (Short et al., 2014b) when analyzing expression levels of two Vtg genes in *Echinogammarus marinus*, highlighting the fact that invertebrate biomarkers must be redeveloped and revalidated *de novo*, rather than vertebrate-derived.

4. Objectives of the thesis

Numerous biological parameters such as hormones, enzymes, and other proteins, are already validated as tools for evaluating toxic effects in aquatic organisms However, these tools are available mainly for vertebrate species, for which there is a great many amount of molecular information available in databases. The majority of biomarkers developed in invertebrate species are straightforwardly transferred from vertebrates, without a previous systematic molecular characterization. In some cases, when proteins are highly conserved during biological evolution, this transfer leads to robust and accurate biomarkers, such as the measurement of cholinesterase and Na^+K^+ ATPase activities. However, for proteins for which sequence and/or physiological function have significantly changed during biological evolution, this cross taxa transfer approach leads to inconsistent results in invertebrates.

Taking into account these obstacles, the proteogenomic approach was applied to *G. fossarum* in order to unravel specific protein sequences of this ecotoxicologically-relevant species. These works allowed obtaining databases of coding gene sequences and proteins (1873 protein sequences) specific of the species (Trapp et al., 2014b). This was achieved even with the use of a RNAseq-derived transcriptome containing assembly errors due to the absence of a reference genome for reliable mapping. Thanks to this innovative work in invertebrate ecotoxicology, it became possible to unravel high-throughput molecular information on any species of interest. By having these tools available, the development, application, and validation of new biomarkers in invertebrate ecotoxicology gained new perspectives.

In this thesis project, I used these available gene and protein catalogues for developing and validating molecular biomarkers of toxicity in *G. fossarum*. The project was based on both laboratory and field experiments. In order to answer to the proposed objectives, the work was divided in two major parts:

Development and application of a multiplexed protein biomarker measurement by mass spectrometry

This axis of the thesis resulted from a collaboration between my host laboratory of ecotoxicology in Irstea Lyon-Villeurbanne, and the “Analytical Sciences Institute” (ISA

Lyon), with the objective of developing an analytical strategy to quantify simultaneously several dozens of proteins by mass spectrometry. This innovative strategy aimed at overcoming the difficulties of the lack of biomarker direct quantification methodologies, and the need for developing multibiomarker strategies for application in field studies. Firstly, an extensive list of peptides/proteins was proposed for SRM method development. After validation of the analytical methodology (Publication nº 1), the method was applied for inferring the implication of the candidate biomarkers in key physiological processes such as female molt cycle and male spermatogenesis (Publication nº2). The suitability of the method as a tool for assessing the health status of organisms was then evaluated through ecotoxicological studies in laboratory (Publication nº2) and in the field (Publication nº3).

Endocrine disruption biomarkers

This axis of the thesis resulted from a collaboration between my host laboratory of ecotoxicology in Irstea Lyon-Villeurbanne, and the "Functional Genomics Institut of Lyon" (IGFL), and the "French Alternative Energies and Atomic Energy Commission". The objective was to develop specific biomarkers of endocrine disruptions in *G. fossarum*. Two approaches were used to answer this objective. The first approach consisted in analyzing the **proteome** of male gonads from organisms exposed to the endocrine disrupting chemical pyriproxyfen (PYR). Through comparative proteomics of control and exposed samples, significantly modulated proteins were proposed as biomarkers of reproductive disorders (Publication nº4).

As highlighted in section 2.1.2, shotgun proteomics methods generally do not allow the detection of low-abundant proteins. However, the most physiologically relevant molecules for biomarker development may be present in low quantities. Therefore, in this part we decided to go back to the transcriptomic information and develop biomarkers following a gene candidate approach. Through gene database and literature mining, we targeted relevant genes implied in the hormonal processes and that were not present in our protein database. We focused on hormonal nuclear receptors involved in ecdysone response. Gene sequencing and gene expression studies were performed in order to 1/ obtain the specific nucleotide sequences of the candidate genes, 2/ evaluate their pertinence as biomarkers of endocrine disruptions through

laboratory exposure to EDCs (Publication nº5), and 3/ to propose proteotypic peptides that can be incorporated into the SRM multiplex biomarker quantification method.

CHAPTER II. MATERIALS AND METHODS

This chapter details the experimental protocols implemented during this thesis.

Section 1 describes the procedures concerning organism sampling, maintenance, and biological measurements. Then, conditions of exposure in laboratory and field experiments are described.

Sections 2 and 3 describe the techniques and methodologies employed for the molecular analyses performed both at the protein and gene level.

1. Biotests in *Gammarus fossarum*

1.1. Sampling and maintenance of organisms

All the organisms used during this thesis were collected from a source population used for several years in our laboratory. This population is located in the "Bourbe" river (France) (Figure II-1), a site that presents a very high density of gammarids (>1000 individuals/m^2) and is considered by the "Réseau National des Bassins" as a pristine site.

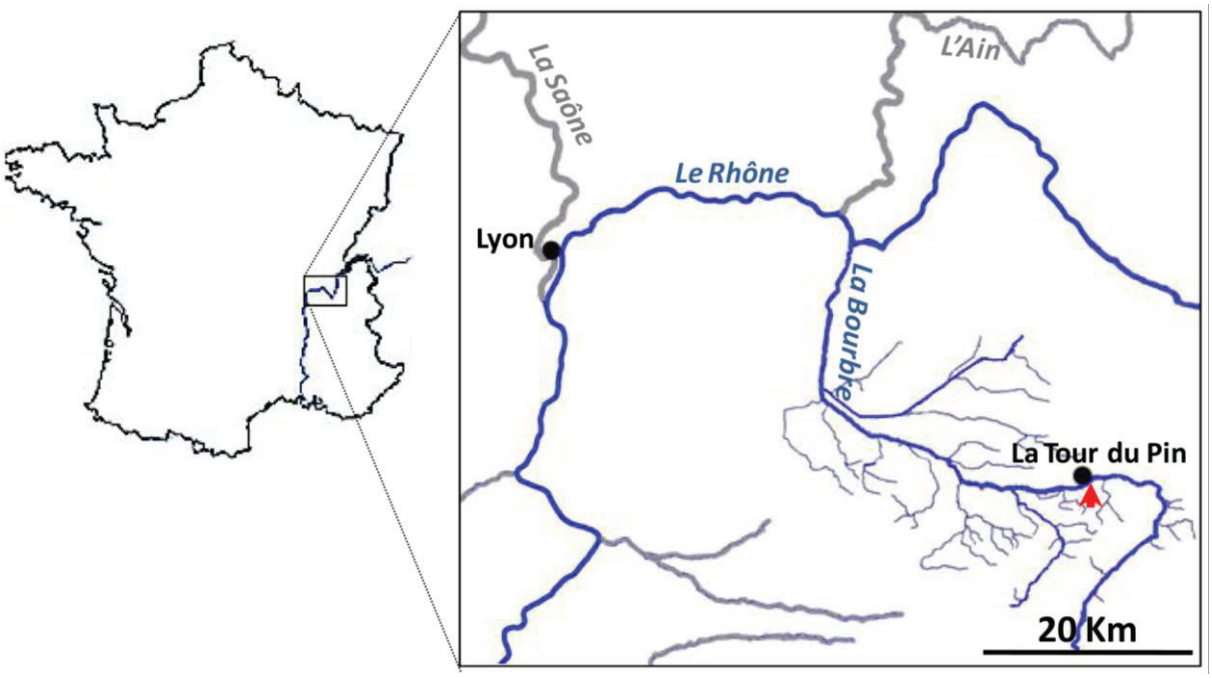


Figure II-1 – Localization of the source population used in this thesis, in La Tour du Pin. (Xuereb, 2009)

Gammarids were collected by kick-sampling, a technique adapted to watercourses of low depth, in which the operator grasps the substrate with the foot (Figure II-2). This provoked organisms to flee, disperse, and to be transported by the river flow into a fishnet previously placed downstream to the working zone. Trapped organisms were then sorted by using a sieve with a 2–2.5 mm net in order to obtain adult organisms with roughly similar sizes. Then, they were quickly transported to the laboratory in plastic buckets containing water from the sampling site.

Figure II-2 – Organism collection using the fishnet and the sieve.

In the laboratory, organisms were separated from organic debris, minerals, and other non-target species that may be present in the collected samples. Then, organisms were acclimatized to specific conditions for a period of two weeks. Gammarids were put in aquariums of 30 liters that were placed inside a thermoregulated bath at 12 ± 1°C and continuously supplied with drilled groundwater adjusted with osmotic water to a conductivity of 500 ± 20 µS/cm. A bubbling system was used for oxygenation, and a photoperiod of 16h day / 8h night was maintained. Organisms were fed *ad libitum* with alder leaves (*Alnus glutinosa*), previously conditioned for 6±1 days in water.

1.2. Biological procedures

For answering the objectives of this thesis, different experiments were performed in laboratory. Two types of experiments were done: (1) contaminant-free exposures to study the effects of physiological states (gametogenesis) in biomarker responses, and (2) contaminant exposures to assess the relevance of the candidate proteins/genes as biomarkers of toxicity in *G. fossarum*.

1.2.1. Reproductive toxicity test

We applied the experimental design of the test protocol proposed by (Geffard et al., 2010) for studying the responses of several biomarkers connected to molting, vitellogenesis, and embryogenesis in *G. fossarum*. As illustrated in Figure II-3, the experiments were launched using mature females at their D2 molt stage (i.e., in amplexus, with visible gonads and with well-developed juveniles in the brood pouch). Exposures were conducted from stage D2, just before fertilization (day 0), until they reach the C2/D1 stage during secondary vitellogenesis and stage III embryos (Figure I-20). This ensured that the female organisms were exposed during a new starting reproductive cycle, up to a determined stage of follicular maturation and marsupial embryonic development (details in (Geffard et al., 2010)), and that males are in a common reproductive status (no copulation after the initial fertilization at the beginning of the exposure). The duration of exposure is temperature-dependent, calculated thanks to a mathematical model developed for predicting *G. fossarum* molt stage according to the temperature of the water.

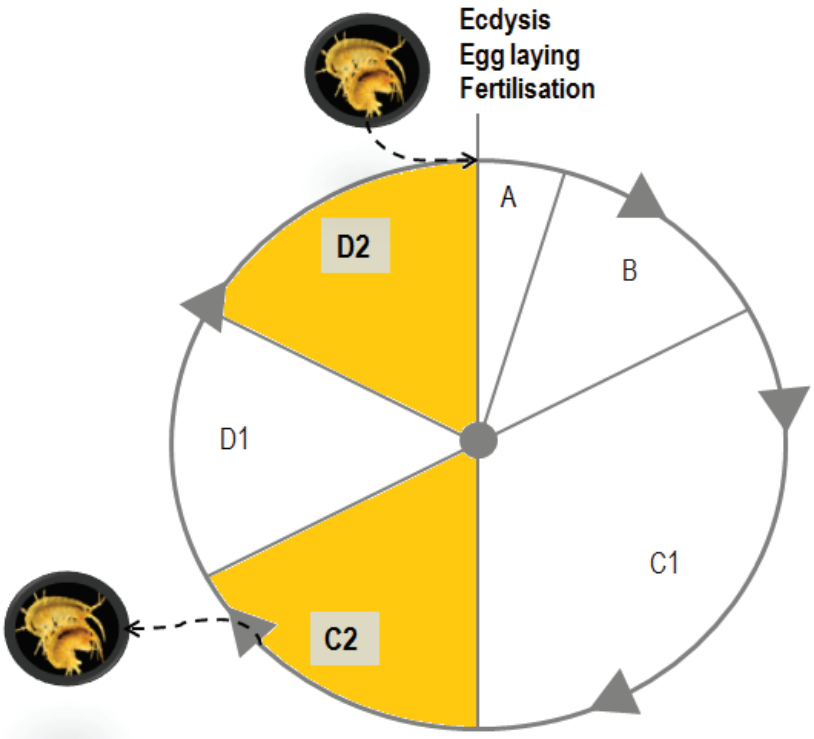


Figure II-3 – Application of the test protocol developed for *G. fossarum* organisms, based on the female molting cycle. (Geffard et al., 2010)

After exposure, several physiological parameters were analyzed, such as the molt stage, counting the number and surface of oocytes (fertility), and number of embryos (fecundity) (Figure II-4). These tests aimed at determining possible reproductive disorders induced by contamination. Molt stage is determined by analyzing the tissues in the extremity of the third and fourth pair of periopods (Figure I-19). The periopods are cut using Wecker scissors and put between two lames for microscopic observation of the detachment of the epidermis and appearance of the new exoskeleton. For determining oocyte surfaces, females are first photographed under a binocular loupe equipped with a photographic camera. The pictures are then analyzed with the SigmaScan Pro software for measuring the surfaces of oocytes in mm^2. Embryo anomalies are described in (Geffard et al., 2010), and are commonly used as biomarker of toxicant effects in females. These anomalies can be linked to detachments of the membrane, abortions, development delays, oedems, or non-specific malformations.

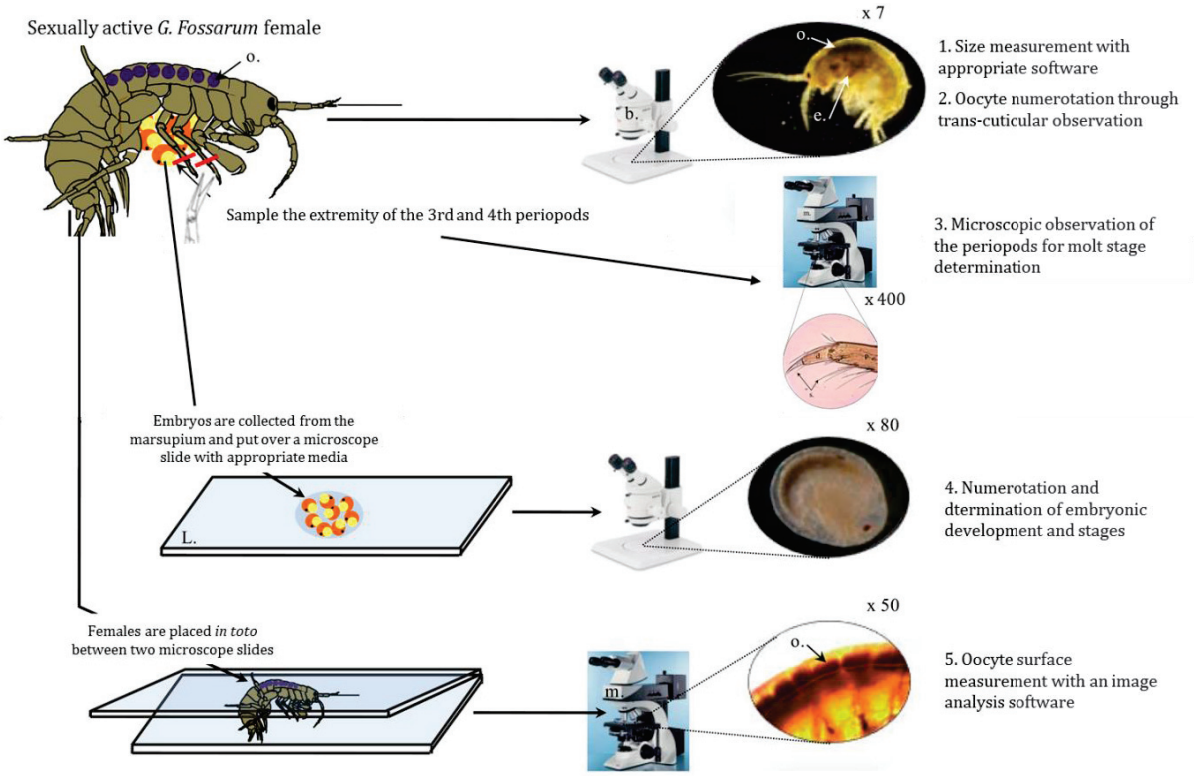


Figure II-4 – Schematic representation of the procedures for determining the gammarid size, molt stage, embryo number and stages, oocyte number and apparent surfaces. (Xuereb, 2009)

1.2.2. Gonad dissection

Male gonads were sampled under stereomicroscopic magnification using fine forceps. Briefly, as illustrated in Figure II-5, the cephalon is removed by cutting the head of the gammarid. The caecum and digestive tube are then removed through the thoracic cavity, and the dorsal and ventral cuticles are cut close to the cuticle so that the internal organs are not touched. Then, gonads, which are normally surrounded by orange lipidic droplets, are collected using fine forceps. Gonads were individually sampled in homogenization buffer, frozen in liquid nitrogen, and stored at -80°C until further analysis.

Figure II-5 – Main steps in the male *G. fossarum* gonad dissection protocol. (Lacaze, 2011)

1.2.3. Sampling of organisms at different physiological conditions

For each experiment performed, organisms were sampled from the aquariums in the stables, after acclimation, at specific reproductive stages. Molt stage of each female was determined through microscopic observation of periopods, similarly to the method described in section 1.2.1. These stages comprise the post-molt stages A and B, inter-molt stages C1 and C2, and pre-molt stages D1 and D2. Since stage A is very short (less than one day), the two post-molt stages A and B are sometimes grouped together. Embryos were removed from all the females used for the further analysis.

Male organisms sampled at different days of their spermatogenesis process were selected according to the protocols from (Trapp et al., 2015). Briefly, mature couples were sampled and individualized in 500 mL glass beakers containing the same water as the aquariums. The couples were selected based on the observation of the female, which

must be in an advanced reproductive stage (D2). Organisms were checked daily for determining the date of copulation. After copulation, males were isolated in punctured polypropylene cylinders (diameter 5cm, length 10 cm), and sampled after 1, 3, 5, and 7 days, which corresponds to the duration of the spermatogenesis cycle in *G. fossarum* (Lacaze et al., 2011a). During the experiment, organisms were fed *ad libitum*.

Finally, food deprived males were sampled after being submitted to a 14-day starvation period. A control condition of *ad libitum* fed males was simultaneously performed (feeding with alder leaves). For each condition, ten males of homogeneous sizes were used. Each individual was isolated in punctured polypropylene cylinders (diameter 5cm, length 10 cm). Throughout the exposure, the conditions were the same as in section 1.1.

After sampling, all organisms were weighed and immediately frozen in liquid nitrogen for storage at -80◦C until further analysis.

1.3. Protocols of contamination exposures

1.3.1. Laboratory exposures to model contaminants

Experiment 1

In the first experiment, male and female organisms were exposed to environmental concentrations of cadmium and lead (Cd 2µg/L and Pb 10µg/L) for seven and twenty-one days. This experiment (Figure II-6) was designed for the first axis of the thesis. We followed the exposure protocol described in section 1.2.1. Briefly, gammarids were placed in nine 500mL plastic beakers (7 couples per beaker, three beakers per condition) with constant aeration. Stock solutions were prepared in milliQ water at concentrations of 37.6 and 130.6 mg/L for cadmium and lead, respectively. The contaminated media was obtained by adding 530 and 770 µL of stock solutions of cadmium and lead, respectively, in 10L of uncontaminated drilled ground water. The constant renewal of the media was achieved by means of a peristaltic pump at a rate of 1L per day and per beaker. A piece of net (6 x 5 cm) was added as a resting surface into each beaker. Organisms were fed *ad libitum* with conditioned alder leaves (*Alnus glutinosa*). At the end of the exposure, male and female organisms were individually

sampled, rapidly weighed, frozen in liquid nitrogen, and stored at -80°C until further analysis.

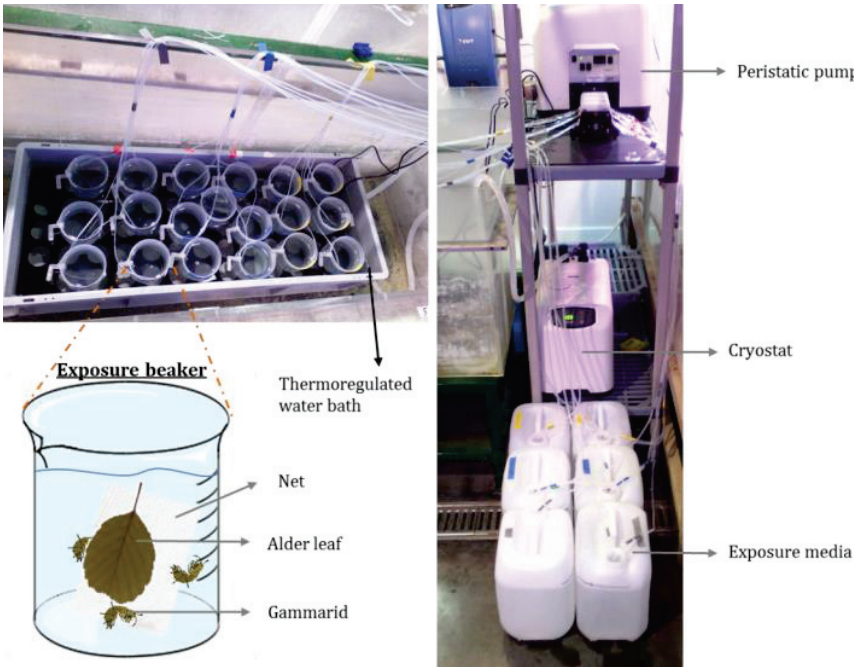


Figure II-6 – General experimental setup for the laboratory-controlled contaminant exposures.

Experiment 2

The second experiment consisted in exposing male gammarids to the insecticide PYR and analyzing the proteome response to contamination through comparative shotgun proteomics. Organisms were exposed to PYR at 0.5µg/L and 50 µg/L during two consecutive spermatogenesis cycles (15 days), following the protocol previously described in (Trapp et al., 2015) and illustrated in Figure II-7. For each condition and after copulation, 35 males were placed into five 500 mL polyethylene beakers (seven animals per beaker), under constant oxygenation. A piece of net (6 x 5 cm) was added as a resting surface into each beaker. Organisms were fed *ad libitum* with conditioned alder leaves (*Alnus glutinosa*). Controls consisted in exposure to 0.005% acetone (solvent). After seven days of exposure (end of spermatogenesis), new mature females were introduced into the beakers, allowing a second fertilization by males and thus the beginning of a second spermatogenesis cycle. During this semi-static exposure, media and nutriment leaves were renewed every two days. Survival was checked daily. Seven

days after the second fertilization, the experiment was ended. At the end of the exposure, the gonads were dissected, individually sampled, rapidly weighed, frozen in liquid nitrogen, and stored at -80°C until further analysis.

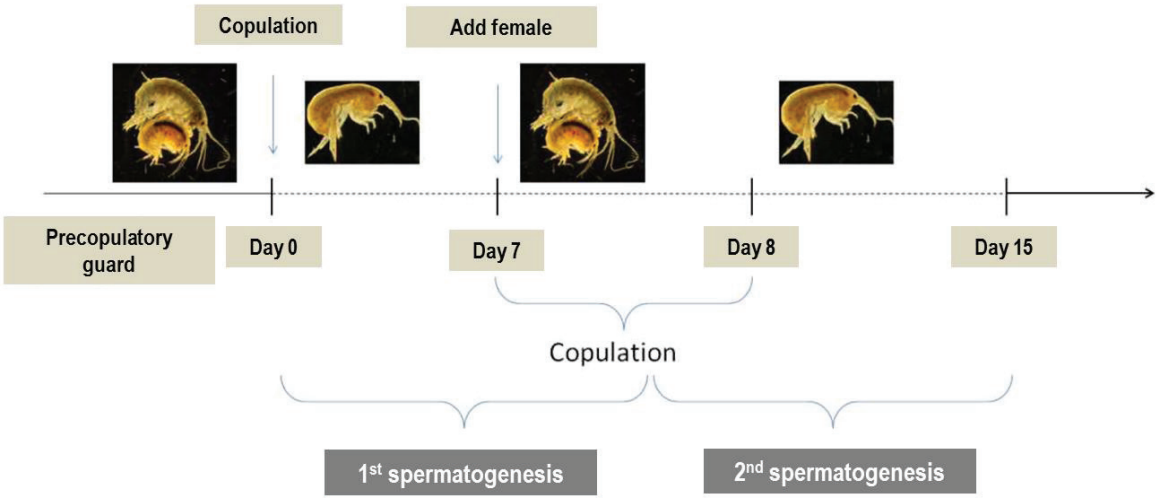


Figure II-7 – Stages of the PYR exposure protocole during two cycles of spermatogenesis. (Trapp et al., 2015)

Experiment 3

In this experiment, we tested the effects of six model molecules in female molt and oogenesis processes, using the protocol proposed by Geffar et al. (Geffard et al., 2010), in order to identify model compounds for validating the endocrine disruption biomarkers proposed in the second axis of this thesis. The contaminants were selected based on their mode of action in target insect species and in daphnids, which have a similar endocrine regulation as amphipods and therefore similar modes of action. We also studied the effects of two naturally present hormones in crustaceans and insects. The molecules chosen for this study were the following:

- Juvenile hormone III (JH): regulates development and reproduction in insects
- Methyl Farnesoate (MF): suspected to play a role similar to that of JH but in crustaceans
- Fenoxycarbe, Pyriproxyfen (PYR): analogues of the juvenile hormone
- Tebufenozide (TEB), Methoxyfenozide: analogues of the molting hormone ecdysone

After bibliographic research, the range of concentrations used for all contaminants was 0.05, 0.5, 5, and 50 µg/L. This range comprises the concentrations that exerted reprotoxic effects in the studied species, without significant mortality.

Females were placed in glass beakers of 500 mL inside a thermoregulated bath (10 females per beaker, 2 beakers per condition). Solvent controls (uncontaminated drilled ground water spiked with 0.005% acetone) were used taking into account possible solvent interferences. Stock solutions were prepared in acetone at 1g/L. The contaminated media was obtained by adding 25 µL of stock solutions to 499.975 mL of uncontaminated drilled ground water (solvent at 0.005% in the media). Renewal of the media was performed manually every two days, and temperature was maintained at 17°C throughout the experiment. After fourteen days, the exposure was ended.

Experiment 4

The objective of this experiment was to evaluate the sensitivity of the gene candidates proposed as biomarkers in *G. fossarum.* Based on the protocol proposed by Geffard et al. 2010, female organisms were expose to pyriproxyfen, tebufenozide, and piperonyl butoxide. Exposure strategy and conditions were the same as for experiment 3. The experiment comprised three different exposure times (4, 9, and 14 days) and the following concentrations: pyriproxyfen (PYR 5 µg/L), tebufenozide (TEB 0.5 µg/L) and piperonyl butoxide (PBO 150 µg/L). The exposure times were chosen in order to evaluate the impact of contamination in the maximum expression peaks of the genes analyzed. At the end of the exposure, five females from each condition were individually sampled, rapidly weighed, and incubated In RNA later overnight at 4°C. After incubation, organisms were frozen in liquid nitrogen, and stored at -80°C until further analysis.

1.3.2. In situ exposures

The objective of this experiment was to assess the suitability for field studies of the multiplexed SRM mass spectrometry technique developed in *G. fossarum*. For this, a high-throughput protein biomarker measurement was carried out during a caging deployment operated throughout a monitoring network from the “Rhône Méditerranée

Corse" (RMC) Water Agency. We analyzed caged male and female organisms in 4 reference sites and 13 sites subjected to chemical contamination, previously monitored and prioritized by the Water Agency. This campaign took place between April and June 2015. A detailed description of the field sites is provided in publication nº3.

Caging strategy

We applied a robust, active biomonitoring strategy which is based on caged organisms (Figure II-8), previously developed for *Gammarus fossarum* (Besse et al., 2013; Coulaud et al., 2011; Jubeaux et al., 2012b; Lacaze et al., 2011b). This strategy involved the use of a reference population sampled in a "clean" site (genotyped organisms always with the same size, and therefore similar age). As described in 1.1, organisms were acclimatized for a minimum of one week to laboratory conditions, prior to their caging in the study sites. This acclimatization period (constant temperature, *ad libitum* feeding) aimed to eliminate organisms weakened during harvesting, to ensure a good physiological status of the organisms used for the test whatever the season, and to acclimatize them to laboratory conditions before beginning the exposures.

"Calibrated" organisms were then caged *in situ* according to the protocol of section 1.2.1, and then brought back to the lab for further analysis. Organisms were placed in punctured polypropylene cylinders to allow free circulation of water, and fed *ad libitum* with alder leaves throughout the exposure.

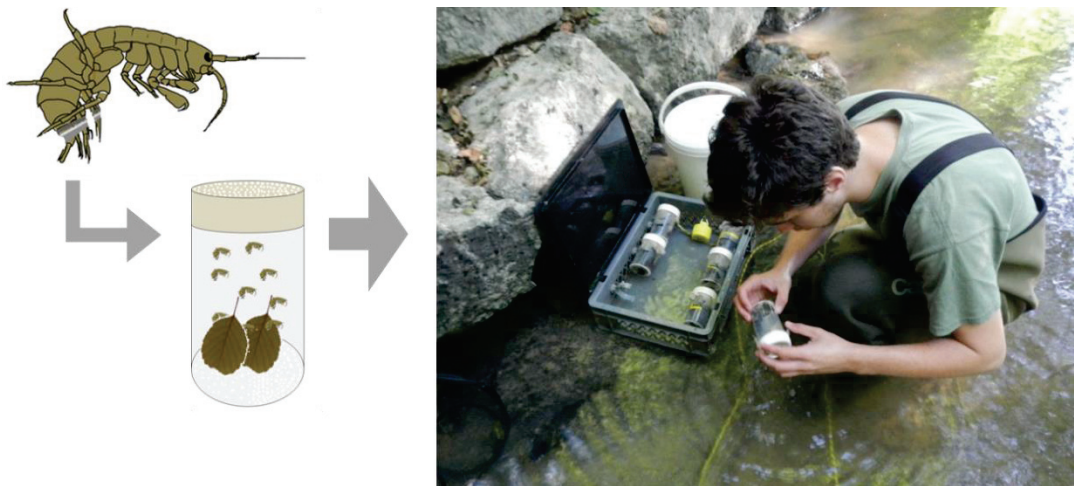


Figure II-8 – Illustration of gammarid *in situ* caging.

Briefly, two polypropylene cylinders containing 10 couples each prepared to be placed in each site. The plastic cylinders were transported in plastic buckets containing the same water as the stables, inside coolers with a bubbling system to oxygenate the water. Once in the exposure site, the cylinders were placed inside pierced PVC boxes (60*40*20 cm), in order to protect them against possible damaging agents brought by the currents (branches, rocks, etc). The boxes were dived into the water and fixated on the margins of the river using a rope. A temperature probe was placed inside each box to record continuously the water temperature. A multimeter probe was used to measure several physical and chemical parameters of water (pH, conductivity and temperature) at the time of installation of the systems, and in the end of exposure. At the end of exposure, organisms were counted (for survival rate assessment), weighed, and directly frozen in liquid nitrogen and stored at -80°C until further analysis.

2. Proteomic analyses

2.1. Targeted proteomic analysis by Selected Reaction Monitoring

2.1.1. Candidate selection based on proteogenomics-derived protein sequences

The theoretical/experimental procedures for the SRM-based assay are represented in the workflow presented in Figure II-9.

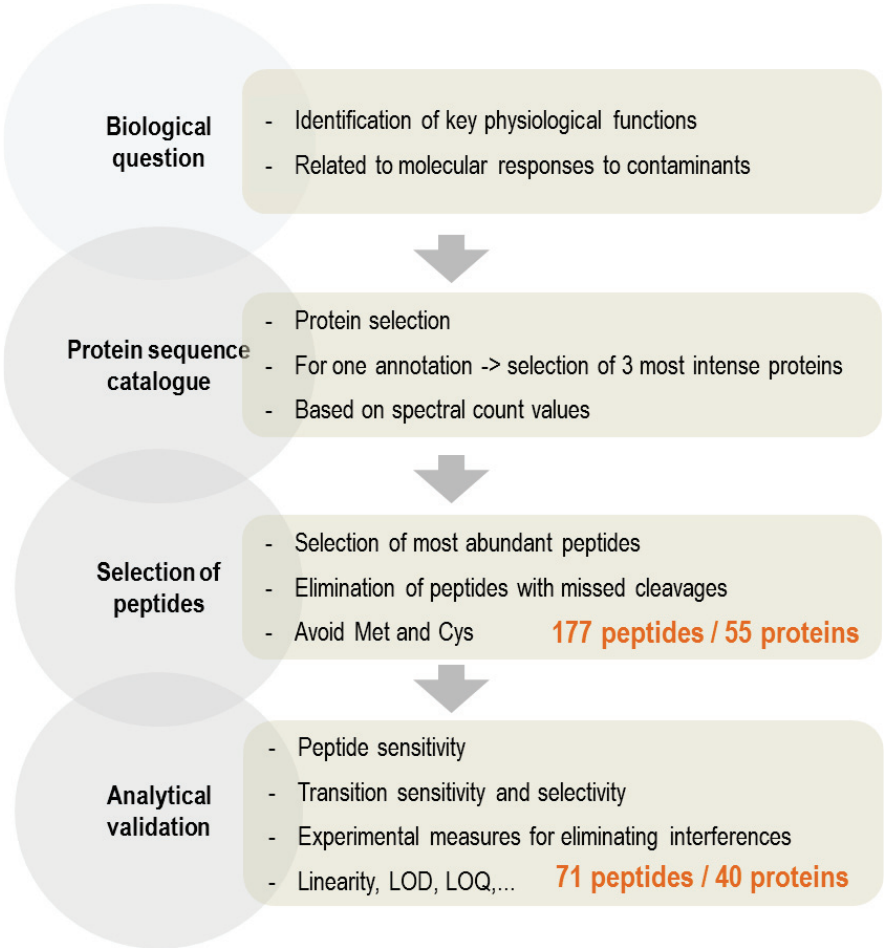


Figure II-9 – Workflow used for the protein/peptide selection, and the following analytical validation. In orange is represented the number of initial proteins/peptides selected, and the final number which can be measured simultaneously after analytical validation.

The selection of target proteins is a hypothesis-driven process, and is based on previous experiments, scientific literature, or prior knowledge. This selection aims at

answering a specific biological question like the analysis of pathways, protein networks, or the evaluation of a set of biomarkers (Gallien et al., 2011). In this work, the selection of proteins and peptides for the SRM experiment was based on sequence information obtained in the previous proteogenomic experiments performed in *G. fossarum* (Trapp et al., 2014b). By mining this protein catalog, we identified proteins whose bioinformatic annotations suggested an implication in key physiological functions: hormonal regulation, molt, immune system, detoxification/antioxidant mechanisms, and osmoregulation. This led to the construction of a subset of candidate proteins listed in Table II-1.

Each protein ID from Table II-1 corresponds to an assembled contig in the GFOSS RNAseq-derived transcriptomic database. Since the *Gammarus* genome is not sequenced, the *de novo* assembled RNAseq GFOSS database is subjected to errors in the assembly. This could lead to the creation of innacurate chimeric sequences of contigs, abusive merging of real duplicated paralog gene sequences, or difficulty to integrate alternative splicing in contig identification. De novo assembled RNASeqs are known to present often a weakness of contig redundancy. Therefore, several protein IDs often present similar functional annotations. For example, 10 proteins are annotated as "chitinase" in the protein catalog obtained from the shotgun proteomics experiments. Functional annotations were performed through cross-species sequence homology searches, i.e., the functions were predicted based on sequence conservation and functional information from various taxonomic groups present in databases, mainly vertebrates. Because of the lack of molecular information available for these species, functional annotations predicted from sequence homology of phylogenetically distant species can sometimes be erroneous.

The shared annotation for several contig IDs can therefore result from several factors: 1/ the existence of distinct paralog genes in the genome of *G. fossarum* ; 2/ isoforms generated by alternative splicing abusively assembled as distinct contig sequences in the RNAseq database; and 3/ from the simple failure to assemble whole length gene sequence (the two contigs are in fact two parts of the same gene). For this proof of concept work, we defined each different protein ID as being indeed different proteins.

Table II-1 – Proteins and proteotypic peptides selected for SRM method development.

Protein ID	Functional annotation	Description	Most abundant proteotypic peptides
153198	pre-amylase	Digestive enzyme. Catalyzes the hydrolysis of starch into sugars.	GFVAINNEGSDMK
			YGAEIGAAMR
4757	α amylase		SGDEAAFTNMVQR
			YLVNFGEGWGFLPSGEALAFIDNHDNQR
166198	α amylase		NCNLVGLTDLYGATDYVR
			LGVAFTLAQPYGYTR
193455	celullase	Digestive enzyme. Catalyzes the decomposition of cellulose into monosaccharides.	SGTIAMFAEVVGGQYVGILQSHVAELR
			AIDLGLDVATNTEWAER
213317	celullase		GDSALSDGSDHGLDLTGGYYDAGDHVK
			ELFDFADAHR
4242	cellulase		DLFEFADQYR
			GDSALGDGDDVGLDLTGGYYDAGDFVK
45375	catalase	Antioxidant enzyme. Catalyzes the decomposition of hydrogen peroxide to water and oxygen.	NLPADQAAALASSDPDYAIR
			YNSSDDDNYTQTGDFYR
			LGSNFLQIPVNCPYR
110912	catalase		LADNIAGHVINTQEFIR
			ADPALGQAIQER
			YNSSDDDNYTQTGDFYR
195622	catalase		LADNIAGHVINTQEFIR
			ADPALGQAIQER
			LGSNFLQIPVNCPYR
			LFAYADTHR
105797	superoxide dismutase	Antioxidant enzyme. Catalyzes the dismutation of the superoxide radical into ordinary molecular oxygen or hydrogen peroxide.	HVGDLGNIQADNAGNAIVNIR
			AIVVHDGVDDLGLGGDHGSR
176048	superoxide dismutase 1, soluble-like, partial		HPGDFGNIEVPTDGR
			EGDCATAGDHFNPHGTVHGGPTDDVR
150268	glutathion S transferase	Detoxification eyzme. Catalyzes the conjugation of the reduced form of glutathione to xenobiotic substrates for detoxification.	SVLLDPTCLSDFSSLQAFQK
			IIGTHNYVAGDEISFADFFLFELLER
40079	glutathion S transferase		DTLGLEFPNLPYYIDGDLK
			FSESSAILR
142711	glutathion S transferase		TSMPGYEQNEEGAQQFGAFAK
			LSAWLAACK
32234	Na/K-ATPase a1 subunit	Osmoregulation-related enzyme. Active pump (uses energy from ATP) that pumps sodium out of the cells while pumping potassium into the cells, both against their concentration gradients.	FVGLISLIDPPR
			SPDFTNDNPLETR
110907	Na+/K+_ATPase		LGAIVAVTGDGVNDSPALK
			VIMVTGDHPITAK
			SVGIISEGNETVEDIANR
			DMTSDQLDDILR
209438	Na+/K+_ATPase		LQTNPDTGLSTAEAR
			SPDFTNDNPLETR
			ALSTIAALCNR
			NLAFFSTNAVEGTAR
			EVNGDASEAALLK
			CCELACGDVLDWR
			GVVINIGDR
134275	farnesoic acid O-methyltransferase	Involved in hormonal metabolism. Catalyses the formation of methyl farnesoate from farnesoic acid in the biosynthetic pathway of juvenile hormone	EVFIGGWSNQNSAIR
			EFWIATDHNEVR
166723	farnesoic acid O-methyltransferase		ETTVIKVPTPDECCGAAK
			DIATEDKLEYR
			EFWIATDHNEVR
191918	juvenile hormone epoxide hydrolase-like protein 5	Involved in hormonal metabolism. Inactivates juvenile hormones through the hydrolysis of the epoxide functional group contained within these hormones.	AAFVDDLLVFPK
			GGHFAAMEEPKLVADDVIK
144144	JHE-like carboxylesterase 1	Involved in hormonal metabolism. Catalyzes the hydrolysis the carboxylic ester bonds of juvenile hormone.	AFWGSLPLR
			ILTTMWADFAR
			AETSEEIVLCLQDVNAHLLDDTMYNLMDWNFQPFK
122081	cytochrome P450 CYP12A2	Thought to be associated with the metabolism of ecdysone.	FNNNLINTR
			TLEELSNEALR
			AYKTLEELSNEALR
100255	cytochrome P450 enzyme, CYP4C39		ILEDFVDVFNR
			VYAEVIEVAGSGPIGLDQLR

212968	chitinase 1 precursor	Essential for molting. Hydrolytic enzymes that break down glycosidic bonds in chitin.	YSQMVSVDTR
			DTDWVGYEDPDSTAIK
			VAIAVGGWAEGGKK
181833	chitinase		LVLGTATYGR
			LDSTANTGMYAPAQQPGSAGQYTR
12415	chitinase		GGQWFGYDDISMIR
			VIDAGYDVPALNR
4315	antimicrobien peptide	Are a part of the innate immune response found among all classes of life, acting as antibiotics.	IFGAGGFGDLGGGGLGGPGVGLGGVGGYGGVSDTCR
			APQLCSNDGACGGIDK
			TDVGQAYCCESGLEEPGLVGTKPGK
7908	hemolectin	Multi-domain protein thought to act in hemostasis and coagulation.	TCQCPAGLLGDLCQFR
			EVICAGSWILDGVQTLSGLSTPLSCK
			TNVVDSGEVLQLSCVNGFWLPDPDHLVR
11145	hemolectin		VECIAGFILPLEFK
			EGSSLVTLILEGDTSCSDGCPR
			GWPLPIGEGQITGCVR
109695	hemolectin		CPVEYGGISCQEKR
			NAGPVLLPSNTSPVLR
17046	clottable protein 2	Proteins experimentally validated as being yolk proteins (vitellogenin-like) in *G. fossarum*	ITMQEDGSGEVQLK
			IATGMQSALPEYAGTGIK
			ILNVMAAADLGAR
			MIFSSYDEPSQQVQSYISWVVDSLSR
			NVEVGEGMAVLPDNGR
			IAQGMMDAMHLLR
			TSEVFLPLTNELYQQTK
			TASIFFNLANNIR
			NQNEPEYSPIFYQR
39606	clotting protein precursor		YLNPVESNIIMPTAPGLFMK
			TLPFTAIAGMMSTLER
			LAVYDELNMDVK
			LMSLLASPTVK
			SQDLIEGYLFPK
			NYDLTIYQLNEPSQTK
			KIQVAVLENNQDSK
			IELNTPTVLFGEVK
			EVGPNPDFETVAR
			IQVAVLENNQDSK
206469	clotting protein precursor		IYTMLDLDMSK
			QLLPIAESHAEDFEIR
			YLEVMAHLNFAK
276	hemolymph clottable protein		IIYPAEALTIVIEK
			NCEVSLLLDNHQSQTK
			HAEFSVNPPLDSTQAVK
			MISVVNSLTAR
			SKIPSVSQTPVQTCITLAMK
			KQGEIQINLLR
			QGEIQINLLR
			TAATVSQSALSMK
194758	Apolipocrustacein		TAQLYSELIQSMDQVEER
			LAGEIYQEMASQTQLQSIK
			EGCLQEEPVQQIMQTISTK
			KTSLLYDYTDSPK
277	vitellogenin		LVLADVDVDVAVGR
			TYHVTSEMEFAK
			KTPGSEGASQYVSQVQLSYNEK
200426	vitellogenin		TLGALELDVFSEAGK
			WEGDLQGNAEGGVEK
			FEAIGVNQHEPALSAR
			IELGGVVSSVNNYDCQGYK
			QAAHAGVLSVLLQTPESK
			QTLAPTLSLAER
64	Vitellogenin		HIEIFSPITK
			NLHFINQISMAK
			VSRPTLAYVQLK
			YLDLASEDESLFFHLHILK

2308	copine-8	Probable calcium-dependent phospholipid-binding protein that may play a role in calcium-mediated intracellular processes. Previously defined as male-specific protein in *G. fossarum.*	EVLEELPAQYMEFTR
			SDPICVVGLSELEGGEVK
18473	copine-8		EVLEELPAQYMEFTR
			VAERDIVQFVELR
34845	copine-8		AVAEIVQDYDSDGFFPALGFGGK
			SDPICVVGLSELEGGEVK
			DLIGSVGLSVGQLMNAGGK
			QGPCEANTQSLVNPK
			QGTDNEYMTAIR
5059	Flotillin-1	Membrane protein, undetermined function. Previously defined as male-specific protein in *G. fossarum*.	RAEAELASELQEAK
			LTGEVMDIVSR
			MTGIDVAKPVR
4227	calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type-like isoform 1	This magnesium-dependent enzyme catalyzes the hydrolysis of ATP coupled with the translocation of calcium from the cytosol to the sarcoplasmic reticulum lumen. Previously defined as male-specific protein in *G. fossarum*.	SASEMVLADDNFASIVAAVEEGR
			LLEFEITGSTYEPIGDVFLGGQR
			EFTLEFSR
14545	epididymal sperm-binding protein 1	Binds to spermatozoa upon ejaculation and may play a role in sperm capacitation. Previously defined as male-specific protein in *G. fossarum*.	VLSVVQNIITR
			RGDCPLLATQLTDSAK
186168	epididymal sperm-binding protein 1		VLSVVQNIITR
			RGDCPLLATQLTDSAK
18609	epididymal sperm-binding protein 1		RGDCPLLATQLTDSAK
			VLSVVQNIITR
2562	prophenoloxidase	Enzyme involved in the proPO-activating system. Precursor of phnoloxidase, responsible for melanization of pathogens and damaged tissues. Previously defined as male-specific protein in *G. fossarum.*	ATQPSYTVAQLELPGVNITR
			APILEGYFSK
			FNQNQTSTQMPLNEQR
			GIDFGTTQSVR
			GIDIIGDAFEADR
15561	prophenoloxidase		FGDTTPNVINVETNTALASALPNLVK
			DVMYLFERPSDPLFLPR
			IVIDLLQQSTTVAQLR
6945	prophenoloxidase		TPNQQVLTIVDLATLENMAVHDITIR
			GIDFGTTQSVR
1917	transglutaminase	Enzyme that participates in blood coagulation, skin formation, and signal transduction. Catalyzes the isopeptide bond formation between Lys and Gln residues to form ε-(γ-glutamyl) lysine bonds between appropriate substrates. Previously defined as male-specific protein in *G. fossarum.*	GRPWAFGQFDDVVLPVAAYVLEMK
			FDQIHDADSPKPVLR
			DLVMQVYIPASAPVGIWR
			GTLAVIPVQNR
2415	transglutaminase		FAMQGGGLFQGGGQYGSQLGR
			AADPSQNPTWLQR
			VVLPVNPGAEFQR
7169	transglutaminase		RGEVSLQYDVPFVLAEVNADLVR
			GTQEVLEVLGKR
			VLAVDILAK
461	hemocytin-like	Ortholog of von Willebrand factor and is expected to be a major mediator of hemocyte aggregation. Previously defined as male-specific protein in *G. fossarum.*	TTLLCSDGIWYLALPK
			SSVFDGFDAPQLWGR
			SCQLPFSYGGR
			CQGSFWSYPDCDTNDPPVCRPVCR
			NGGVCIKPNECSCPENFSGPR
1975	hemocytin-like		GLCLPTDACPR
			GPQCDEPCEEDPIGCR
			LLLAAELDCLPLCPMGCR
160381	hemocytin-like		GPQCDEPCEEDPIGCR
			CEEGYALQSGVQSVR
			SGDTSFSPCSPVCDPLCVNR

The choice of the protein ID for each annotation was based on their spectral count intensity in the shotgun experiments. Hence, from the ten protein sequences annotated as “chitinase” (Table II-2), the three more intense were chosen for the SRM experiment. In the case where the spectral count value was similar between proteins, the one that was identified with a higher number of peptides was chosen (like protein 212968).

Table II-2 – The ten protein IDs annotated as having chitinase function, and their respective spectral count values from previous shotgun experiments.

GFOSS contig_ID	Mass	Best Match			Total non-redundant peptides	Total Spectral Count
		NBCI_ID	Probable function	Blast Evalue		
181833	**13303**	**AFC60660**	**chitinase**	**8,09E-24**	**2**	**21**
12415	**50796**	**gb|ADG22164.1|**	**chitinase 2**	**6,00E-154**	**2**	**7**
42149	41659	AFC60662	chitinase	3,77E-35	2	7
183925	48770	AFH58810	chitinase, partial	1,24E-27	2	7
212968	**40290**	**ACG60512**	**chitinase 1 precursor**	**3,43E-11**	**3**	**7**
202609	32281	dbj|BAA12287.1|	chitinase	3,00E-32	3	4
145	177776	ref|NP_001036035.1|	chitinase 7 precursor	0,00E+00	2	3
216470	89950	gb|ACG60513.1|	chitinase 1 precursor	1,00E-33	2	3
105035	29544	AFC60662	chitinase	6,00E-21	2	2
144157	41994	gb|ACR23314.1|	chitinase 4 precursor	1,00E-139	2	2

For each protein included in an SRM analysis, a set of proteotypic peptides (*i.e.,* unique peptides) was selected. Desirable peptides must be unique to the protein and easily detectable by mass spectrometry. For proteins identified with more than three peptides, peptide selection was also based on their spectral count values obtained in the shotgun proteomics experiment. This value gives an important insight into the high- or low-abundant peptides in the organism. Moreover, peptides detected in shotgun experiments will be easily detected in the posterior SRM analysis. Indeed, we focused on the best peptides observed during discovery phase. A set of two to three peptides per protein were selected, resulting in a list of 177 peptides (Table II-1). These signature peptides were synthesized with heavy labelled arginine or lysine [$^{13}C_6$, $^{15}N_2$] at the C-terminus of the peptides to serve as internal standards for absolute quantification in the SRM assay.

2.1.2. Analytical validation of the methodology

The results obtained from the validation procedure will inform us about the quality, reliability, and consistency of the method. The most common validation criteria used for method development and validation are its specificity/selectivity, linearity, precision, limit of detection, and limit of quantification.

Standards and control quality preparation

Stock standard peptide solutions of isotopically-labeled peptides were prepared by dissolving accurately weighed standard compounds in an H2O/ACN (50/50, v/v) mixture containing 0.1% of formic acid to yield a concentration of 20 mg/mL. Solutions were prepared from the stock solutions and diluted further with either an H2O/ACN (90/10, v/v) mixture containing 0.1% of formic acid or the matrix (pool of extracted and crushed G. fossarum) to prepare several solutions to reach the desired concentrations of 2500, 1000, 500, 400, 250, 100, 50, 10, and 5 and 5 ng/mL, which was used to build the calibration curves. Quality controls were prepared from the stock solutions to obtain concentrations of 150 ng/mL, 750 ng/mL, and 1500 ng/mL.

LOD, LOQ, linearity, precision, and accuracy

For the determination of the limit of detection (LOD), limit of quantification (LOQ), and linearity of the assays, a standard curve from area was produced, based on nine samples containing equal amounts of *G. fossarum* protein extract digests as background matrix. Each sample was spiked with an increasing amount of labeled peptides from 5 to 2500 ng/mL, spanning a 500-fold range. For linearity, each standard sample was analyzed three times. Determination of the signal-to-noise ratio was performed by comparing measured signals from samples with known low concentrations of labeled peptide with those of blank samples and by establishing the minimum concentration at which the peptide can be reliably detected. A signal-to-noise ratio between 3:1 and 10:1 was considered acceptable for estimating the LOD and LOQ, respectively.

Precision and accuracy of the method were evaluated by performing three runs on three separate days. Each run consisted of one set of calibration standards, three (intrabatch) or nine (inter-batch) replicates of each QC concentration (150, 750, and 1500 ng/mL). Least square linear regression with a weighting factor of 1/x2 was used to plot the curve of the calibration standards. Criteria of precision determined at each concentration level as the relative standard deviation or coefficient of variation should not exceed 20% and the average accuracy in determining the concentration must be

within 80–120% of the expected value. Calibration standards and the final calibration line contain at least five calibration concentrations.

2.1.3. Protein extraction and digestion

After selection of the peptides/transitions (section 2.1.2.) and method validation, the final experimental procedure followed the scheme in Figure II-10. Total protein extraction was done using the whole organism. Since no laborious dissection is needed, this comprises a major advantage in terms of time gain. After extraction, proteins were digested to obtain smaller peptides. Since mass spectrometry was originally directed at the analysis of small molecules, its implementation in protein analysis implicated protein fragmentation into peptides (nowadays there are also methods for analyzing entire proteins). This step is commonly performed using trypsin, an enzyme with a very well defined specificity, since it hydrolyzes peptide bonds in which the carbonyl group belongs to either an arginine or a lysine residue. A digestion protocol normally involves reduction of disulfide bridges using a reducing agent such as dithiothreitol (DTT), thus linearizing the protein. After reduction, an akyl group is added to the reduced cysteines to prevent the reformation of the bonds. Then the tryptic digestion is performed. Protein digestion was performed in solution, and the heavy-labelled peptides added to the solution after peptide purification by solid phase extraction. The AQUA (absolute quantification using heavy-labeled peptides) strategy was used for quantification. Heavy labelled peptides were synthesized with incorporated stable isotopes as internal standards to mimic native peptides formed by proteolysis. AQUA internal standard peptides are then used to quantify the absolute levels of proteins after proteolysis by SRM. The details concerning the LC-MS/MS analysis are presented in section 2.1.4.

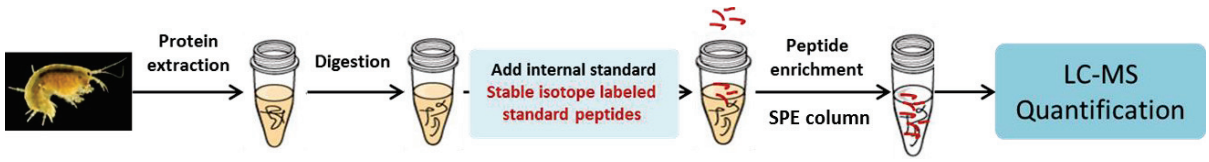


Figure II-10 – Main stages of sample preparation for SRM analysis. SPE – *solid phase extraction*.

The procedure used for total protein extraction was adapted from Simon et al. (Simon et al., 2010). Whole-body organisms were homogenized in Tris Buffer (Tris 50mM, 100mM NaCl, 0.01 mM EDTA, 0.1% v/v Triton X-100) buffered at pH 7.8 and complemented with 10 µg/L of each leupeptin and aprotinin, with a bead mill homogenizer. For each organism, the volume of buffer was adapted according to their weight (25 µL buffer/mg male; 50 µL buffer/mg female). The homogenates were centrifuged at 10,000 g at 4 °C for 15 min, and 250 µL of the clear resulting supernatant were collected to new tubes. A volume of 750 µL of an ethanol/diethyl ether delipidating solution (1:1 v/v) was added to the supernatant, and the mixture was vortexed and incubated on ice for 10 min. After a 10 min centrifugation at 10,000 g, the resulting supernatants were removed, and the pellets resuspended in 250 µL Tris Buffer.

The protein extracts were then treated with 3 mL of ammonium bicarbonate (AMBIC) 50mM and 362 µL of DTT 150 mM for 40 min at 60 °C. After cooling to room temperature, samples were incubated in the dark with iodoacetamide (final concentration of 15 mM), during 40 min at room temperature. A volume of 150 µL of a 2 mg/ml TPCK treated trypsin solution prepared in 50 mM of AMBIC was added, and the samples were incubated for 60 min at 37 °C. A volume of 20µL of formic acid (FA) was added to stop trypsin reaction. A volume of 10 µL of the solution of heavy peptides at 1 µg/ml was added to each digest. Each mixture was then purified with an Oasis HLB 3 cc solid phase extraction columns (60 mg). After peptide elution with 1 mL of methanol/0.5 % FA, 100 µL of a glycerol 10%/methanol were added, and the samples evaporated under a flow of nitrogen. The pellet was then resuspended in 90 µL of a H_2O/ACN (90/10) + 0.1% FA solution.

2.1.4. LC-MS/MS analysis

LC–MS/MS analysis was performed on an HP1200 series HPLC device (Agilent Technologies, Waldbronn, Germany) coupled either to a QTRAP® 4000 or a QTRAP® 5500 LC/MS/MS System hybrid triple quadrupole/linear ion trap mass spectrometer (Applied Biosystems/ MDS Analytical Technologies, Foster City, CA, USA) equipped with a Turbo VTM ion source. The LC separation of the 20µL injected sample was carried out

on an Xbridge C_{18} column (100mm×2.1mm, particle size 3.5 µm) from Waters (Milford, MA, USA). Elution was performed at a flow rate of 300µL/min with water containing 0.1% (v/v) formic acid as eluent A and acetonitrile containing 0.1% (v/v) formic acid as eluent B, employing a linear gradient from 2% B to 33% B in 19 min, followed by a second linear gradient from 33% B to 64% B in 6 min. Then, column washing and re-equilibration was performed for 6 min. The injection duty cycle was 35 min, taking into account the column equilibration time. Instrument control, data acquisition, and processing were performed using the Analyst 1.5 software. The mass spectrometer was initially tuned and calibrated using polypropylene glycol, reserpine, and Agilent Tuning Mix (all Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. MS analysis was carried out in positive ionization mode using an ion spray voltage of 5500 V. The nebulizer and the curtain gas flows were set at 50 psi using nitrogen. The Turbo V™ ion source was set at 550 ◦C with the auxiliary gas flow (nitrogen) set at 40 psi. The SRM transitions were monitored and acquired at unit resolution in Q1 and Q3, with a dwell time of 10 ms used for each transition, to obtain 10 data points per chromatographic peak minimum.

2.1.5. Data exploitation and absolute quantification

WIFF files containing all the MS/MS spectra obtained from the SRM analysis were exported from the Analyst® software. WIFF files were then imported into a previously prepared document in the Skyline v3.1 software (MacCoss Lab Software, USA). This document contained all the information about the analytical run, peptides, and transitions that were analyzed in the mass spectrometer, such as peptide modifications (cysteine carbamethylation, isotope modifications), precursor and product masses, instrument used, ion types and charges, ion match tolerance, method match tolerance. A skyline document has the outline represented in Figure II-11.

In order to perform a quality control of the large number of SRM runs performed, spectra quality, and reproducibility between samples were evaluated using Skyline. This analysis comprised manual peak picking for reliable peak identification, verification of retention time between samples, and transition refinement regarding quantification: check if the three transitions selected for each peptide were of good quality and

reproducible over the samples (constant transition intensity and contribution to the total area of the peak in all runs).

Absolute quantification was performed using the ratio between the peak areas of peptides in the sample and the corresponding heavy labeled peptide (with known concentration of 1000ng/ml, 10 ng added to the sample). Peptide areas were acquired by integrating the peptide peaks in Skyline v3.1 software (MacCoss Lab Software, USA). Peptide quantities were normalized to their corresponding molecular weight (pmol), and to the weight of the organism (pmol/mg organism).

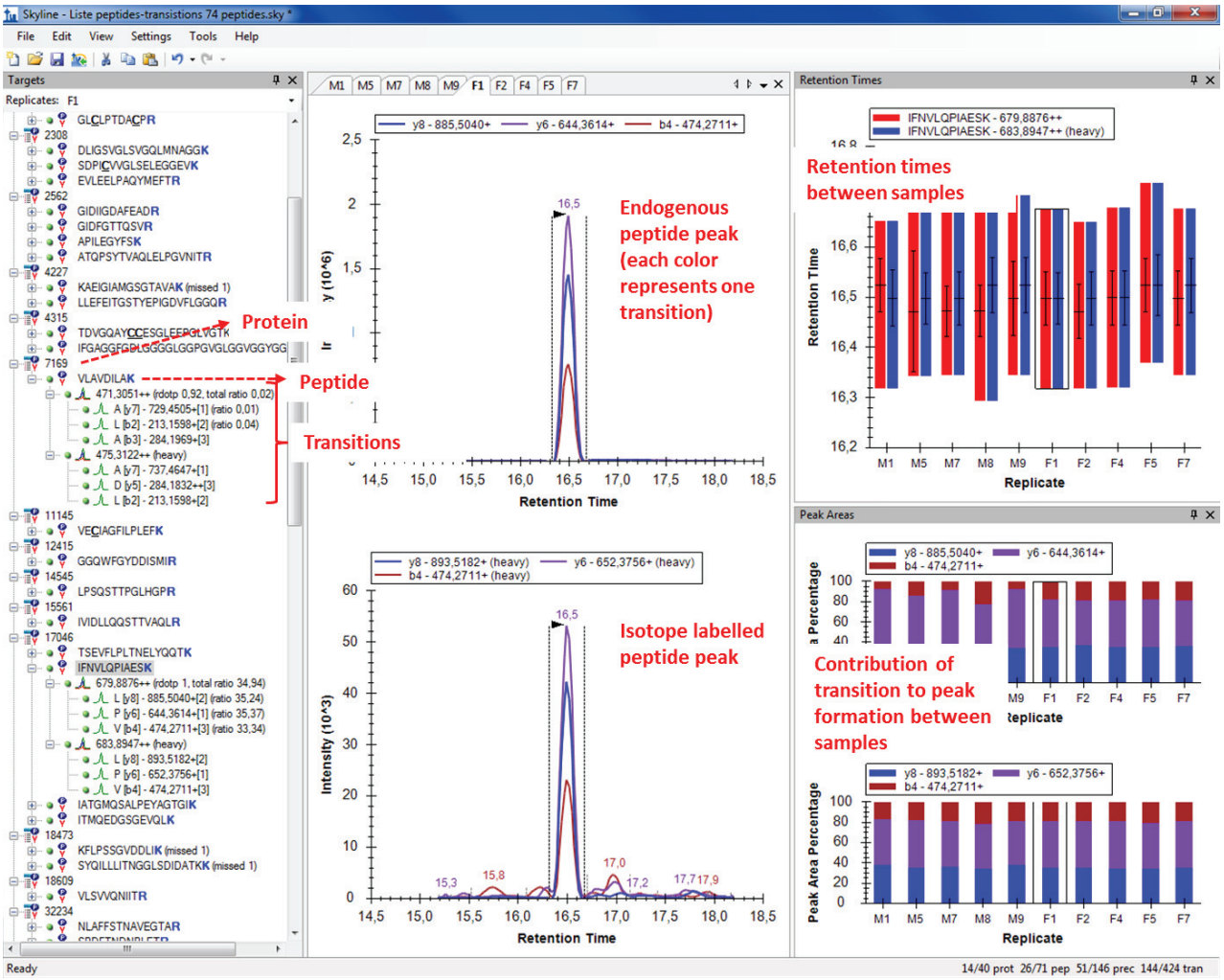


Figure II-11 – Overview of skyline window for analysis of SRM MS/MS peaks.

2.2. Shotgun proteomics of male gonads

2.2.1. Protein extraction and digestion

Each pair of testis was directly dissolved in 40 µL of lithium dodecyl sulfate LDS sample buffer (Invitrogen). After sonication for 1 min (transonic 780H sonicator), samples were boiled for 5 min at 95 °C. A volume of 35 µL of each sample was then subjected to SDS-PAGE for a short electrophoretic migration as previously described (Trapp et al., 2015). The whole protein content from each well was extracted as a single polyacrylamide band, processed, and proteolyzed with trypsin (Roche) using 0.01% ProteaseMAX surfactant (Promega) as described (Hartmann, Allain et al. 2014).

2.2.2. LC-MS/MS analysis

The peptide mixtures were analyzed in data-dependent mode with two tandem mass spectrometers. The parameters used with the LTQ Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings) have been previously described (Trapp, Almunia et al. 2015). For the analysis with the Q Exactive HF mass spectrometer (Thermofisher) operated also coupled to an UltiMate 3000 LC system (Dionex-LC Packings), the parameters were as follows. Peptides were first desalted on line on a reverse phase precolumn C18 PepMap 100 column. Then, peptides were resolved with a 90 min gradient of CH_3CN, 0.1% formic acid, at a flow rate of 0.2 µl. They were analyzed with a data-dependent Top15 method consisting in a scan cycle initiated with a full scan of peptide ions in the Orbitrap analyzer, followed by high-energy collisional dissociation and MS/MS scans on the 15 most abundant precursor ions after. Full scan mass spectra were acquired from *m/z* 350 to 1800 with an Automatic Gain Control (AGC) Target set at $3x10^6$ ions and a resolution of 60,000. MS/MS scan was initiated when the ACG target reached 10^5 ions with a threshold intensity of 83,000 and potential charge states of 2^+ and 3^+ after ion selection performed with a dynamic exclusion of 10 sec. The sample analysis with the LTQ Orbitrap XL used the following tune method including the full scan spectra range from m/z 350 to 1800, the AGC target value set at $5x10^5$, and the mass resolving power set at

30000. The MS2 were analysed with a minimum signal required set at 15000 and a Top5 method.

2.2.3. Protein identification and relative quantification by spectral counting

MS/MS spectra were assigned to peptide sequences with the Mascot Daemon 2.3.2 search engine (Matrix Science) against the customized RNA-seq-derived database GFOSS described in (Trapp, Geffard et al. 2014). This database contains 1,311,444 putative protein sequences totalling 289,084,257 amino acids. The following parameters were used for MS/MS spectra assignation: full-trypsin specificity, maximum of two missed cleavages, mass tolerances of 5 ppm on the parent ion and 0.5 Da on the MS/MS analysed with the LTQ Obitrap XL and 0.02 Da on the MS/MS analysed with the Q Exactive HF, static modification of carboxyamidomethylated cysteine (+57.0215), and oxidized methionine (+15.9949) and deamidation of asparagine and glutamine as dynamic modifications. All peptide matches with a MASCOT peptide score below a *p* value of 0.05 were filtered and assigned to a protein according to the principle of parsimony. A protein was validated when at least two different peptide sequences were detected. The false discovery rate for protein identification was estimated by employing the decoy search option of MASCOT (Matrix Science) to be <0.1%. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaino, Cote et al. 2013) via the PRIDE partner repository with the dataset identifier PXD002XY?XX.

Three protein datasets were recorded: Control with solvent (TS); pyriproxyfen treatment at 0.5 µg/L; and pyriproxyfen treatment at 50 µg/L. Each of them was carried out in quintuplets as separate biological replicates. The Tfold module of the PatternLab program for proteomics version 4.0.0.59 (Carvalho, Lima et al. 2016) was used for normalizing and comparing proteomics datasets based on spectral count, as previously described (Hartmann & Armengaud, 2013). Based on the mean spectral count calculation, selection of proteins included in the analysis was based on the L-stringency parameter set at 0.6 and taking into account proteins detected in at least three of the different biological replicates. The F-stringency parameter was optimized as recommended (Carvalho, Lima et al. 2016). To define the set of proteins whose expression was modulated, the fold change was calculated by comparing spectral counts

of each protein dataset. Proteins satisfying both the fold change and statistical criteria were considered as potential protein signatures of the pyriproxyfen treatments.

Functional annotation was carried out with the open source program DIAMOND (Buchfink, Xie et al. 2015) targeting the database SWISSPROT downloaded on 2016/09/30 to extract proteins homologs. Only alignments with an expected value of ≤0.001 were considered and alignments with a bit score <40 were not reported. The GO annotation was then performed from the data provided by the GO Consortium at the Gene Ontology (GO) website with the web application Amigo. Two levels of GO annotation were extracted, level 1 and level 2.

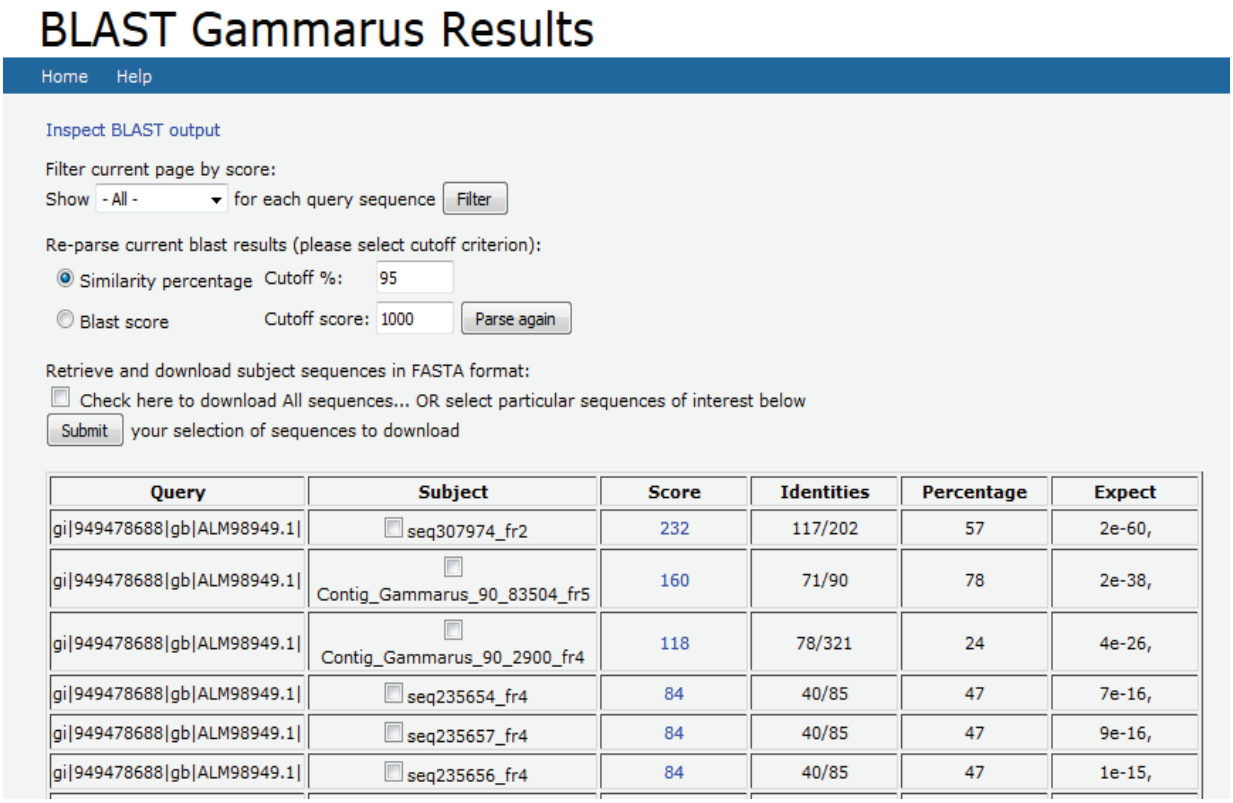
3. Genetic analyses

3.1. Sequence alignments and phylogenetic analyses

The bioinformatic annotations from the three ecdysone-responsive genes chosen as ED biomarker candidates were verified via sequence alignments and phylogenetic analysis. For each candidate, a BLASTP (Altschul et al., 1997) was firstly performed against our RNAseq-derived transcriptome database GFOSS, using protein sequences from the closest species available. The BLAST search was executed against both assembled contigs and unassembled reads present in GFOSS. The top three blast scores with Evalues inferior to E^{-10} and alignment length superior to 25% were chosen for the following phylogenetic analysis. An example of a BLASTP using an RXR sequence as query is represented in Figure II-12. Query sequences used for the blast search were obtained from the NCBI database. For instance, sequences from the recently genome-sequenced amphipod *Hyallela azteca* were available for EcR and BR and therefore used for blast. The RXR query sequence belonged to a decapod crab *Eriocheir sinensis*. For this gene, decapods were the closest order to have sequenced RXR sequences in NCBI protein database.

In order to select *G. fossarum* ortholog genes, phylogenetic analyses were conducted to validate the annotations. For each candidate, a homologous sequence dataset was constructed using ortholog and paralog sequences from the same multigenic family as the candidate gene. The deduced *G. fossarum* sequences were aligned with

other known sequences from a diversity of arthropod species, obtained from GenBank (National Center for Biotechnology Information – NCBI). Multiple sequence alignments and phylogenetic trees were performed in the SeaView software version 4.6.1 (Gouy et al.). Alignments were achieved using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004) program using default parameters. Phylogenetic tree model was determined using the IQ-TREE web server model selection (Trifinopoulos et al., 2016). Trees were built in SeaView, using the PhyML program based on the maximum-likelihood principle (Guindon et al., 2010), with the LG model with 4 substitution rate categories to estimate the gamma parameter shape, and 100 bootstrap replicates for branch support. Default settings were used for all other parameters.



Query	Subject	Score	Identities	Percentage	Expect
gi|949478688|gb|ALM98949.1|	seq307974_fr2	232	117/202	57	2e-60,
gi|949478688|gb|ALM98949.1|	Contig_Gammarus_90_83504_fr5	160	71/90	78	2e-38,
gi|949478688|gb|ALM98949.1|	Contig_Gammarus_90_2900_fr4	118	78/321	24	4e-26,
gi|949478688|gb|ALM98949.1|	seq235654_fr4	84	40/85	47	7e-16,
gi|949478688|gb|ALM98949.1|	seq235657_fr4	84	40/85	47	9e-16,
gi|949478688|gb|ALM98949.1|	seq235656_fr4	84	40/85	47	1e-15,

Figure II-12 – Example of the results obtained from a BLASTP search against the GFOSS theoretical protein database.
Assembled contigs have the nomenclature "Contig_Gammarus_90_proteinID_frame", and unassembled reads "seqProteinID_frame".

3.2. Protocols for gene expression

3.2.1. Total RNA extraction and cDNA synthesis

Whole-body organisms were disrupted with a TissueRuptor (Qiagen). For total RNA extraction, the RNeasy® Fibrous Tissue Mini Kit was used according to the manufacturer's instructions (Qiagen) (Figure II-13). DNase I was used during the protocol to remove possible genomic contaminations. Briefly, organisms were homogenized in the presence of 300 µL of buffer RLT. 590 µL of water plus 10 µL of proteinase K were then added to the homogenate and mixed. After a 10-minute incubation at 55°C, the solution was centrifuged for 3 minute at 10,000g. The supernatant was transferred to a new tube (~900 µL), 450 µL of absolute ethanol were added, and the solution well mixed. 700 µL of sample were then transferred to the RNeasy Mini column and centrifuged 15s at 8,000g (this step was repeated until the complete lysate was used). The flow-through is always discarded until the final step. 350 µL of buffer RW1 were added to the column and the tubes centrifuged for 15s at 8,000g. 10 µL of DNase stock solution is added to the RNeasy membrane and the tube incubated for 15 minutes at ambient temperature. After incubation, the membrane is washed with 350 µL of RW1, 500 µL of buffer RPE, and again 500 µL of buffer RPE (with the centrifugation between the washes). To elute the RNA from the membrane, a 1-minute centrifugation at 8,000g using 50 µL of RNA-free water was performed.

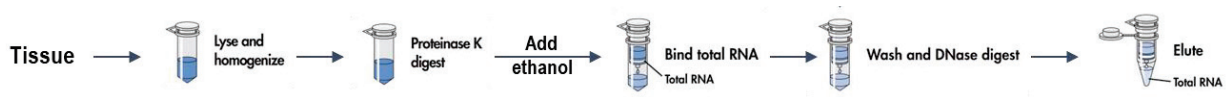


Figure II-13 – Outline of the RNA extraction procedure using the RNeasy® Fibrous Tissue Mini Kit from QIAGEN. (Source: RNeasy® Fibrous Tissue Mini Kit HANDBOOK)

Total RNA was subsequently quantified using a NanoDrop 2000 spectrophotometer (Thermo Scientific), followed by a qualitative analysis using an Agilent Bioanalyzer 2100 RNA Nanochip (Agilent technologies, Santa Clara, CA). 1 µg of RNA from each sample was converted into cDNA using the SuperScript® III First-Strand Synthesis System (Thermo Fisher Scientific) following the manufacturer's protocol (Figure II-14), and conserved at -20 °C until further analysis.

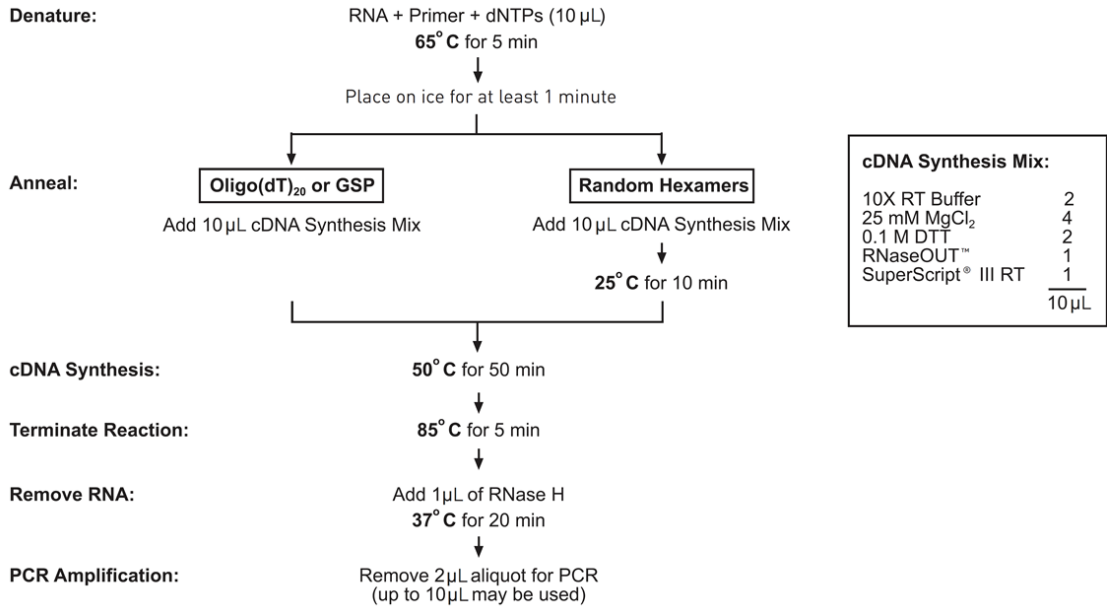


Figure II-14 – Protocol used for cDNA synthesis using the SuperScript® III First-Strand Synthesis System from Thermo Fisher Scientific. In this work we followed the protocol with oligo(dT).

3.2.2. PCR amplification, cloning and sequencing

After phylogenetic validation, we started developing the experimental procedures for sequencing the target genes specific of *G. fossarum.* The objectives were 1/ to obtain reliable nucleotide sequences, since the sequences obtained from the GFOSS database can have potential assembly errors; and 2/ to have functional primers for the subsequent gene expression studies. The experimental strategy is presented in Figure II-15.

Primers were designed manually, based on sequence alignments of *G. fossarum* deducted sequences with other arthropod sequences of the same gene. Four primers were designed for each target gene: two forward primers plus two reverse primers. For the nuclear receptors RXR and E75, primers were designed in the conserved DBD and LBD domains. For BR, since only a small conserved zone was available, primers were designed in that zone in order to obtain the longest sequences possible. For optimal specificity and annealing temperatures, the final primers had between 19 – 21 nucleotides long.

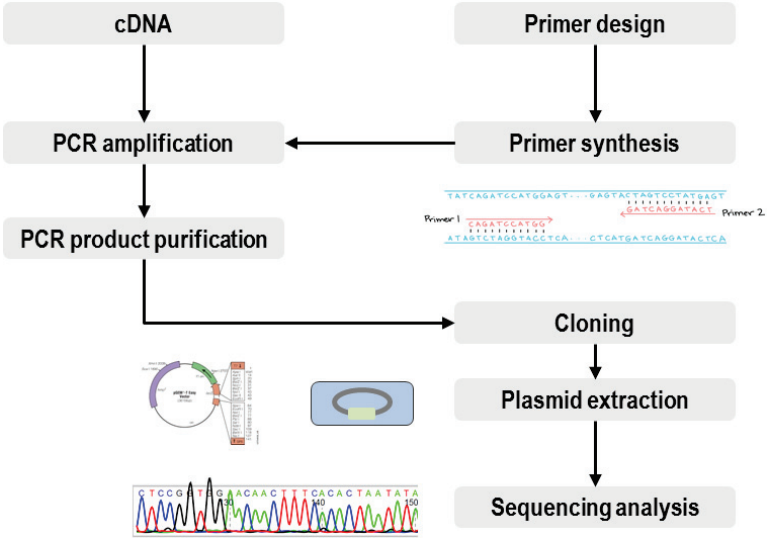


Figure II-15 – Schematic representation of the strategy used for gene sequencing.

Primers were subjected to a quality evaluation using two online softwares: PCR Primer Stats (http://www.bioinformatics.org/sms2/pcr_primer_stats.html), and NetPrimer (Premier Biosoft). These softwares provide information about important parameters such as melting and annealing temperatures of primers, GC content, thermodynamic properties and stability of primer secondary structures, and single or dinucleotide repetitions within the sequences. The GC content of a primer sequence will dictate its melting temperature. In general, primers must have between 40-60% GC in the sequence, and melting temperatures (T_m) around 55-65°C. The species-specificity of the primers was verified through NCBI PrimerBlast. We performed a PCR with a temperature gradient between 56 and 62 °C to determine the optimal annealing temperature (Figure II-16).

PCR was performed with a 50 µL mix of Millipore water containing 1 µL of cDNA, 2 µL of 10mM dNTP, 2 µL of each primer at 10 µM, 5 µL of RT buffer, 3 µL of 25mM $MgCl_2$, and 0.5 µL of native Taq DNA polymerase (Thermo Fisher Scientific) (200U/µL). The following PCR program was used: initial denaturation step at 94°C for 3 min, followed by 39 amplification cycles (denaturation 30s at 94°C; annealing 30s at 56, 58.3, 59.7, and 62°C; elongation 2min at 72°C). A final extension step was made with 10 min incubation at 72°C. The PCR products were run on a 1% agarose gel.

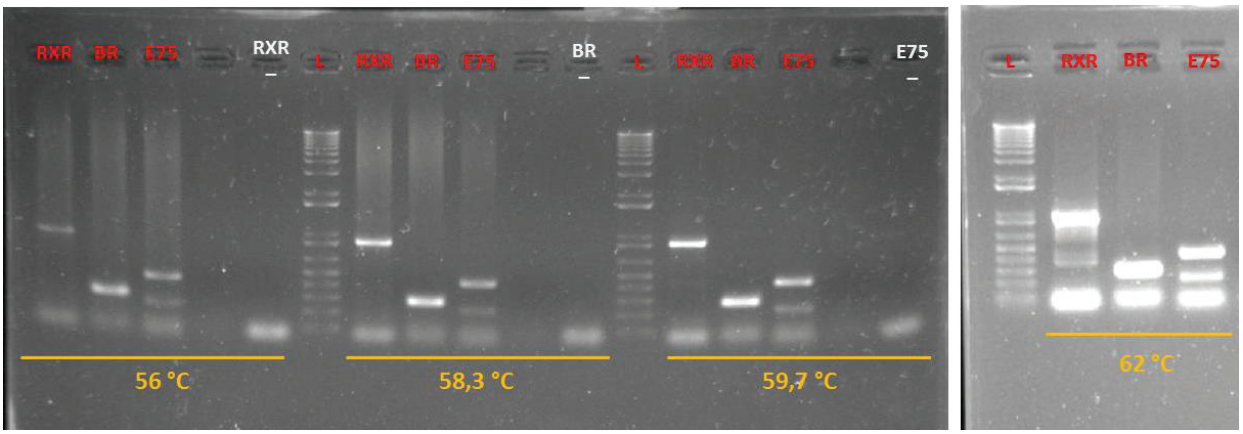


Figure II-16 – Gel image of the amplicons obtained from the temperature gradient PCR for the three target genes. L corresponds to a DNA ladder sizes, and gene name with a "-" means negative control (without cDNA template).

The run at 59.7°C presented intense bands and no smear in the gel. Therefore, the bands from this PCR were selected for the subsequent steps for the purification, and 59.7°C was selected as the optimal temperature for the PCR runs using these primers. PCR amplicons were purified using either the QIAquick®PCR Purification Kit (Qiagen), or the QIAquick® Gel Extraction Kit (Qiagen). Purified fragments were inserted into pGEM®-T Vectors (pGEM®-T Vector System I, Promega), and transformed into DH5α competent cells. Clones containing target fragments were isolated and submitted for PCR verification using the following program: initial denaturation step at 94°C for 10 min, followed by 40 amplification cycles (denaturation 30s at 94°C; annealing 30 s at 55°C; elongation 1 min at 72°C). A final extension step was made with 5 min incubation at 72°C. The PCR products were run on a 1% agarose gel. To obtain high-purity plasmids, positive clones were purified with the QIAprep®Spin Miniprep Kit (Qiagen) following the manufacturer's protocol. Samples were sent to Beckman Coulter Genomics (GENEWIZ®, UK) for sequencing.

3.2.3. Quantitative real-time PCR

Primers for reference genes were chosen based on previous studies that also performed qPCR studies in *Gammarus* species, and are listed in Table II-3. Three reference genes were chosen: Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), 18S RNA, and elongation factor 1 α (EF).

Table II-3 – Nucleotide sequences of the primers retained for gene expression studies.

	Target genes	5' -> 3' nucleotide sequence	Fragment (bp)
Target genes	RXR FW1	AAACCTTTGTGCCATCTGTG	220
	RXR RV3	GAACAGCCTCCCTCTTCATG	
	BRC FW1	GGGAGCAGACCAACAGTTCT	278
	BRC RV2	TGACCTCGCCATGGTACATA	
	E75 FW1	TCGTCAATGCAGCAAGAATC	419
	E75 RV2	GGATGTTCTTGGCGAAGGT	
Reference genes	EF FW	TTCAAGTATGCCTGGGTGCT	82
	EF RV	CGAACTTCCAGAGAGCAATGTC	
	GADPH FW	CGCTGGCCAGAACATCATTC	?
	GADPH RV	CGGCCTTGATGTCGTCGTAA	
	18S FW	TGGGGGAGGTAGTGACGAAATC	?
	18S RV	CCTGCGCTCGATACAGACATTC	

In order to select a suitable reference gene, the stability of these three genes was tested in samples from the two experiments (cDNA from females from each reproductive stage and from pesticide exposure experiment). In order to select the more suitable reference gene, their stability was tested on the cDNAs from the two experiments (female reproductive cycle and contamination). EF was considered as the most stable, presenting the lowest standard deviation for all Ct values obtained in the different samples (Ct standard deviation of 0.42 and 0.33 for each experiment). The EF gene was therefore chosen to normalize expression data in both experiments.

For the quantitative PCR reactions, the iTaq™ Universal SYBR® Green Supermix (Biorad) kit was used. For the reactions, 25 ng of template cDNA and 300 mM of each primer were used, and the procedure was according the manufacturer's instructions. PCR was performed in a CFX96 Touch™ Real-Time PCR Detection System (Biorad), using the following program: 30 sec at 95°C, 40 cycles of 95°C for 5 s, 60°C for 30 s then a temperature increment was programmed for the melting curve (65–95 °C with 0.5°C increments at 5 s/step). Samples for each condition were a pool of five biological replicates (RNAs extracted from the individual females were pooled together for the reverse transcription and subsequent qPCR). All samples were analyzed in triplicate, and controls without DNA were used in every plate.

3.2.4. Data exploitation and relative quantification of gene expression

The analysis of the raw data file extracted from the CFX96 Touch™ Real-Time PCR Detection System (Biorad) was performed in CFX Manager™ v3.1 (Biorad). The software allows one to perform a quality control of the qPCR run by analyzing the amplification curves of both positive and negative controls, the PCR efficiency, and the deviation between technical replicates. In Figure II-17 is represented the work environment of this software.

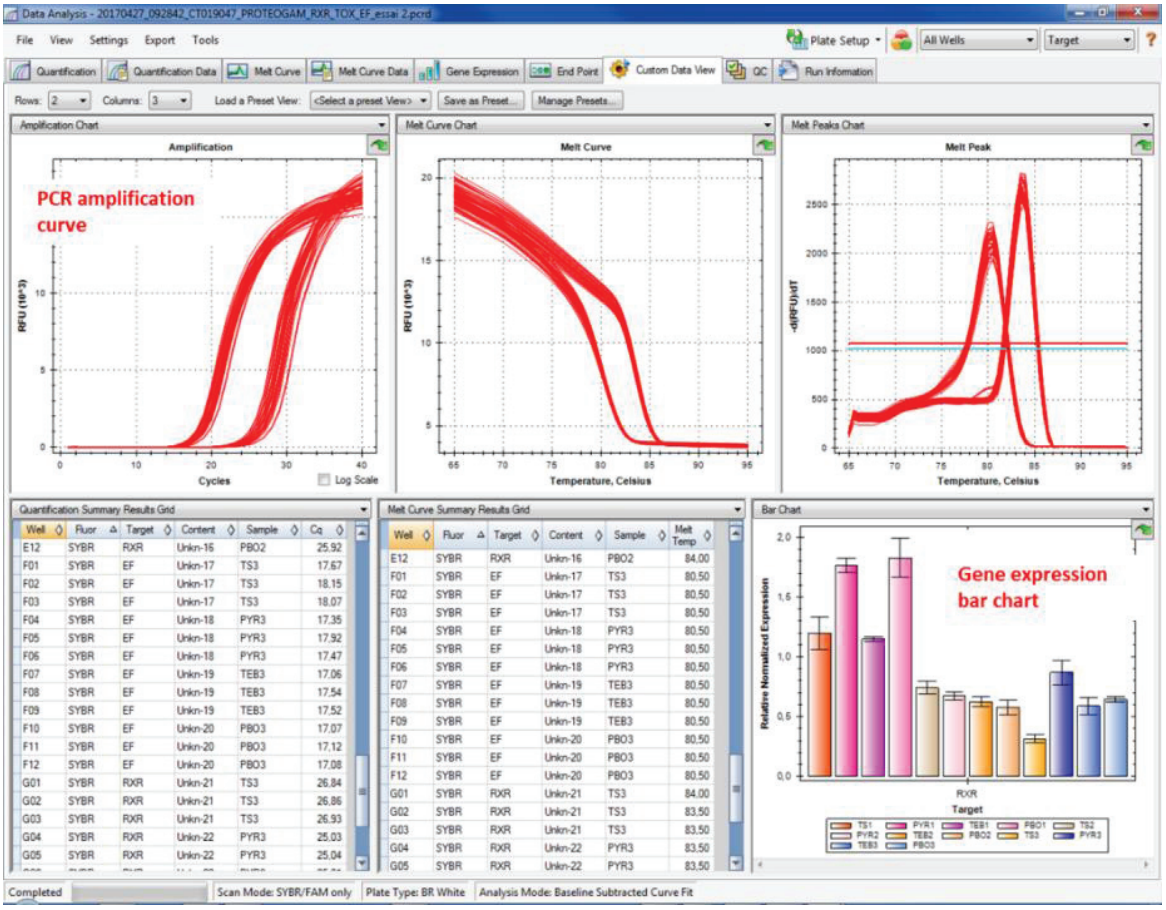


Figure II-17 – Overview of CFX Manager window for gene expression data analysis.

Gene expression levels were analyzed using the relative quantification method (ΔΔCt), illustrated in Figure II-18 (Livak and Schmittgen, 2001). This very popular method compares results from experimental samples with both a calibrator (control sample) and a normalizer (housekeeping gene expression). With this method, the Ct

value for the target gene is adjusted in relation to the Ct of the normalizer gene, in both test and calibrator samples.

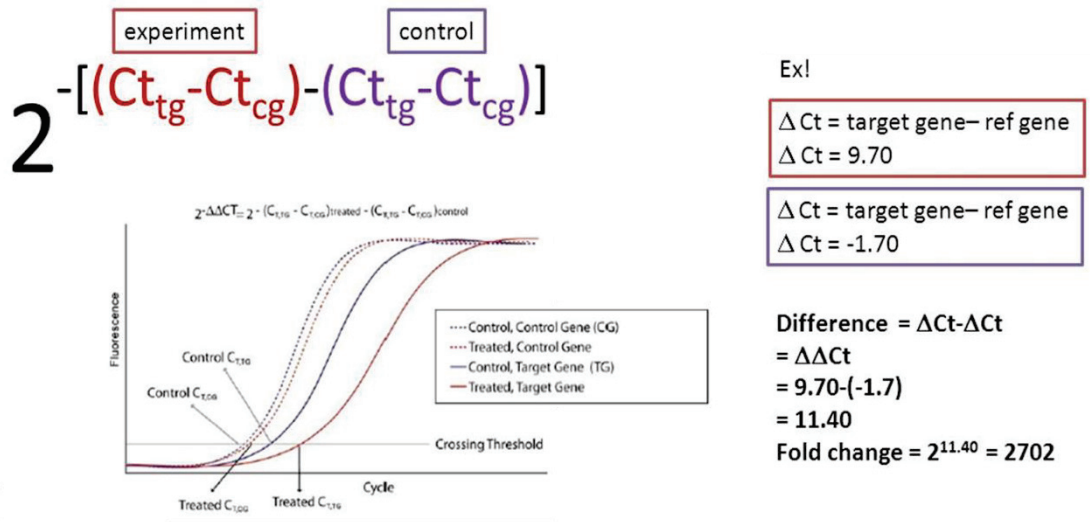


Figure II-18 - Application of the relative quantification method ΔΔCt for calculating the fold change expression of a reference and target genes, in a control (blue lines) and treated sample (red lines).

CHAPTER III. DEVELOPMENT AND APPLICATION OF A MULTIPLEXED PROTEIN BIOMARKER MEASUREMENT BY MASS SPECTROMETRY IN *GAMMARUS FOSSARUM*

This chapter assembles the studies that lead to the development and application of a MS-based SRM assay for the simultaneous quantification of a set of dozens of specific protein biomarkers. These works are presented in the form of three publications.

The first publication describes the development and validation of the LC-MS/MS-SRM analytical methodology employed for multiplexed biomarker measurement. This involved determining a wide range of parameters such as precision, accuracy, and reproducibility of the methodology, as well as the limits of detection and quantification for each biomarker protein.

The second publication comprises the experimental application of the SRM methodology in physiological and ecotoxicological studies in order to assess the biological significance of the biomarker measurements, along with their suitability to track toxicological effects in *Gammarus*.

The third publication presents the results obtained in a first test for assessing the potential applicability of the SRM multibiomarker methodology in field studies. Biomarker responses were analyzed and compared in gammarids exposed in four reference sites and thirteen sites subjected to chemical contamination during a caging deployment of a regional river monitoring network.

Publication nº1

Anal Bioanal Chem (2017) 409:3969–3991
DOI 10.1007/s00216-017-0348-0

RESEARCH PAPER

Multiplexed assay for protein quantitation in the invertebrate *Gammarus fossarum* by liquid chromatography coupled to tandem mass spectrometry

Aurore Charnot[1] · Duarte Gouveia[2] · Jean Armengaud[3] · Christine Almunia[3] · Arnaud Chaumot[2] · Jérôme Lemoine[1] · Olivier Geffard[2] · Arnaud Salvador[1]

Received: 15 November 2016 / Revised: 20 March 2017 / Accepted: 28 March 2017 / Published online: 17 April 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract: A highly-multiplexed liquid chromatography mass spectrometry - selected Reaction Monitoring (SRM) based assay for determination of 40 potential protein biomarkers from *Gammarus fossarum,* an ecotoxicological relevant species was described. The assay relies on 71 stable isotope-labeled reported peptide standards for the quantitation of proteins of interest in relation to essential physiological functions such as reproductive cycle, defense mechanism, enzymes involved in homeostasis process and in energy. A direct linear relationship between the spiked peptide concentration and the area under the peak was clearly demonstrated in biological extracts. Precision and accuracy was determined between 1.1 and 21 % and between 79 and 120% respectively depending on the selected protein in a few samples after optimization of digestion conditions. The validity of the assay was documented for several biomarkers linked with reproduction and molting process was performed with the assessment of protein levels throughout contrasted physiological process (sex, reproductive status). This assay is easy to use, robust, sensitive, and has high-throughput capabilities. The proposed strategy may be extended to any non-model organisms relevant in environmental science.

Keywords: *multiplex, protein quantification, mass spectrometry, biomarkers, invertebrate*

Introduction

In aquatic ecosystems, organisms are exposed to physical and chemical stressors. To evaluate and predict their biological effects, the need for relevant tools has increased considerably in the last decades. Biological monitoring, or biomonitoring, is the use of biological responses to assess changes in the environment, generally due to anthropogenic pressure. This valuable assessment tool is receiving increased use in water quality monitoring programs. The measurement of molecular biomarkers is one relevant approach for biomonitoring (e.g. OSPAR marine strategy), giving information at the sub-organismal level (biochemical, physiological and histological). Changes in molecular biomarkers can be related to an exposure to chemical compounds and some of them provide early warning indicators of possible effects on the ecosystem. As proteins are the molecular effectors of biological processes, their use as biomarkers is of strong interest because their modulation is likely associated with the effective impact on the physiology of the organism. However, the routine use of these tools in biomonitoring is confronted with several limitations, specifically in invertebrate sentinel organisms. Among these limitations are the lack of specific quantification methods and the fact that each biomarker measurement involves a specific method, leading to very expensive biomonitoring strategies in terms of time, cost and number of biological samples [1].

Lack of specific method in invertebrates mainly results from the lack of molecular knowledge in relevant environmental sentinel species, limiting the development of specific and direct method for the quantification of biomarkers. Recently, a novel approach, namely proteogenomics has emerged as a straightforward strategy for discovering proteins in non-model organisms [2, 3]. It relies on fast genome or transcriptome sequencing, six-reading frames translation of coding nucleic acids and creation of a molecular database comprising protein sequences validated by high-throughput proteomics data obtained by next-generation proteomics. For instance, based on a genomic and proteomic combined methodology, Trapp et al. [4] created a database consisting of 2,257 mass spectrometry (MS)-certified proteins from the amphipod crustacean Gammarus fossarum, a sentinel species for continental water biomonitoring.

Based on a set of proteins of interest extracted from the Gammarus fossarum protein sequence database [4], the objective of the present study was to setup a multiplex assay using an innovative approach to improve their specific identification and simultaneous absolute quantification.

For a reliable and sensitive method of protein quantification, enzyme-linked immunosorbent assays (ELISA test) is the golden standard. However, its development is cost and time expensive, but the main limit is related to the detection which is often based on its tridimensional conformation. Consequently, this method is sensitive to phylogenetic distance among species and poorly transferable, which is problematic in a biomonitoring perspective. Consequently, and mainly in invertebrates, the development of protein biomarkers is restricted to measurement of enzymatic activities, via indirect assays and expressed in nmol of substrate hydrolyzed or DO units.min−1.mg−1 proteins, often making difficult inter-laboratory comparisons.

Targeted MS in the Selected Reaction Monitoring (SRM) mode has emerged as an alternative to immunoassays for accurately quantifying proteins in human biological matrices [5, 6]. We already applied this method to Gammarus fossarum for measuring a key protein involved in reproduction, the vitellogenin [7, 8]. SRM, named method of the year 2012 [9], is a hypothesis-driven method offering outstanding multiplexing capabilities, as demonstrated recently for systems biology studies and the clinical evaluation of biomarker candidates [10]. SRM-based assays are developed on a hyphenated LC-triple quadrupole platform, and consist in monitoring at least three precursor fragment ion pairs and one or more peptides per enzymatically digested protein, those being selected among the "best flyer" and proteotypic peptide sequences. Heavy labeled isotopologs of the targeted peptides are concomitantly monitored to ensure precise and accurate measurements [10, 11]. Although not as sensitive as the most sensitive ELISAs, SRM routinely detects concentrations in the low to sub-nanogram/milliliter range after target peptide or protein enrichment or after extensive sample fractionation to reduce the likelihood of co-eluted interferences [12–16].

For an integrative approach and for a relevant interpretation in term of impact in regard to an exposure or in term of health status of monitored organisms, a multi-biomarker strategy must be used [17]. Indeed, such assay is required in some integrated indicators developed and proposed in recent years , as the Integrated Biomarker Response Index (IBR) [18]. In a perspective of a multi-biomarker deployment in routine biomonitoring, a multiplex methodology to monitor a panel of biomarkers in a single biological sample, in a single run, will be high-throughput and cost-effective. However, for any biomarker, a specific analytical procedure in terms of homogenization buffer, reaction medium or wavelength measurement is currently required. Multiplexing by mass spectrometry is in this case straightforward and offers innovative perspectives in ecotoxicology with lower resource investment in terms of organism and experimental maintenance.

Herein, we describe the development and application of a highly-multiplexed SRM-based assay for determination of 40 potential protein biomarkers in Gammarus fossarum. We also describe the precision, accuracy, linearity, and reproducibility of the method after peptides and SRM transition validation. To the best of our knowledge, this is the first multiplex protein quantitation assay in ecotoxicology.

Materials

Chemicals and Reagents

Acetonitrile, methanol and water (LC-Ms grade) were obtained from Fisher Scientist (Strasbourg, France). Dithiothreitol (DTT), iodoacetamide (IAM), formic acid (FA) (LC–MS grade), trypsin (treated TCPK from bovine Pancreas and not treated type 1X-S from porcine pancreas), urea and TRIS, EDTA, Triton X, sodium chloride, leupeptin and aprotinin were purchased from Sigma–Aldrich (St Quentin-Fallavier, France). Isotopically labelled peptides containing either a C terminal [$^{13}C_6$, $^{15}N_2$] lysine or arginine were synthesized by Fisher Scientist (Strasbourg, France) (purity > 97%). Absolute

ethanol and ethylic ether were obtained from Carlo Erba (Val de Reuil, France).

Candidate Marker Proteins

Protein sequences were extracted from the protein database previously created with a proteogenomics approach combining shotgun proteomics and RNA-seq data acquired on G. fossarum [4]. Hence, this database comprises the mass-spectrometry-certified sequences of 1873 proteins and their putative function as deduced from sequence similarities searches.

Biological part

Test organism collection and maintenance

Specimens of G. fossarum were sampled using a net (by kick sampling) from La Tour du Pin (Isère, France), upstream of the Bourbre River (mid-eastern France) and quickly transported to the laboratory (IRSTEA, Lyon) in ambient fresh water from the station. The organisms were kept during an acclimatization period of at least 10 days in 30-L tanks continuously supplied with drilled groundwater adjusted to the conductivity of the sampling site (i.e.,600 µScm−1) and under constant aeration. A 16/8 h light/dark photoperiod was maintained and the temperature was kept at 12±1 ◦C. Organisms were fed ad libitum with alder leaves (Alnus glutinosa). The leaves were conditioned for at least 6±1 days in water. Tubifex worms were given as a dietary supplement twice a week.

To validate the specificity of our peptides, we monitored their concentrations in organisms with contrasting reproductive status. Males and females in various reproductive statuses were obtained following partial life-cycle test protocols and determination criteria based on moulting process, as previously described in [19]. The reproducibility between all the samples (homogeneity) and the modulation of peptide concentrations in the tested conditions were evaluated. All the chosen organisms were weighed and immediately frozen in liquid nitrogen for storage at -80ºC until MS analysis.

Sample extraction and preparation

Specimens (whole bodies) were homogenized in Tris buffer (Tris-HCL 50mM; 100mM NaCl; 1mM EDTA; Triton X-100 0.1 % v/v; adjusted to pH 7.8. plus 10 µg/mL of each leupeptin and aprotinin), in a volume of 25 µL/mg of organism, with a bead homogenizer. All the homogenates were centrifuged at 10,000×g at 4 ◦C for 15 min. A volume of 250µL of clear supernatant was collected and delipidated by adding 750µL of ethanol/diethylether solution (1/1, v/v). The mixture was vortexed and kept on ice for 10 min. Then, samples were centrifuged at 10,000×g at 4 ◦C for 10 min. Clear supernatant was removed and the pellet was resuspended with 250µL Tris buffer. A volume of 3mL of ammonium bicarbonate 50mM and DTT at a final concentration of 15mM was added to the supernatant. Samples were incubated for 40 min at 60 ◦C. After cooling to room temperature, iodoacetamide was added (final concentration of 15mM) and the samples were placed in the dark at room temperature for 40 min. Then, 300µg of trypsin was added and samples were incubated for 1 h at 37◦C. Finally, 20µL of formic acid (FA) was added to stop the reaction. A volume of 10µL of a solution containing all the isotopically labeled peptides at 1µg/mL was added to the sample. Afterwards, a solid phase extraction was carried out using the 60mg hydrophilic-lipophilic balance (HLB) cartridge Oasis from Waters (Millford, MA). Cartridges were preconditioned with 1 mL of methanol and 1 mL of water acidified with 0.5% FA. Samples (3 mL) were then deposed on the cartridges. Afterwards, cartridges were rinsed with 1 mL of solution of water/methanol (95/5 v/v) acidified at 0.5% of FA. The analytes were eluted using 1 mL of methanol acidified at 0.5% of FA into an Eppendorf tube. Then the solvent was evaporated under nitrogen flow until a volume of 10 µL, which was diluted back with 90 µL of water/acetonitrile (90/10, v/v) with 0.1% of FA. The tubes were then vortexed for several seconds and the contents were transferred to glass autosampler vials.

Analytical part

Liquid chromatography and mass spectrometry

LC–MS/MS analysis was performed on an HP1200 series HPLC device (Agilent Technologies, Waldbronn, Germany) coupled to either a QTRAP® 5500 or 4000 LC/MS/MS System hybrid triple quadrupole/linear ion trap mass spectrometer (Applied Biosystems/ MDS Analytical Technologies, Foster City, CA, USA) equipped with a Turbo VTM ion source. The LC separation of the 20 µL injected sample was carried out on an Xbridge C_{18} column (100mm×2.1mm, particle size 3.5 µm) from Waters (Milford, MA, USA). Elution was performed at a flow rate of 300µL/min with water containing 0.1% (v/v) formic acid as eluent A and acetonitrile containing 0.1% (v/v) formic acid as eluent B, employing a linear gradient from 2% B to 33% B in 19 min, followed by a second linear gradient from 33% B to 64% B in 6 min. Then, column washing and re-equilibration was performed for 6 min. The injection duty cycle was 35 min, taking into account the column equilibration time. Instrument

control, data acquisition and processing were performed using the Analyst 1.5 software. The mass spectrometer was initially tuned and calibrated using polypropylene glycol, reserpine and Agilent Tuning Mix (all Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. MS analysis was carried out in positive ionization mode using an ion spray voltage of 5500 V. The nebulizer and the curtain gas flows were set at 50 psi using nitrogen. The Turbo V™ ion source was set at 550 ◦C with the auxiliary gas flow (nitrogen) set at 40 psi. The software Skyline v3.1 (MacCoss Lab Software, USA) was used to product a list of suitable SRM transitions which was used to scout and select the best transitions after analysis using isotopically labelled peptides. Q1 and Q3 masses as well as collision energy (CE) and declustering potential (DP) values were predicted by using Skyline. From the initial set of candidate SRM transitions, three transitions by peptides were selected for the final assay. The collision energies, declustering potential and collision cell exit potential (CXP) were optimized for each SRM transition chosen. The SRM transitions were monitored and acquired at unit resolution in Q1 and Q3, with a dwell time of 10 ms used for each transition, to obtain 10 data points per chromatographic peak minimum.

Trypsin proteolysis

Samples were prepared as described previously (section 2.2.3), except that they were removed from the bath at 37°C following different incubation time (0.5, 1, 2, 4, 6 and 12). The digestion was stopped by adding formic acid (20 µL). The substrate to enzyme ratio was kept constant (1:30). Two different types of trypsin were tested along these kinetics: normal trypsin and trypsin treated with anti-chymotrypsin effect (TCPK trypsin).

Standards and control quality preparation

Stock standard peptide solutions of isotopically labelled peptides were prepared by dissolving accurately weighed standard compounds in an H_2O/ACN (50/50, v/v) mixture containing 0.1% of formic acid to yield a concentration of 20 mg/mL. Solutions were prepared from the stock solutions and diluted further with either an H_2O/ACN (90/10, v/v) mixture containing 0.1% of formic acid or the matrix (pool of extracted and crushed *G. fossarum*) to prepare several solutions to reach the desired concentrations of 2500, 1000, 500, 400, 250, 100, 50, 10 and 5 and 5 ng/mL, which will be used to build the calibration curves. Quality controls were prepared from the stock solutions too to obtain concentrations of 150 ng/mL (i.e. QC1), 750 ng/mL (i.e. QC2) and 1500 ng/mL (i.e. QC3).

Method validation

For the determination of the limit of detection (LOD), limit of quantification (LOQ) and linearity of the assays, a standard curve from area was produced, based on nine samples containing equal amounts of *Gammarus fossarum* protein extract digests as background matrix. Each sample was spiked with an increasing amount of labelled peptides from 5 to 2500 ng/mL, spanning a 500-fold range. For linearity, each standard sample was analyzed three times. Determination of the signal-to-noise ratio was performed by comparing measured signals from samples with known low concentrations of labelled peptide with those of blank samples and by establishing the minimum concentration at which the peptide can be reliably detected. A signal-to-noise ratio between 3:1 and 10:1 were considered acceptable for estimating the LOD and LOQ, respectively. Precision and accuracy of the method were evaluated by performing 3 runs on 3 separate days. Each run consisted of one set of calibration standards, three (intra-batch) or nine (inter-batch) replicates of each QC concentration (150, 750 and 1500 ng/mL). Least square linear regression with a weighting factor of $1/x^2$ was used to plot the curve of the calibration standards. Criteria of precision determined at each concentration level as the relative standard deviation or coefficient of variation should not exceed 20 % and the average accuracy in determining the concentration must be within 80–120% of the expected value. If standard points for any concentration level felt outside these ranges, the entire concentration level would be removed from the curve and the linear regression equation would be recalculated. Calibration standards and the final calibration line contain at least 5 calibration concentrations.

Results and discussion

Development of a multiplex SRM assay

General strategy and selection of protein candidate biomarkers

Quantitative mass spectrometry of proteins has evolved dramatically over the last decade. Selected reaction monitoring (SRM) coupled with stable isotope dilution mass spectrometry (SID-MS) is now a powerful method for quantitative measurement of target proteins. It has been a principal tool for quantification of small molecules

in environment for number of decades. However, it was not until recently, that this approach benefitted from the rapid advances in mass spectrometry instrumentation and proteomics technologies; with these improvements, the mass spectrometry-based quantification techniques are becoming practically available to detect candidate proteins in a complex biological sample with sensitivity, high specificity, sensitivity, multiplexing capability, and precision. This approach capitalizes on the high specificity of SRM detection, and is capable of multiplexed protein quantitation in a rapid and inexpensive manner.

Selected Reaction monitoring performed on triple quadrupole mass spectrometer, was used for the identification of targeted protein. Peptides mixtures were analysed by LC-ESI-MS/MS. In SRM assay, the first quadrupole (Q1) and the last quadrupole (Q3) of triple quadrupole instrument filters ions and isolate specifically a list of targeted proteotypic peptides: the precursor ion on Q1 and the corresponding fragment on Q3. The second quadrupole serves as collision cell. The resulting precursor fragment ion pairs (transitions) are highly specific for a given proteotypic peptide and are therefore unique for a given protein in the analyzed proteome. SRM methods provide both absolute structural specificity for the analyte and relative or absolute measurement of analyte concentration when stable, isotopically-labelled standards are added to a sample in known quantities. Absolute quantification may be done by introducing a defined amount of chemically synthesized standard peptides, an artificial protein made of concatenated peptides, or a isotopically labelled, recombinantly expressed analogues of analyte proteins used at a known concentration: a PSAQ standard (Protein standards for absolute quantification). PSAQ are proteins that conserve the native context in which the quantified peptides exist; therefore, any differences between analyte and standard in proteolytic cleavage and procedural losses are minimized [20, 21]. This technique can therefore yield a true absolute and very high quality quantification. Despite that PSAQ is the best choice, it was not retained for cost and difficulties to obtain the proteins. Quantification with a synthetic peptide was preferred. When a synthetic, stable isotope labeled peptide is used as an internal standard, the relative concentration can be measured by comparing the signals from the exogenous labeled and endogenous unlabeled species. This can be done because they have the same physicochemical properties and differ only by their masses. The approximate retention time (RT) information can be used to restrict the time devoted for their detection of a specific transition and therefore allows the detection of multiple peptide ions in one measurement, a process that is referred to as scheduled SRM for multiplex assay[22].

Based on the G. fossarum protein database previously identified and annotated by proteogenomics (1873 MS-certified sequences) [3], a list of 55 proteins has been selected for quantification. This list included proteins identified as sex-specific (yolk proteins including vitellogenins; copine-8; Ca-transporting ATPase; and Prophenoloxidase), as well as proteins related to moult and hormonal regulation (Farnesoic acid methyltransferase FaMET; juvenile hormone esterase carboxylesterase JHE carboxylesterase; juvenile hormone epoxide hydrolase; Chitinase; and Cytochromes), immunity (Peptide antimicrobien; and Hemolectin) and proteins with annotation related to biomarkers currently used in ecotoxicology (superoxide dismutase SOD; glutathione-S-transferase GST; Catalase; Cellulase; Amylase; and Na+K+ATPase).

Development and Optimization of multiplex SRM assays

Optimization and Selection of high-responding signature peptide

The peptide selection process for targeted MS is summarized in Figure 1. The major challenge of SRM assay development is deciding which of the multiple potential proteotypic peptides should be chosen as signature peptides to stoichiometrically represent the protein of interest. Spectral count data obtained from the previous shotgun experiences [4] were the starting points for the choice of our proteotypic peptides. For each of our 55 protein candidates, peptides with the highest spectral count values and with no possible chemical modifications were selected as the high-responding signature peptides. Indeed, we focused on the best peptides observed during discovery phase. A set of 2 to 3 peptides per protein were selected, resulting in a list of 177 peptides. These signature peptides were synthesized with heavy labeled Arginine or Lysine [$^{13}C_6$,$^{15}N_2$] at the C-terminus of the peptides to serve as internal standards in the SRM assay.

The next step in SRM assay configuration is the optimization of MS parameters (m/z selection for Q1 and Q3, lens voltage,…) which will result in SRM assays with the best detection sensitivity. For this, Skyline software was used to generate in silico a complete b- and y-product ion series for doubly and triply charged precursor ions spanning the m/z 300–1200 range. A set of 20 SRM transitions per peptide was generated. The SRM selection was evaluated with isotopically labelled peptides. A

total of 3540 transitions were evaluated. Since labelled peptides have nearly identical chemical and physical properties to their endogenous counterparts, they elute at the same retention time, ionization, and fragmentation behaviors. Therefore, the SRM acquisition parameters and retention times used for the two forms of the same peptide were identical. The difference between the labelled and unlabeled peptides was the precursor and/or product ion m/z values (these values are different for y-ions, but identical for b-ions at the product ion level). The mass shift between homogenous and isotopically labelled peptides is shown in Table 1. Product ion fragments are mostly singly or doubly charged ions, with y-ions that are dominant in collision-based dissociation mechanism of tryptic peptides or with b-ions.

Upon these preliminary results, 3 transitions per peptide were selected for a total of 1062 transitions. To enable the detection of sensitive and interference-free ion pairs in the SRM assays, a series of optimizations were conducted. These involved optimizing the collision energy CE, declustering potential for each peptide, LC gradient slope, as well as screening the target analytes for interfering ions from the matrix (Data not shown). Together, this enabled the most intense interference-free SRM precursor/product ion pair to be monitored, which improved the accuracy and reliability of the obtained quantitative results. These optimizations were carried out on two instruments available: a 5500 QTRAP and a 4000 QTRAP 4000 mass spectrometers. This optimization step eliminated 52 peptides because of poor ionization or excessive charge state distribution (peptides not reported in table 1). All the final optimized parameters are presented in Table 1. As reported in this table, the determined CE voltage range from 10 to 60 Volts for on the 5500 QTRAP and from 20 to 64 Volts on the 4000 QTRAP. Our goal in this study was to develop a simplified assay for multiplex protein quantitation. In the perspective of the evaluation of large numbers of samples, our initial analytical specifications involved both sufficient throughput and robustness. Prompted by the analytical work flow applied for small molecule assays, only the SPE desalting method was considered for peptide fractionation. Similarly, a robust conventional bore liquid chromatography configuration (*i.e.* 2-mm-internal diameter C_{18} reverse phase column) was immediately preferred to the nanoflow chromatography device (75-µm internal diameter) used in most previous proteomics studies. The LC gradient was optimized to produce a final chromatographic run time of 35 min to ensure high throughput. Figure 2a shows coherent peptide elution from 10 to 25 min.

Validation of transitions and peptides for protein quantitation in biological samples

High specificity is an important feature of multiplex SRM assays. A transition that is shared by other analytes can potentially lead to a false positive peak assignment and thereby an inaccurate quantification. So, interference detection is a crucial step used to determine which SRM ion pairs are free interferences from other components which might be present in the sample. Once the SRM transitions have been developed and optimized, the specificity of the assays should be evaluated on a real sample such as *G. fossarum* protein extract. Unspecific signals may derive from other peptides with precursor/fragment ion pairs of similar masses including isotopic distribution. These peptides might have closely related sequences so that part of the transitions is identical. To conclude on the best peptide (sensitive and specific) for multiplex assay, SRM transitions were tested against about fifty protein extracts of *G. fossarum*, corresponding to various biological conditions (males, females, successive stages in reproductive cycle).

The first step of transition validation was the parallel acquisition of several transitions for a targeted peptide (3 transitions for 125 peptides light and heavy for a total of 750 transitions). At the time of peptide elution, such transitions should yield a perfect set of 'co-eluting' intensity peaks, if they are derived from the same peptide, and the area ratio from the same precursor should be constant between biological samples. As illustrated in figure 2, at the elution time of the missed cleavage peptide KAEIGIAMGSGTAVAK of the protein 4227, we have six transitions eluted at the same time (figure 2 b): three transitions corresponding to the heavy peptides and three heavy and light for the light peptide. This shows that heavy and light peptides have the same retention time. Moreover, the area ratio between the three transitions of the heavy peptide (figure 2 d) and the three transitions of the light peptide (figure 2 c) is constant through the ten samples analyzed. The peptide KAEIGIAMGSGTAVAK was then validated and chosen as a reporter peptide for the protein 4227. However, peptides with area ratios not constant between samples were rejected, as illustrated in Electronic Supplementary Material figure S1.

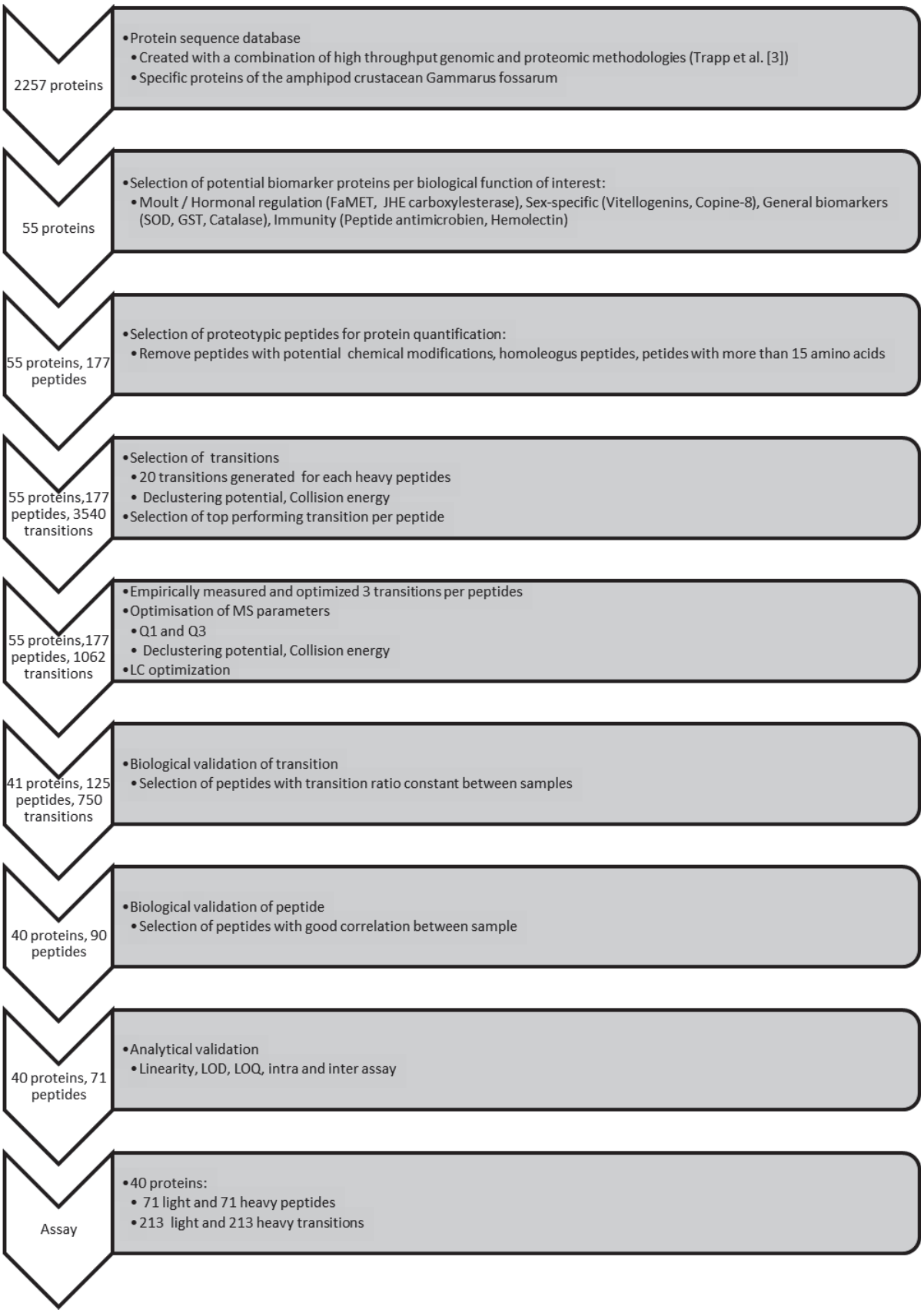


Figure 1: Flow diagram depicting the steps involved to reduce an initial list of candidate protein biomarkers to a viable multiple reaction monitoring assay and workflow.

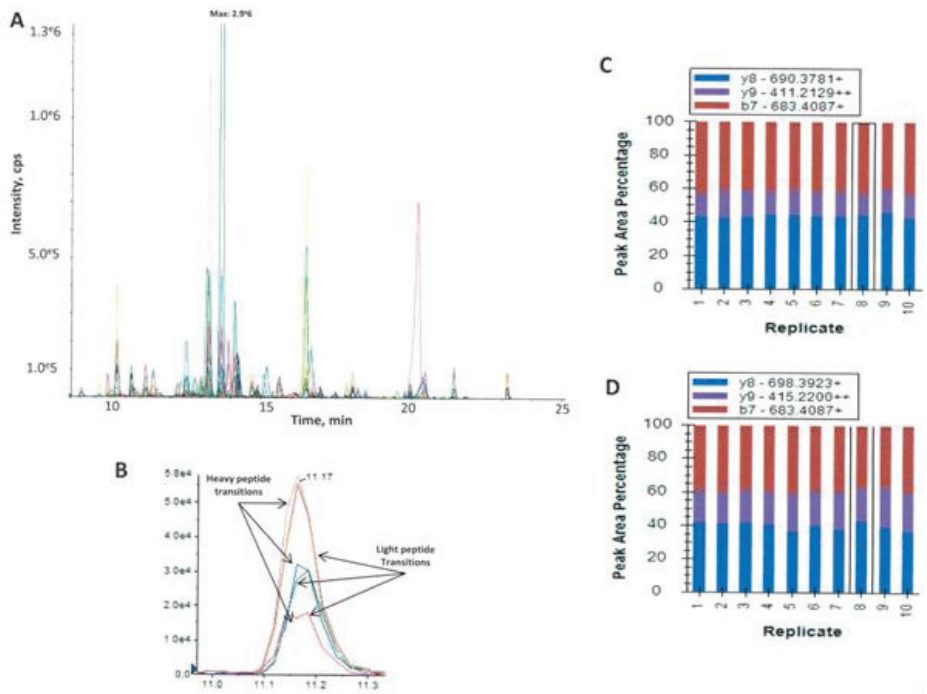


Figure 2: Extracted ion chromatograms for a total of 426 MRM ions pairs for the natural and labelled peptides (**A**), chromatographic peak for the peptide KAEIGIAMGSGTAVAK of the protein 4227 with six followed transitions (three for the natural and three for the labelled) (**B**) and area ratio between the three transitions of light peptide (**C**) and heavy peptide (**D**) KAEIGIAMGSGTAVAK of the protein 4227 for 10 samples.

The second step was the biological validation of peptides in terms of common assignment to the protein of interest. Indeed, for each protein, signature peptides should be correlated between themselves regardless of biological sample. For each protein, reporter peptide areas were correlated between themselves for all sample extracts. If peptides are from the same protein, the ratio between their areas must be the same regardless the concentration of the peptide and the correlation coefficient (r^2) between the areas must be >0.90. It is the same for the chosen transition for one peptide, where the ratio between areas must be unchanged with the concentration and have a good correlation. As seen in figure 3, for signature peptides of prophenoloxidase (protein 2562), good correlations were observed, plotting the peptide area of 4 peptides with corresponding correlation coefficients higher than 0.99 (figure 3a and b). Likewise, good correlation coefficients are found for the three other peptides between themselves (r^2>0.99). The EVLELPAQYMEFTR and KFLPSSGVDDLIK peptides of the protein 18473 (annotated as copine-8 function) are exemplifying bad signatures for reporter peptides. As illustrated in Fig. 3c, the correlation between peptides is low (regression coefficient of 0.73 and 0.85). For other proteins, all correlation coefficients were superior to r^2>0.94 (Electronic Supplementary Material Table 1) which prove that all chosen peptides are specific for the associated proteins and could be used as quantification peptides. These results demonstrate that whatever the sample, the ratios between reporter peptides are constant. This means that (1) specificity, ionization conditions and, more generally, matrix effects, are similar for each peptide and that (2) the sample preparation and above all digestion conditions are reproducible. Finally, after all the optimization process, more peptides are detected with API55000 Qtrap than API4000 Qtrap instrument: 71 peptides instead of 46 peptides, as reported in Table 1. Such result was expected because of higher analytical performances of the former instrument. MS and LC optimization (API5500 QTRAP and 4000 QTRAP) allowed to detect respectively 19 and 16 proteins based on only one peptide, and 21 and 13 with at least two or more measured peptides, respectively. Mainly data from the API5500 QTrap will be discussed in the next paragraphs.

Proteolysis optimization

In PSAQ experiment, the primary structure of PSAQ standards is identical to the endogenous target proteins. Thus, PSAQ standards are used to correct analytical variabilities due to upstream sample handling or incomplete proteolysis trypsin and cleaves both forms of proteins. In our study, selected peptides are used as surrogates of intact

proteins in SRM assay based on the assumption that digestion is complete and reproducible.

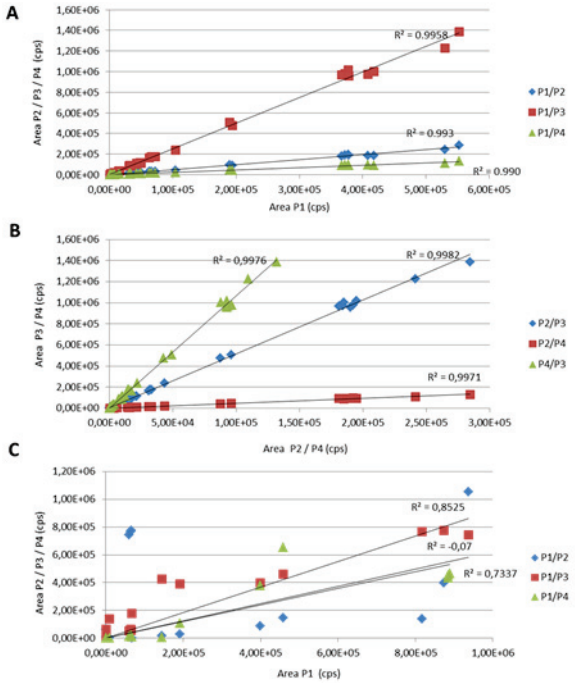


Figure 3: Correlation of the chromatographic peak area between the peptides of: **(A and B)** protein 2562 (P1: GIDIIGDAFEADR, P2: GIDFGTTQSVR, P3: APILEGYFSK and P4: ATQPSYTVAQLELPGVNITR) and **(C)** protein 184473: (P1: EVLELPAQYMEFTR, P2: KFLPSSGVDDLIK, P3: SYQILLLITNGGLSDIDATKK and P4: VAERDIVQFVELR).

Concentrations are reported on the basis of the relative peak areas of the internal standard and natural peptides. Incomplete enzymatic digestion could result in inaccurate concentration determination — thus if the efficiency of digestion and natural peptide recovery are less than 100%, the reported concentrations could be underestimated. Thanks to its reliability and specificity, trypsin is the most common proteolytic enzyme, which cleaves peptides on the C-terminal side of lysine and arginine amino acid residues. Proteins can differ greatly in their susceptibility to proteolysis, and this requires a digestion protocol that is optimal for all proteins for accurate measurement. The optimal digestion protocol has been shown to be analyte-dependent and may be unattainable for all proteins if a single digestion protocol is used [23]. To determine the optimal digestion conditions, we first evaluated the total protein concentration used for tryptic digestion and adjusted the trypsin-to-substrate ratio. This was challenging to attain in practice, especially if we consider performing quantification from an entire organism because of the need to add an extremely large amount of trypsin to reach the conventionally accepted ratio of trypsin over protein, which must be in the range 1/20–1/100. The final ratio was 1/30 (w/w). Modification of this ratio did not significantly influence the signal intensity. The most critical parameter is the digestion time. A time-course evaluation for digestion efficiency has been done to demonstrate digestion completeness, the abundance of target signal and stability of signal over the length of time of the digestion. G. fossarum proteins were extracted in replicates, digested at 37◦C and quenched with formic acid at 6 time points. Signal peak areas were plotted versus digestion time (Figure 4).

Two type of bovine pancreas trypsin were evaluated: non-treated and treated and dialyzed with L-1-Tosylamide-2-phenylethyl chloromethyl ketone (TPCK). Treatment with TPCK reduces the slight but notable chymotrypsin activity which is usually present in trypsin. The digestion was found to be complete after 0.5 or 1 h for most peptides, as shown in Figure 4. Depending on peptides, five profiles were observed. Except, for one peptide (peptide EVFIGGWSNQNSAIR - protein 134275) in figure 4c, longer digestion time did not improve the detection. Evaluation of kinetics of the individual signals reveals that once the fragments have been released from the protein their signals are falling off either quickly (Figure 4d) or slowly (figure 4 a,e) due probably to their specific susceptibility regarding degradation or chemical modification. Peptide concentration is more important with trypsin TCPK and the variations of peptide concentration upon time are lower. A digestion time of 1h and Trypsine TCPK were the parameters selected for the final method.

Assay validation

After selection of peptides and transitions for protein quantification assay, SRM assay is performed. Proteins from entire G. fossarum organisms were analyzed after grinding step. Protein quantitation was performed by comparing the extracted ion signal (peak height or peak area) of the isotope-labeled and native forms of the proteotypic peptides. As a general rule, isotope dilution-based quantification methods display good linearity and precision, regardless of the quantification standard used. Application of the isotope dilution principle to MS quantification of proteins by using isotope-labeled peptide analogues is termed AQUA (24) for “Absolute Quantification”. AQUA standards correspond to synthetic peptides that are spiked into the samples after proteolysis step. As previously mentioned, the retention times and fragmentation patterns of the native peptides formed after tryptic digestion are strictly equivalent to AQUA standards. Since an

absolute amount of the AQUA peptides is added, the ratios of both peak areas are used to determine the precise amount of proteins in the sample extract. The internal standard is present after digestion as native peptides are formed. Peptide extraction efficiency, absolute losses during sample handling (including sample concentration), and variability during introduction into the LC–MS system do not affect the ratio of native and AQUA peptide abundances. Since the concentrations of the different proteins were unknown, a wide range of concentrations for the standard was evaluated. The range of the calibration curve was defined as wide as possible from 5 to 2500 ng/mL. Linearity, LOD, LOQ, intra-run precision and accuracy were established to describe performance evaluation method as long as matrix effect.

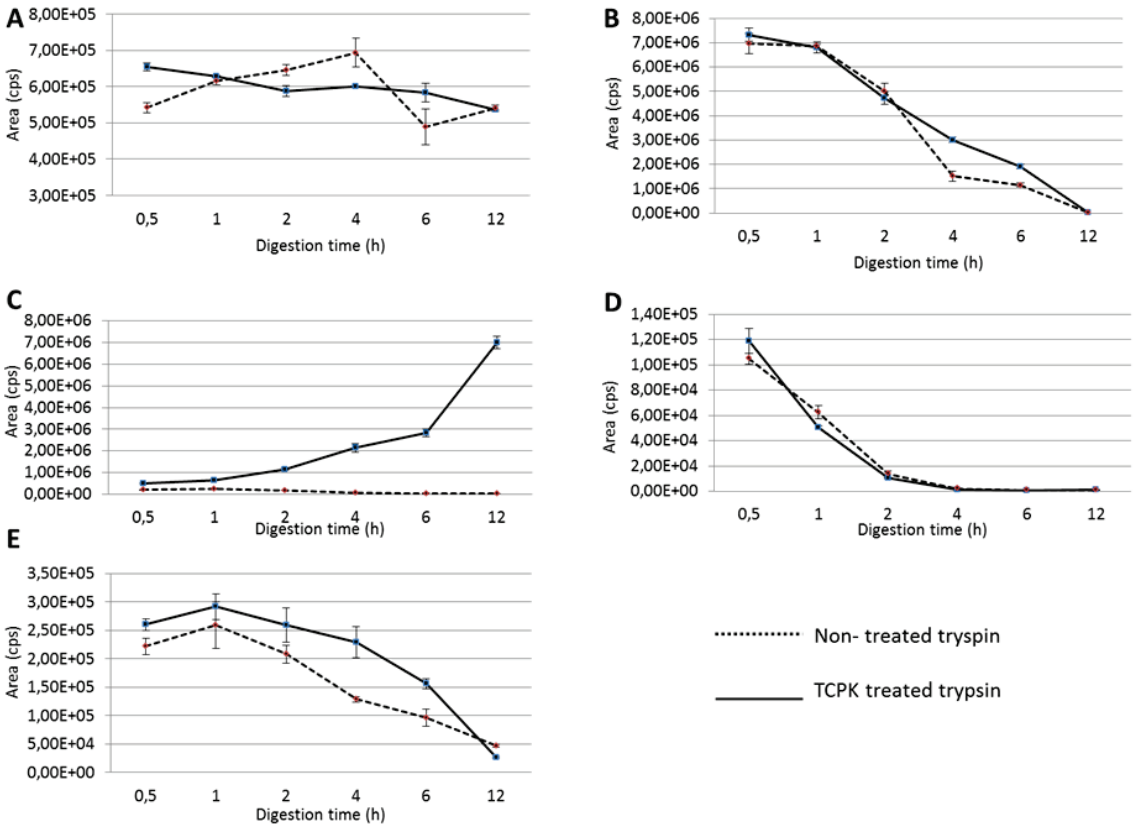


Figure 4: Digestion efficiency at six different incubation times of 30 min and 1, 2, 4, 6 and 12 h for two type of trypsin: Non-treated trypsin (dotted line) and TCPK treated trypsin (solid line). Five different profiles were obtained after analysis. Profile (A) shown is for peptide ADPALGQAIQER (protein 110912) and is the same for 8 other peptides, (B) peptide AAIETAFVNHLK (protein 206469) and for 17 other peptides, (C) is only for peptide EVFIGGWSNQNSAIR (protein 134275), (D) peptide SYQILLLITNGGLSDIDATKK (protein 18473) and for 8 other peptides and (E) is for peptide AVAEIVQDYDSDGFFPALGFGGK (protein 34845) and 27 other peptides.

Linearity, LOD, LOQ

Proteins were quantified by comparison to the peptide area response curves. These curves were generated from the LC-MS-MS analysis of labeled standard peptide samples, and required that a given concentration level exhibit a high precision and accuracy. A weighted (1/x2) least-square linear regression of area response versus concentration was used for the calibration. A set of 15 proteins among the 55 targets did not meet qualification and quantification criteria (i.e., an average precision below 20% CV per concentration level, an average accuracy below 80–120% per concentration level) for a minimum of 3 consecutive concentration levels, as well as a co-eluting noise in the SRM channels (see section 3.1), and thus remained not quantified. The remaining 40 proteins were quantified based on 71 validated peptides, their curves exhibiting a strong linear correlation as evaluated with a correlation coefficient between 0.964 and 1.000 as reported in table 2 (for API4000 Qtrap, 28 proteins with 45 peptides and 0.961< R2<1.000, see Electronic Supplementary Table 2).

For the determination of the LOD and LOQ several strategies are described in the literature [24]. Among them, the one relying on the signal-to-noise ratio is faster and therefore preferred. The LOD and LOQ were determined as the analyte concentration that produced a peak signal higher than three and 10 times the background noise from the extracted ion current chromatograms, respectively. For instance, LOQ are between 9 and 446 ng/mL for Glutathion S transferase protein like n°142711 and Prophenoloxidase n° 15561, respectively. The mean LOD is approximately 40 ng/mL (Table 2).

Precision and accuracy

Intra-day accuracy (mean determined concentration/nominal concentration × 100) and precision (SD/mean concentration × 100) (n = 3 for each level) were evaluated by analysis of QC samples the same day. Inter-day accuracy and precision were determined by repeated analysis over 3 consecutive days. The concentration of each sample was determined using calibration curve. The intra- and inter-day precision and accuracy are shown in Table 2 and Electronic Supplementary Material Table 2 for API4000Qtrap. The intra-day accuracy ranged from 76.7 to 119.1% while precision ranged from 5 to 19%. The inter-day accuracy ranged from 80.3 to 118.9% and precision from 8 to 22% (for the QC at 150 ng/mL). These results reveal that precision, accuracy and reproducibility of the method are acceptable for the 40 proteins assayed, taking into account that the assay is carried out on total organisms. This protocol has then been applied for quantification of 71 potential peptide biomarkers (see section 3.1).

Matrix effects

It is important to determine the matrix effects with the use of the LC–ESI–MS system. Indeed, electrospray source ionization can reduce or enhance the analyte response. Moreover, the absence of high sample preparation steps as multidimensional SPE optimization with reverse phase and ionic retention mechanism can result in significant matrix effects despite the time savings. Indeed, some endogen compounds from the matrix can cause ionization competition. The (US Food and Drug Administration) FDA guidelines recommend evaluating matrix effects but do not provide guidance on how to determine them. Several methods have been proposed for the detection and assessment of matrix effects, including post-extraction spikes [25]. The post-extraction spike method evaluates matrix effects by comparing the signal response of an analyte in neat mobile phase with the signal response of an equivalent amount of the analyte in the blank matrix sample spiked post-extraction. The difference in response determines the extent of matrix effect. A comprehensive assessment of matrix effects and recovery studies were carried out for all peptides. The matrix effect was determined for a concentration level of 1500 ng/mL (i.e. QC 3), with the following formula:

Matrix effect=((Heavy peptide peak area in matrix)/(Heavy peptide peak area in solvent)-1)*100

Based on the results obtained (data not shown), the amount of ion enhancement ranges from 5.6% to 192% and the amount of ion suppression ranges from -1.6% to -186.2%. Matrix effects depend on each peptide. Results are summarized in Figure 5 as a Pareto chart. Matrix effects can be reduced simply by injecting small amounts of samples or by diluting samples. Histograms represent the number of peptides per interval.

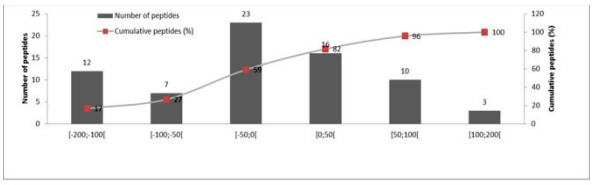


Figure 5: Distribution of matrix effect (%) at a concentration level of 1500 ng/mL with LC-MS/MS method using a 5500 Qtrap triple quadrupole.

The abscissa represents effect matrix gain/loss intervals (in %) for the 71 peptides most intense SRM transitions (see Electronic Supplementary Material Figure S1 for API4000Qtrap). Figure 5 shows that there are 42 peptides for whose ion are suppressed and 29 enhanced. The curve shows that it represents 59 % and 41 %, respectively. However, this approach was not retained as the sensitivity of the assay was not high enough for all peptides. Another method used in our lab to reduce or eliminate matrix effects include optimizing sample preparation to remove interfering compounds from the samples [26], changing chromatographic parameters to avoid coelution of analytes and interfering compounds to reduce the occurrence of matrix effects in the ion source [12]. Such methods were not retained. Indeed, in a multiplex assay, a generic SPE sample clean-up should be used. Although there are no methods to eliminate matrix effects in a multiplex assay, the most well-recognized technique available to correct matrix effects is the internal standardization using stable isotope–labeled versions of the analytes. Indeed, the internal standard, structurally analogous to the analytical

target, is expected to experience a matrix effect similar to that of the target in any batch of biological samples. So, in this context, the accuracy/precision of the method is considered unaffected by the matrix effect, or by variations in sample preparation in which the internal standard is added. Matrix effects are important in the present assay as the proteome from a whole organism is prepared and injected. We have shown the interest and the need to use isotopically labelled peptides which are spiked at known concentrations in samples and allow quantifying peptides despite matrix effects.

Biological validation

To validate this proof of concept SRM-method in environmental science and to determine the relevance of the methodology regarding future ecotoxicological applications, 22 candidate biomarkers of interest (sex-specific proteins and/or proteins with key physiological functions) were simultaneously monitored in several biological samples.

First, the levels of proteins annotated as involved in the reproduction process were compared in sexually mature male and female organisms (Figure 6). Males were isolated after copulation and sampled at day seven of the spermatogenesis process, and females were sampled based on the developmental stage of their embryos (well-developed juveniles in the marsupium). Electronic Supplementary Material Table 3 shows the concentrations obtained for the entire set of peptides after the LC-MS/MS analysis and peptide absolute quantification for each organism. Based on these raw data, Figure 6 illustrates the levels of the sex-specific proteins between male and female organisms. These proteins were previously classified as being sex-specific by comparative shotgun proteomic analysis of male and female reproductive tissues and functional annotation [3]. As expected, yolk proteins are present at much higher levels in mature females than in males, with fold changes ranging between 42 (for the peptide HIEIFSPITK of the protein 64) till 388 (for the peptide ISPLINSPSDLPK of the protein 39606). Male-specific proteins were also detected at higher levels in male organisms, but with fold changes significantly lower than for the yolk proteins – range between 1.5 for the peptide GTLAVIPVQNR (protein 1917) and 5.7 for the peptide KAEIGIAMGSGTAVAK (protein 4227).

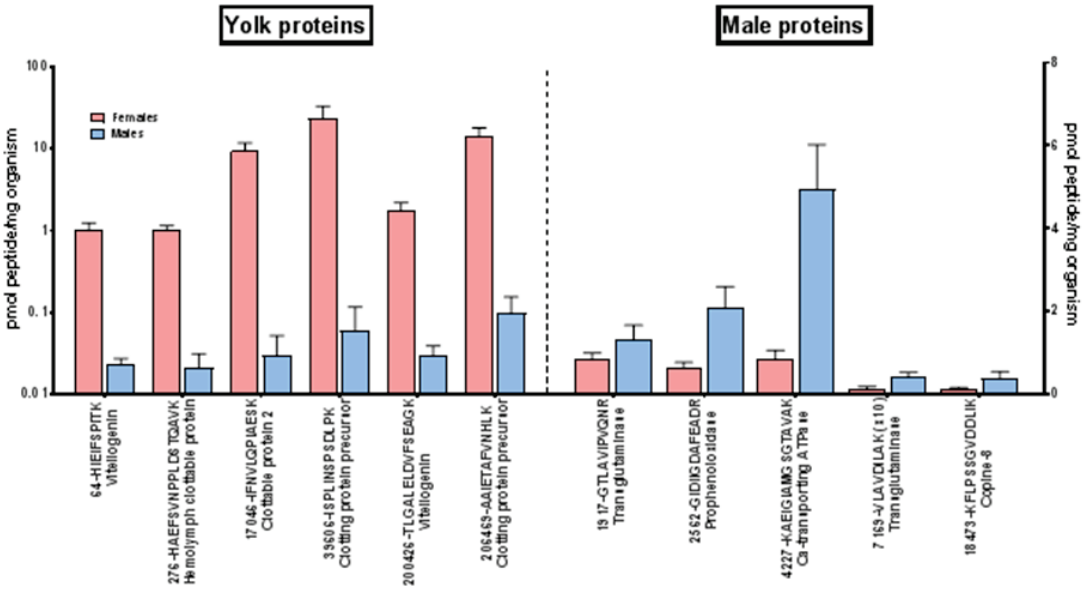


Figure 6: Concentration (in pmol/mg of organism) of female- and male-specific proteins in sexually mature organisms. Statistical significance determined using the Holm-Sidak method, with alpha = 0.05. Each row was analyzed individually, without assuming a consistent SD.

Besides male/female comparison, we also compared peptide levels between organisms at contrasting stages of their reproductive cycles (Figure 7). Males were sampled 7 and 21 days after copulation corresponding to pre-moult D2 and post-moult B stages, respectively. Females were also sampled at the B and D2 moulting stages, corresponding to periods of early and maximum

vitellogenin synthesis respectively [7,11]. Electronic Supplementary Material Table 4 shows the concentrations obtained for the entire set of peptides after the LC-MS/MS analysis and peptide absolute quantification for each organism. In figure 7 are represented examples of pertinent proteins that were quantified at different levels in the two-contrasting reproductive status analyzed. As shown in figure 7A, yolk proteins are more abundant (fold changes between 21 and 50) in females at the D2 reproductive stage, where the oocytes are in the later developmental stages and consequently much bigger than in the B stage. This agrees with the previous observations of a strong increase in oocyte growth beginning at the C2 stage and reaching a maximum oocyte size in the D2 stage [7,11]. Besides, the protein 144144 (JHE-carboxylesterase) was more abundant in females at the D2 stage as shown in figure 7B (fold change of 2.15, p value =0.006403). This protein is annotated as an enzyme related to the degradation of methyl farnesoate, a crustacean hormone implied in moutling process and control of oogenesis.

In male organisms, the proteins 2562 (annotated as Prophenoloxidase -PPO), 18473 (annotated as copine-8), and 213317 (annotated as cellulase) are present at higher levels in the post-moult stage B than at the pre-moult stage D2 (only the protein 2562 has a statistical significance, with a fold change of 2 and p value = 0.0441). The pattern for the copine-8 protein which has been previously shown to be a male specific protein, but with no known function, asks for deeper functional investigation. The modulation of the two other proteins are in accordance with what is known about their potential functional role in crustaceans. PPO is the precursor form of the enzyme phenoloxidase (PO), which participates in the moulting process of crustaceans [23,24]. A study in the crustacean decapod M. rosenbergii [25] has shown that the expression of the PPO gene reaches a peak during the post-moult stages, and is significantly decreased during the inter- and pre-moult stages. Our results show that for G. fossarum the PPO levels also decreased in the pre-moult stage D2, probably because PPO is continuously converted to PO during the inter- and pre-moult in order to accomplish the moulting process. The decrease in cellulase levels in the pre-moult stage D2 is also in accordance with previous observations in Gammarus fossarum [26], which reported a peak of digestive enzymatic activity concomitant to energy acquisition in the early stages of the moult cycle, followed by a gradual decrease during the inter- and pre-moult stages, in relation to copular behavior.

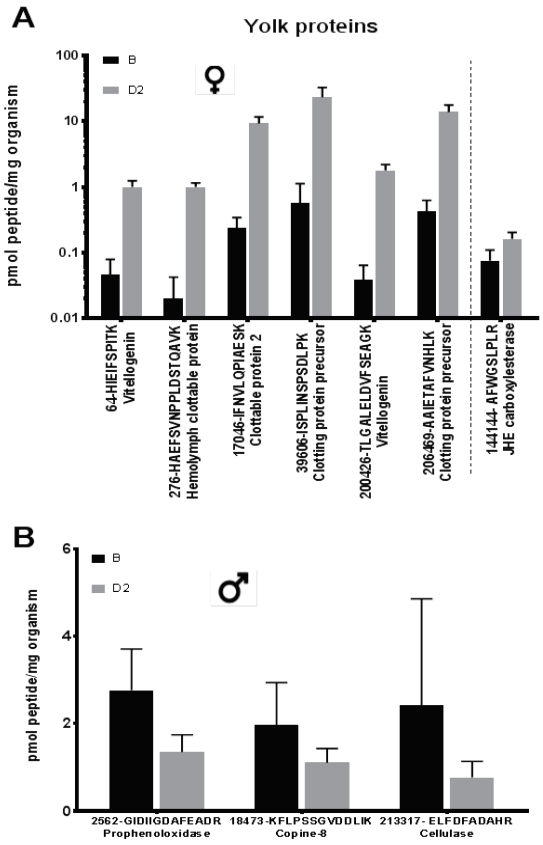


Figure 7: Concentrations (pmol/mg OF organism) of several proteins of interest in (A) female and (B) male organisms at contrasting reproductive cycles. Statistical significance determined using the Holm-Sidak method, with alpha = 0.05. Each row was analyzed individually, without assuming a consistent SD.

Globally, the application of the SRM multiplex quantification allowed for the multiplex quantitation of a set of 22 proteins in several organisms that differed in sex and in reproductive status. We managed to simultaneously quantify, in one whole organism, peptides at levels as little as 0.0033 pmol/mg and as higher as 33.61 pmol/mg of organism. The targeted quantitation values are consistent with the literature and with the physiological roles of the proteins studied. The fact that we could follow the different modulations of some proteins related to the moulting cycle and embryonic development is promising for future field applications. The female gammarid moulting cycle is indeed used as a biomarker in the monitoring of water bodies contamination [5]. Contaminated waters may cause delays in the moulting cycles with adverse outcome on reproduction of these organisms. These results show that the multiplexed SRM protein

quantification could be a pertinent method to complement the data obtained from the physiological measures used in the analysis (moult stage, oocyte size). [27]

Conclusion

We have developed a robust and accurate SRM-LC/MS-MS based method for quantifying a highly multiplexed set of 40 proteins in *Gammarus fossarum*. The method uses internal labeled standard peptides and does not require the costly development of specific antibodies. This is an important breakthrough approach in ecotoxicology. This is the first time that such innovative strategy is proposed for the development of a multiplex assays for protein biomarkers in aquatic species. Once several proteins are measured, overall multiplex assay costs are lower than if one chooses to obtain the same information using separate indirect assays. In addition to requiring smaller sample volumes, multiplex technology also offers savings in terms of time required to complete the assay. The development of SRM multiplex assay will allow new monitoring approaches that could be applied rapidly in a near future in environmental issue for other model species as no antibody are required. The development of a reliable immunoassay for one target protein is expensive, has a long development time, and is dependent upon the generation of high quality protein antibodies which is incompatible with most common basic funding in the environmental sphere. However, in *G. fossarum,* further functional studies are needed before proposing these 40 proteins as biomarkers in gammarids (eg.an indicator of reproductive impairment). Similarly, analytical developments are still further possible for increasing its throughput, such as for example automation of sample preparation and extraction process.

Acknowledgments

We thank the French “Ministère de la recherche et de l’enseignement supérieur” for the fellowship of Aurore Charnot. We also thank the ANR program “ProteoGam” (ANR-14-CE21-0006-02) and the Regional Water Agency Rhône-Méditerranée-Corse for financial support.

Compliance with ethical standards

Conflict of interest: The authors declare that they have no conflict of interest and ethical approval was not required for this work.

References

1. Trapp J, Armengaud J, Salvador A, Chaumot A, Geffard O (2014) Next-generation proteomics: toward customized biomarkers for environmental biomonitoring. Environ Sci Technol 48:13560–72. doi: 10.1021/es501673s

2. Renuse S, Chaerkady R, Pandey A (2011) Proteogenomics. Proteomics 11:620–630. doi: 10.1002/pmic.201000615

3. Armengaud J, Trapp J, Pible O, Geffard O, Chaumot A, Hartmann EM (2014) Non-model organisms, a species endangered by proteogenomics. J Proteomics 105:5–18. doi: 10.1016/j.jprot.2014.01.007

4. Trapp J, Geffard O, Imbert G, Gaillard J-C, Davin A-H, Chaumot A, Armengaud J (2014) Proteogenomics of Gammarus fossarum to document the reproductive system of amphipods. Mol Cell Proteomics 13:3612–25. doi: 10.1074/mcp.M114.038851

5. Addona TA, Shi X, Keshishian H, Mani DR, Burgess M, Gillette MA, Clauser KR, Shen D, Lewis GD, Farrell LA, Fifer MA, Sabatine MS, Gerszten RE, Carr SA (2011) A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. Nat Biotechnol 29:635–43. doi: 10.1038/nbt.1899

6. Anderson L (2006) Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. Mol Cell Proteomics 573–588. doi: 10.1074/mcp.M500331-MCP200

7. Simon R, Jubeaux G, Chaumot A, Lemoine J, Geffard O, Salvador A (2010) Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: application to vitellogenin in Crustacea (Gammarus fossarum). J Chromatogr A 1217:5109–15. doi: 10.1016/j.chroma.2010.06.015

8. Jubeaux G, Simon R, Salvador A, Lopes C, Lacaze E, Quéau H, Chaumot A, Geffard O (2012) Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: inconclusive experience. Aquat Toxicol 122–123:9–18. doi: 10.1016/j.aquatox.2012.05.007

9. Gillette MA, Carr SA (2013) Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. Nat Methods 10:28–34. doi: 10.1038/nmeth.2309

10. Surinova S, Schiess R, Hüttenhain R, Cerciello F, Wollscheid B, Aebersold R (2011) On the development of plasma protein biomarkers. J Proteome Res 10:5–16. doi:

10.1021/pr1008515

11. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A 100:6940–5. doi: 10.1073/pnas.0832254100

12. Simon R, Passeron S, Lemoine J, Salvador A (2014) Hydrophilic interaction liquid chromatography as second dimension in multidimensional chromatography with an anionic trapping strategy: Application to prostate-specific antigen quantification. J Chromatogr A. doi: 10.1016/j.chroma.2014.05.063

13. Jaffuel A, Lemoine J, Aubert C, Simon R, Léonard JF, Gautier JC, Pasquier O, Salvador A (2013) Optimization of liquid chromatography-multiple reaction monitoring cubed mass spectrometry assay for protein quantification: Application to aquaporin-2 water channel in human urine. J Chromatogr A 1301:122–130. doi: 10.1016/j.chroma.2013.05.068

14. Whiteaker JR, Zhao L, Anderson L, Paulovich AG (2010) An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. Mol Cell Proteomics 9:184–96. doi: 10.1074/mcp.M900254-MCP200

15. Shi T, Fillmore TL, Gao Y, Zhao R, He J, Schepmoes AA, Nicora CD, Wu C, Chambers JL, Moore RJ, Kagan J, Srivastava S, Liu AY, Rodland KD, Liu T, Camp DG, Smith RD, Qian W-J (2013) Long-gradient separations coupled with selected reaction monitoring for highly sensitive, large scale targeted protein quantification in a single analysis. Anal Chem 85:9196–203. doi: 10.1021/ac402105s

16. Qian W-J, Kaleta DT, Petritis BO, Jiang H, Liu T, Zhang X, Mottaz HM, Varnum SM, Camp DG, Huang L, Fang X, Zhang W-W, Smith RD (2008) Enhanced Detection of Low Abundance Human Plasma Proteins Using a Tandem IgY12-SuperMix Immunoaffinity Separation Strategy. Mol Cell Proteomics 7:1963–1973. doi: 10.1074/mcp.M800008-MCP200

17. Depledge MH, Galloway TS (2005) Healthy animals, healthy ecosystems. Front Ecol Environ 3:251–258. doi: 10.1890/1540-9295(2005)003[0251:HAHE]2.0.CO;2

18. Vieira CED, Costa PG, Lunardelli B, de Oliveira LF, da Costa Cabrera L, Risso WE, Primel EG, Meletti PC, Fillmann G, Bueno dos Reis Martinez C (2016) Multiple biomarker responses in Prochilodus lineatus subjected to short-term in situ exposure to streams from agricultural areas in Southern Brazil. Sci Total Environ 542:44–56. doi: 10.1016/j.scitotenv.2015.10.071

19. Geffard O, Xuereb B, Chaumot A, Geffard A, Biagianti S, Noël C, Abbaci K, Garric J, Charmantier G, Charmantier-Daures M (2010) Ovarian cycle and embryonic development in Gammarus fossarum: Application for reproductive toxicity assessment. Environ Toxicol Chem 29:2249–2259. doi: 10.1002/etc.268

20. Brun V, Masselon C, Garin J, Dupuis A (2009) Isotope dilution strategies for absolute quantitative proteomics. J Proteomics 72:740–9. doi: 10.1016/j.jprot.2009.03.007

21. Carr SA, Abbatiello SE, Ackermann BL, Borchers C, Domon B, Deutsch EW, Grant RP, Hoofnagle AN, Hüttenhain R, Koomen JM, Liebler DC, Liu T, MacLean B, Mani DR, Mansfield E, Neubert H, Paulovich AG, Reiter L, Vitek O, Aebersold R, Anderson L, Bethem R, Blonder J, Boja E, Botelho J, Boyne M, Bradshaw RA, Burlingame AL, Chan D, Keshishian H, Kuhn E, Kinsinger C, Lee JSH, Lee S-W, Moritz R, Oses-Prieto J, Rifai N, Ritchie J, Rodriguez H, Srinivas PR, Townsend RR, Van Eyk J, Whiteley G, Wiita A, Weintraub S (2014) Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. Mol Cell Proteomics 13:907–17. doi: 10.1074/mcp.M113.036095

22. Fillâtre Y, Rondeau D, Jadas-Hécart A, Communal PY (2010) Advantages of the scheduled selected reaction monitoring algorithm in liquid chromatography/electrospray ionization tandem mass spectrometry multi-residue analysis of 242 pesticides: a comparative approach with classical selected reaction monitoring mode. Rapid Commun Mass Spectrom 24:2453–61. doi: 10.1002/rcm.4649

23. Loziuk PL, Wang J, Li Q, Sederoff RR, Chiang VL, Muddiman DC (2013) Understanding the Role of Proteolytic Digestion on Discovery and Targeted Proteomic Measurements Using Liquid Chromatography Tandem Mass Spectrometry and Design of Experiments. J Proteome Res 12:5820–5829. doi: 10.1021/pr4008442

24. Vial J, Jardy A (1999) Experimental Comparison of the Different Approaches To Estimate LOD and LOQ of an HPLC Method. Anal Chem 71:2672–2677. doi: 10.1021/ac981179n

25. González O, Blanco ME, Iriarte G, Bartolomé L, Maguregui MI, Alonso RM (2014) Bioanalytical chromatographic method validation according to current regulations, with a special focus on the non-well defined parameters limit of quantification, robustness and matrix effect. J Chromatogr A 1353:10–27. doi: 10.1016/j.chroma.2014.03.077

26. Jaffuel A, Lemoine J, Aubert C, Simon R, Léonard J-F, Gautier J-C, Pasquier O, Salvador A (2013) Optimization of liquid chromatography-multiple reaction monitoring cubed mass spectrometry assay for protein quantification: application to aquaporin-2 water channel in human urine. J Chromatogr A 1301:122–30. doi: 10.1016/j.chroma.2013.05.068

27. Mant CT, Jiang Z, Boyes BE, Hodges RS (2013) An improved approach to hydrophilic interaction chromatography of peptides: salt gradients in the presence of high isocratic acetonitrile concentrations. J Chromatogr A 1277:15–25. doi: 10.1016/j.chroma.2012.12.044

Publication nº2

Aquatic Toxicology 190 (2017) 199–209

Contents lists available at ScienceDirect

Aquatic Toxicology

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

Research Paper

Assessing the relevance of a multiplexed methodology for proteomic biomarker measurement in the invertebrate species *Gammarus fossarum*: A physiological and ecotoxicological study

D. Gouveia[a,c], A. Chaumot[a], A. Charnot[b], H. Queau[a], J. Armengaud[c], C. Almunia[c], A. Salvador[b], O. Geffard[a,*]

[a] IRSTEA, UR MALY, Laboratoire d'ecotoxicologie, centre de Lyon-Villeurbanne, F-69616 Villeurbanne, France
[b] UMR 5180, Institut des Sciences Analytiques, Université de Lyon 1, F-69100 Villeurbanne, France
[c] CEA-Marcoule, DRF/Joliot/DMTS/SPI/Li2D, Laboratory—Innovative Technologies for Detection and Diagnostics, Bagnols-sur-Ceze, F-30207, France

Abstract

Recently, a protein sequence database was built specifically for the sentinel non-model species *Gammarus fossarum* using a proteogenomics approach. A quantitative multiplexed targeted proteomics assay (using Selected Reaction Monitoring mass spectrometry) was then developed for a fast and simultaneous quantification of dozens of biomarker peptides specific of this freshwater sentinel crustacean species. In order to assess the relevance of this breakthrough methodology in ecotoxicology, the response patterns of a panel of 26 peptides reporting for 20 proteins from the *Gammarus fossarum* proteome with putative key functional roles (homeostasis, osmoregulation, nutrition, reproduction, molting,…) were recorded through male and female reproductive cycles and after exposure to environmental concentrations of cadmium and lead in laboratory-controlled conditions. Based on these results, we validated the implication of annotated vtg-like peptides in the oogenesis process, and the implication of Na^+/K^+ ATPase proteins in the organisms molt cycle. Upon heavy metal contamination, peptides belonging to proteins annotated as involved in antioxidant and detoxification functions, immunity and molting were significantly down-regulated. Overall, this multiplex assay allowed gaining relevant insights upon disruption of different main functions in the sentinel species *Gammarus fossarum*. This breakthrough methodology in ecotoxicology offers a valid and high throughput alternative to currently used protocols, paving the way for future practical applications of proteogenomics-derived protein biomarkers in chemical risk assessment and environmental monitoring.

Keywords: *Biomarkers, mass spectrometry, proteomics, Gammarus fossarum, multiplex quantitation*

Introduction

The control of water ecosystem quality in Europe is done via the Water Framework Directive (2000/60/EC), which states that disturbed aquatic systems must be returned to good states. To fulfill these requirements, the most common approaches are water chemical analysis and ecological indicators that provide information about contamination levels and community structures of aquatic systems monitored. These tools, although very useful, do not easily allow the identification of the sources of the disturbance nor the distinction between effects related to modifications of natural habitats from effects related to chemical stress. For this purpose, biological indicators, observed at the individual and sub-individual levels, are of utmost importance. Among them, biomarkers are recognized as relevant tools for diagnostic and hazard assessment of aquatic systems, in complementarity with chemical approaches and bio-indicators (Amiard-Triquet and Berthet, 2015). Substantial efforts were made during the past decades to develop biomarker-based methodologies for ecological risk assessment. In this way, active biomonitoring strategies, as developed in *Gammarus fossarum*, allows the establishment of reference values integrating the natural variability of biomarkers, leading to

accurate interpretation in terms of contamination and/or toxicity levels (Besse et al., 2013; Coulaud et al., 2011; Jubeaux et al., 2012b; Lacaze et al., 2011b; Xuereb et al., 2009). Nonetheless, the use of biomarkers in field survey is still scarce. Few absolute quantification methods are available, due to the scarce genomic and/or proteomic data. Moreover, each biomarker measurement involves a unique protocol (e.g. GST, CAT or ChE (Jemec et al., 2010)), leading to very expensive biomonitoring strategies in time, cost and biological samples. It has been established that monitoring health status of organisms using several sensitive biomarkers covering a broad range of effects is necessary (Connon et al., 2012; Dagnino et al., 2007). Hence, the development of multi-biomarker strategies allowing the simultaneous monitoring of a wide range of biological responses is required for improving environmental hazard assessment.

Proteins, being the molecular effectors of biological processes, are pertinent candidate biomarkers because their modulation is more likely to be associated with the effective impact on the physiology of the sentinel organism (Trapp et al., 2014a). In ecotoxicology, the majority of proteomic studies are based on the classical approach, in which proteomes from a control and contaminated conditions are resolved by 2D-PAGE, and proteins differentially expressed are then identified by mass spectrometry (reviewed in (Sanchez et al., 2011)). "Protein fingerprints" or "Protein patterns" are then established for a certain contaminant or stress (e.g. (Sharma et al., 2004; Silvestre et al., 2006)). However, these 2D-gel-based approaches are not adapted for protein quantification due to several limitations related to gel-to-gel variation, overlapping spots and low abundant proteins (Trapp et al., 2014a). Gel-free methodologies based on the shotgun proteomics strategy (Spectral Count, ICAT, iTRAQ, SILAC) are available as alternatives to gel-based protein quantification (Martyniuk et al., 2012; Martyniuk and Denslow, 2009), each with some advantages and limitations. However, for absolute quantification of a given set of proteins, the method of choice is the mass spectrometric analysis using the selected reaction monitoring (SRM) mode (Gallien et al., 2011). SRM is a LC-MS-based technique for the quantitation of one or more proteins of interest (multiplexing capability) in complex biological samples. SRM is highly sensitive and uses protein-specific reporter peptides to quantify the amount of the whole protein in the samples. When using isotopically-labelled internal standards, absolute quantification can be achieved (Anderson and Hunter, 2006). In invertebrate ecotoxicology, it was already applied as proof of concept for the quantitation of a protein, one vitellogenin (vtg), in the amphipod crustacean *Gammarus fossarum* (Jubeaux et al., 2012a; Simon et al., 2010b). This was the first study to report a methodology for the absolute quantification of a protein from the whole-body of an aquatic invertebrate.

Due to the few sequenced genomes of aquatic invertebrates, sequence-based identification of proteins is very difficult, and proteins that exhibit the most significant changes in expression are not always identified (Sanchez et al., 2011). Recent technological advances in high-throughput "omics" methodologies, in which thousands of genes, proteins or metabolites can be assessed simultaneously, provided researchers new tools for developing biomarkers (Benninghoff, 2007). Recently, an approach called "proteogenomics" emerged as a relevant strategy for the discovery of proteins in non-model organisms (reviewed in (Armengaud et al., 2014; Renuse et al., 2011)). With the combination of high-throughput mRNA sequencing (RNA-seq) and tandem mass spectrometry proteomics, Trapp et al. (Trapp et al., 2014b) created a specific database consisting of 1,873 experimentally validated proteins of the amphipod crustacean *Gammarus fossarum*, a non-model sentinel species for continental water biomonitoring. With these new resources, *Gammarus fossarum* appeared as a good case study to setup an innovative multiplexed SRM-based approach, which could improve specific identification and simultaneous quantification of several proteins of interest in invertebrate species (Charnot et al., 2017). In this first study a list of peptide candidates was built based on their putative function determined by automatic bioinformatics annotation, seeking to cover different important physiological functions potentially subject to disruption by exposure to contaminants. This list included 172 peptides reporting for 55 proteins previously seen either as being sex-specific or whose annotations suggested roles in immunity, homeostasis, detoxification and defense mechanisms (Trapp et al., 2014b). The multiplexed SRM mass spectrometry strategy was then designed in order to quantify peptides from this initial list, using *G. fossarum* whole-body. Method validation led to the possibility of quantifying simultaneously a maximum of 71 reporter peptides from 40 proteins of interest in *G. fossarum* (Charnot et al., 2017).

Based on this analytical methodology, the aim of the present study was to appraise the use of the quantitation of selected peptides as reporters of major physiological functions presumed from the annotation of associated proteins. By this mean we aimed to assess the interest of these peptides as

biomarkers for ecotoxicological assessment. To achieve this, changes in peptide levels were studied through male and female reproductive cycles (regarding male, female, molt-related and osmoregulation annotated peptides), and during exposure to two metal contaminants (regarding immunity biomarkers and antioxidant/detoxification enzymes). The potential applications offered by this methodological breakthrough for protein biomarker measurements in the field of ecotoxicology and biomonitoring are discussed.

Materials and Methods

Reagents and chemicals

Acetonitrile, methanol and water (LC-Ms grade) were obtained from Fisher Scientist (Strasbourg, France). Dithiothreitol (DTT), iodoacetamide (IAM), formic acid (FA) (LC–MS grade), trypsin (treated TCPK from bovine Pancreas and not treated type 1X-S from porcine pancreas), urea and TRIS, EDTA, Triton X, sodium chloride, leupeptin and aprotinin were purchased from Sigma–Aldrich (St Quentin-Fallavier, France). Isotopically labelled peptides containing either a C terminal [$^{13}C_6$, $^{15}N_2$] lysine or arginine were synthesized by Fisher Scientist (Strasbourg, France) (purity > 97%). Absolute ethanol and ethylic ether were obtained from Carlo Erba (Val de Reuil, France).

Sampling and maintenance of organisms

Gammarids were collected from the Bourbre River in France and acclimatized to laboratory conditions as previously described (Jubeaux et al., 2012c). Organisms were collected by kick sampling using a net, and quickly transported to the laboratory in plastic buckets containing freshwater from the station. In the laboratory, organisms were kept in 30 L tanks continuously supplied with drilled groundwater adjusted to the conductivity of the sampling site (600µS/cm) and under constant aeration for at least 10 days. A 16/8h light/dark photoperiod was maintained and the temperature was kept at 12±1°C. Organisms were fed *ad libitum* with alder leaves (*Alnus glutinosa*), previously conditioned for 6±1 days in water.

Selection and exposure of organisms

In order to validate that the candidate peptides report for proteins implied in the different functions expected from automatic bioinformatic annotation, we monitored their concentrations during the male and female reproductive cycles, and after a heavy metal exposure.

Female reproductive cycle

Five organisms from each stage of the molt cycle (B, C1, C2, D1 and D2 – described in (Geffard et al., 2010)) were selected, weighed and immediately frozen in liquid nitrogen for storage at -80ºC until the Mass Spectrometry (MS) analysis.

Spermatogenesis

The concentrations of our target peptides were determined at five different spermatogenesis stages (described in Lacaze et al. (Lacaze et al., 2011a)): days 1, 3, 5 and 7 after copulation. For each stage, five organisms were selected, weighed and immediately frozen in liquid nitrogen for storage at -80ºC until the MS analysis.

Laboratory exposure to toxic compounds

Gammarids were exposed during 7 and 21 days to environmentally relevant concentrations of two heavy metal contaminants – cadmium (Cd 2µg/L) and lead (Pb 10µg/L). These concentrations were chosen based on previous experiments from our research team, in which Cd (Dayras et al., 2017; Geffard et al., 2010; Issartel et al., 2010; Trapp et al., 2015) and Pb (unpublished data) induced toxic but non-lethal physiological effects in *Gammarus fossarum*. They were placed in nine 500mL plastic beakers (7 couples per beaker, three beakers per condition), and continuous exposure conditions were the same as in the previously described and validated method (Felten et al., 2008; Geffard et al., 2010; Pellet et al., 2009; Vigneron et al., 2015). Briefly, stock solutions were prepared in milliQ water at concentrations of 37.6 and 130.6 mg/L for cadmium and lead, respectively. The contaminated media was obtained by adding 531 and 770 µL of stock solutions of cadmium and lead, respectively, in 10L of uncontaminated drilled ground water. The constant renewal of the media was achieved by means of a peristaltic pump at a rate of 1L per day and per beaker. At the end of the exposure, male and female organisms were individually sampled, rapidly weighed, frozen in liquid nitrogen and stored at -80°C until the MS analysis.

Protein extraction and trypsin digestion

The procedure used for total protein extraction was adapted from Simon et al. (Simon et al., 2010a). Whole-body organisms were homogenized in Tris Buffer (Tris 50mM, 100mM NaCl, 0.01 mM EDTA, 0.1% v/v Triton X-100) buffered at pH 7.8 and complemented with 10 µg/L of each leupeptin and aprotinin, with a bead mill homogenizer. For each organism, the volume of buffer was adapted according to their weight (25 µL buffer/mg male;

50 µL buffer/mg female). The homogenates were centrifuged at 10,000 g at 4 °C for 15 min, and 250 µL of the clear resulting supernatant were collected to new tubes. A volume of 750 µL of an ethanol/diethyl ether delipidating solution (1:1 v/v) was added to the supernatant, and the mixture was vortexed and incubated on ice for 10 min. After a 10 min centrifugation at 10,000 g, the resulting supernatants were removed, and the pellets resuspended in 250 µL Tris Buffer.

The protein extracts were then treated with 3 mL of ammonium bicarbonate (AMBIC) 50mM and 362 µL of dithiothreitol (DTT) 150 mM for 40 min at 60 °C. After cooling to room temperature, samples were incubated in the dark with iodoacetamide (final concentration of 15 mM), during 40 min at room temperature. A volume of 150 µL of a 2 mg/ml TPCK treated trypsin solution prepared in 50 mM of AMBIC was added, and the samples were incubated for 60 min at 37 °C. A volume of 20µL of formic acid (FA) was added to stop trypsin reaction. A volume of 10 µL of the solution of heavy peptides at 1 µg/ml was added to each digest. Each mixture was then purified with an Oasis HLB 3 cc solid phase extraction columns (60 mg). After peptide elution with 1 mL of methanol/0.5 % FA, 100 µL of a glycerol 10%/methanol were added, and the samples evaporated under a flow of nitrogen. The pellet was then resuspended in 90 µL of a H_2O/ACN (90/10) + 0.1% FA solution.

Multiplexed LC-SRM/MS analysis and absolute quantification

LC–MS/MS analysis was performed on an HP1200 series HPLC device (Agilent Technologies, Waldbronn, Germany) coupled to a QTRAP® 4000 LC/MS/MS System hybrid triple quadrupole/linear ion trap mass spectrometer (Applied Biosystems/ MDS Analytical Technologies, Foster City, CA, USA) equipped with a Turbo VTM ion source. The LC separation of the 20µL injected sample was carried out on an Xbridge C_{18} column (100mm×2.1mm, particle size 3.5 µm) from Waters (Milford, MA, USA). Elution was performed at a flow rate of 300µL/min with water containing 0.1% (v/v) formic acid as eluent A and acetonitrile containing 0.1% (v/v) formic acid as eluent B, employing a linear gradient from 2% B to 33% B in 19 min, followed by a second linear gradient from 33% B to 64% B in 6 min. Then, column washing and re-equilibration was performed for 6 min. The injection duty cycle was 35 min, taking into account the column equilibration time. Instrument control, data acquisition and processing were performed using the Analyst 1.5 software. The mass spectrometer was initially tuned and calibrated using polypropylene glycol, reserpine and Agilent Tuning Mix (all Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. MS analysis was carried out in positive ionization mode using an ion spray voltage of 5500 V. The nebulizer and the curtain gas flows were set at 50 psi using nitrogen. The Turbo V™ ion source was set at 550 ◦C with the auxiliary gas flow (nitrogen) set at 40 psi. The SRM transitions (previously selected in (Charnot et al., 2017)) were monitored and acquired at unit resolution in Q1 and Q3, with a dwell time of 10 ms used for each transition, to obtain 10 data points per chromatographic peak minimum.

In order to perform a quality control of the large number of SRM runs performed, spectra quality and reproducibility between samples were evaluated using the Skyline v3.1 software (MacCoss Lab Software, USA). This analysis comprised manual peak picking for reliable peak identification, verification of retention time between samples, and transition refinement regarding quantification: check if the three transitions selected for each peptide were of good quality and reproducible over the samples (constant transition intensity and contribution to the total area of the peak in all runs). After this qualitative analysis, 26 peptides reporting for 20 proteins were kept for quantification and data analysis.

Reporter peptides were separated into 4 main groups, according to their representative putative protein functions: yolk-related, molt-related, osmoregulation-related Na^+/K^+ ATPases, and general ecotoxicological biomarkers (Table 1). Some peptides included here can also be immune-related (peptides from proteins annotated as prophenoloxidase, transglutaminase and hemolectin). Yolk peptides report for 6 different female specific proteins. These proteins were originated from different sub-groups of the large lipid transfer protein superfamily, and were previously characterized by shotgun proteomics of reproductive systems. They were described as having a strong sexual dimorphism (Trapp et al., 2014b) and vitellogenin function in *Gammarus fossarum* (Trapp et al., 2016), playing a crucial role in yolk formation and oogenesis. During discussion, these peptides will be referred to as "vtg-like peptides", although only 3 of these 9 peptides were blast-annotated as vitellogenin.

Absolute quantification was performed using the ratio between the peak areas of peptides in the sample and the corresponding heavy labeled peptide (with known concentration of 1000ng/ml).

Table 1 – List of biomarker peptides simultaneously quantified in the SRM experiments from this study.

Peptide Modified Sequence	Protein ID	Functional Annotation	Peptide MW (Da)	Functional Group
HIEIFSPITK	64	Vitellogenin	1184,4	Female yolk proteins
VVPSLSAEDTLSQR	276	Hemolymph clottable protein	1501,6	
HAEFSVNPPLDSTQAVK	276		1839,9	
IIYPAEALTIVIEK	276		1572,9	
IFNVLQPIAESK	17046	Clottable protein 2	1358,8	
ISPLINSPSDLPK	39606	Clotting protein precursor	1380,8	
TLGALELDVFSEAGK	200426	Vitellogenin	1549,8	
ELTSAAEVVSSLLK	200426		1446,8	
AAIETAFVNHLK	206469	Clotting protein precursor	1313,7	
FVGLISLIDPPR	32234	Na+K+ ATPase α1 subunit	1326,8	Osmoregulation
VIMVTGDHPITAK	110907	Na+/K+ ATPase	1381,8	
LGAIVAVTGDGVNDSPALK	110907		1797,0	
NLAFFSTNAVEGTAR	209438	Na+/K+ ATPase	1597,8	
LQTNPDTGLSTAEAR	209438		1573,8	
GIDIIGDAFEADR	2562	Prophenoloxidase	1391,7	Molt-related
APILEGYFSK	2562		1124,6	
ILEDFVDVFNR	100255	Cytochrome P450 enzyme, CYP4C39	1366,7	
AFWGSLPLR	144144	JHE-like carboxylesterase 1	1046,6	
DTDWVGYEDPDSTAIK	212968	Chitinase 1 precursor	1811,8	
GTLAVIPVQNR	1917	Transglutaminase	1167,7	General biomarkers
VLAVDILAK	7169	Transglutaminase	941,6	
VEC[+57]IAGFILPLEFK	11145	Hemolectin	1578,9	
ADPALGQAIQER	110912	Catalase	1268,7	
FSESSAILR	40079	Gluthation-S-transferase	1009,5	
LSAWLAAC[+57]K	142711	Gluthation-S-transferase	962,5	
ELFDFADAHR	213317	Cellulase	1220,6	

Statistical analysis

Results are presented individually in the form of heatmaps (heatmap.2 function in R software), or as the mean of five biological replicate samples with the associated standard deviations. Statistical analyses for testing differences in peptide concentrations between conditions were performed with the GraphPad Prism Version 7.02 software, using non-parametric Mann-Whitney t-tests. Significant differences were accepted at $p<0.05$.

Results and discussion

The objective of this study was to validate the proof of concept of a new multiplexed methodology for the simultaneous quantification of a large number of biomarkers of interest in *Gammarus fossarum* whole-body. To accomplish this, the concentration of peptides selected according to the functional annotation of the protein they belong to were monitored through different physiological situations: assessment of their levels during the organisms' reproductive process, after a food privation stress and after heavy metal exposure. Molecular responses obtained through reproductive and contamination processes were correlated with our physiological knowledge of the species, used to propose biomarker peptides, and to infer about the pertinence of pursuing future applications in biomonitoring.

Peptide profiles through molt cycle and gametogenesis in males and females

The concentrations obtained after peptide quantification for the entire set of proteins in each female organism during the different stages of the molt cycle are shown in Table A.1. These data are represented via a color heat map (Figure 1) that provides a global overview of the results. Individual organisms are represented in columns (five biological replicates for each reproductive stage) and each signature peptide in lines. In general there's a slight trend to higher protein concentrations for females in the pre-molt phase, especially for two specific protein sets: yolk

proteins and proteins annotated as Na^+/K^+ ATPase. The nine signature peptides of the vtg-like proteins have a gradual increase in concentration throughout the reproductive cycle. In Figure 2 are shown the mean levels of each vtg-like peptide during the different conditions. In general, the sharpest variation of vtg occurs between the C2 and D1 stage, which is in accordance with the increase in follicular size and oocyte volume during the C2 stage previously stated by Geffard et al. 2010 (Geffard et al., 2010). In *Gammarus fossarum* females, oogenesis occurs simultaneously with the molting cycle, and is characterized by the progressive accumulation of vitellogenin-like proteins in oocytes during secondary vitellogenesis (normally in the C2 molt stage). Vitellogenin is the precursor of major egg yolk protein and is essential to provide the energy required for embryonic growth.

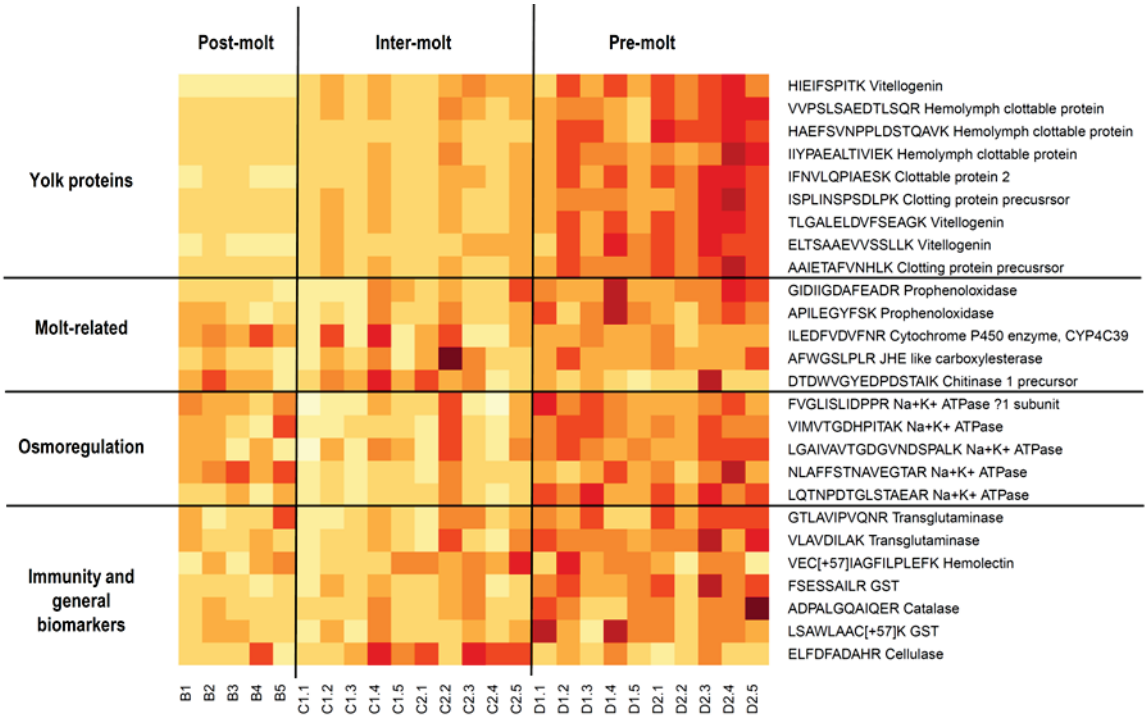


Figure 1: Heatmap of the quantification results in pmol peptide/mg organism of the entire set of peptides during five stages of the female molt cycle. For each molt cycle stage (B, C1, C2, D1 and D2) there are 5 individuals, represented in the y axis. The color gradient ranges from yellow (lowest concentrations) to red (highest concentrations) and signature peptides are clustered in accordance with their putative function: yolk peptides, molt-related peptides, osmoregulation-related Na^+K^+ ATPases, and general ecotoxicological biomarkers.

The quantification results also show that different vtg-like reporter peptides are present at different levels, with maximum concentrations ranging between 1.16 pmol/mg for peptide HAEFSVNPPLDSTQAVK and 33.6 pmol/mg for peptide ISPLINSPSDLPK. The different levels observed between different peptides mean that some vtg proteins are more expressed than others, surely related to their final role in embryo development. During the experiment that lead to the identification of these proteins in *Gammarus fossarum* (Trapp et al., 2016), yolk proteins were also quantified by means of shotgun proteomic analysis. In this previous experiment, data were collected from oocytes and embryos, and quantification was performed using the label free spectral count approach. Despite being different matrices and quantification procedures, the concentration differences obtained for the different yolk proteins are still similar between this experiment and our study. With the exception of the 17046 protein (the most intense observed in Trapp et al., the third most intense in our study), the other five yolk proteins follow exactly the same intensity order, with the protein 39606 being the most intense and protein 64 the less intense (Table A.2). This demonstrates the methodologies high accuracy regarding the quantitation of the most intense proteins.

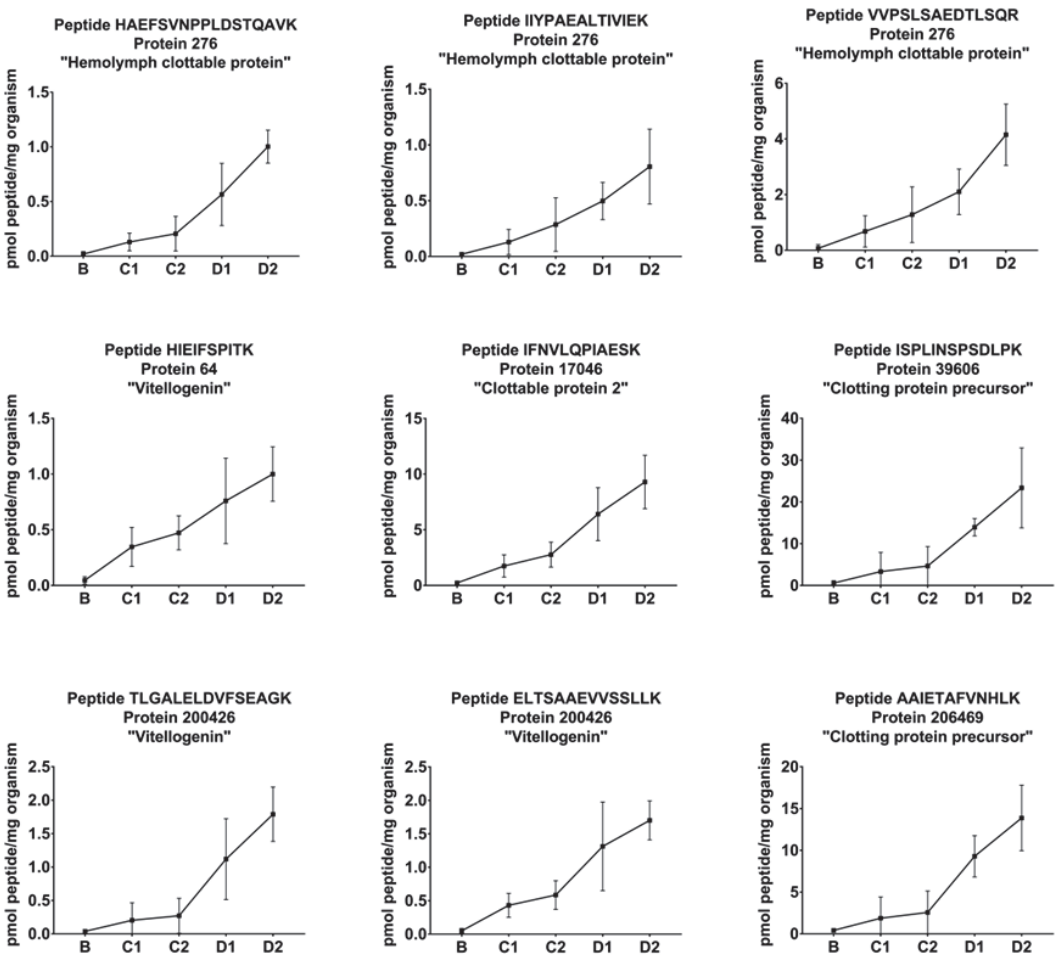


Figure 2: Quantification results in pmol peptide/mg organism of the nine vtg-signature peptides of *G. fossarum*.

Concerning molt-related and immunity/general proteins, it is possible to observe other interesting patterns in Figure 1, especially for peptides GIDIIGDAFEADR and APILEGYFSK derived from the protein annotated as Prophenoloxidase, and peptides GTLAVIPVQNR and VLAVDILAK derived from the protein annotated as Transglutaminase. The concentrations of these peptides are molt stage-dependent, having higher concentrations in the pre-molt phase.

Prophenoloxidase is the precursor of the phenoloxidase (PO) enzyme, responsible for melanization (proPO system). The activation of PO leads to the synthesis of melanin, substance involved in sclerotization, wound healing of the cuticle and defense reactions against invading pathogens (Sritunyalucksana and Soderhall, 2000). In the literature there are several studies that show that these proteins are involved in crustaceans' molt cycle, probably in cuticle formation. For example, a study in the crustacean decapod *M. rosenbergii* (Liu et al., 2006) has shown that the expression of the PPO gene reaches a peak during the post-molt stages, and is significantly decreased during the inter- and pre-molt stages. Yeh et al. (Yeh et al., 2009) sequenced two PPO genes in the shrimp *L. vannamei* and observed that both genes had significant modulations during a pre-molt and post-molt stages, indicating a role in ecdysis.

Transglutaminases (TG) are present mainly in hemocytes, and play a crucial role in the immune responses of crustaceans (Fan et al., 2013), being an active part in the coagulation process. There's not a lot of information about the role of these proteins in the molting cycle of crustaceans, but the study by Liu et al. (Liu et al., 2011) showed that both transcript levels and enzymatic activity of TG were modulated during the molt cycle of *M. rosenbergii*, with a continuous increase followed by a sudden decrease right before ecdysis.

Peptides from proteins annotated as Na^+/K^+ ATPase (osmoregulation) were also modulated throughout the reproductive cycle. Na^+/K^+ ATPase proteins are present in the plasma membranes of animal cells and are responsible for pumping Na^+ ions out of the cell while pumping K^+ into the cell, in an energy dispendious process (ATP hydrolysis). The activity of these enzymes is essential for crustaceans and plays a crucial role in the mechanisms of osmoregulation by which these organisms regulate osmotic and ionic concentrations (Lucu and Towle, 2003). In Figure 3 are represented the mean values of five organisms obtained for the different Na^+/K^+ ATPase proteins during five stages of the female molt cycle. In general, after a slight decrease during inter-molt, there's an increase in the concentration of these pumps in the pre-molt stages. As reviewed in (Charmantier, 1998), osmolality varies throughout the molting cycle of crustaceans, with a more pronounced increase in the pre-molting phase. This ensures the entry of water into the body and an increase in volume at the time of ecdysis.

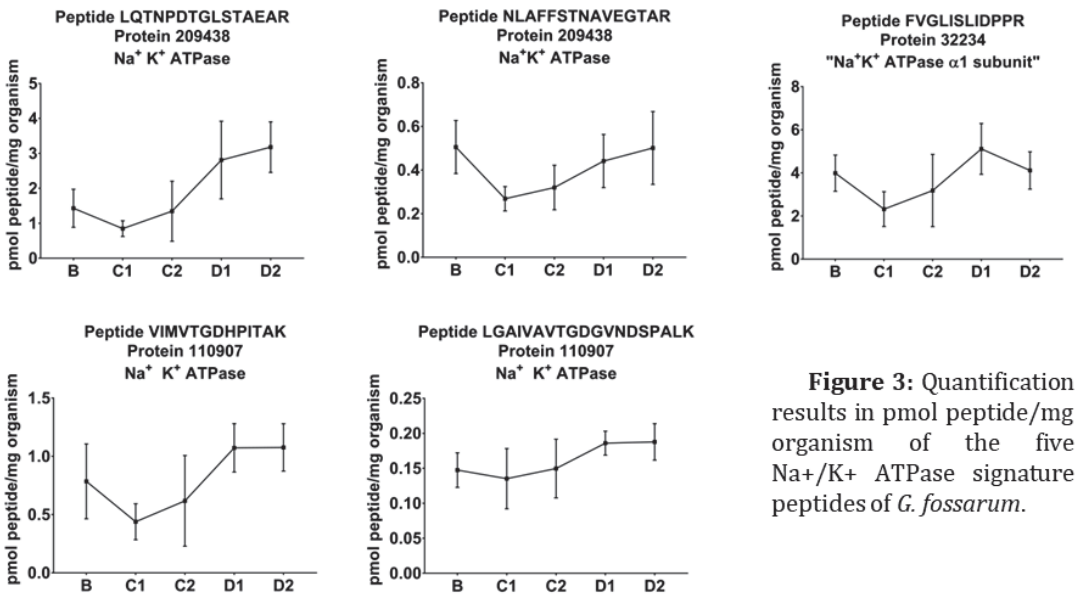


Figure 3: Quantification results in pmol peptide/mg organism of the five Na+/K+ ATPase signature peptides of *G. fossarum*.

Our results follow this pattern, with higher concentrations of the sodium pump enzymes right before ecdysis as a consequence of the effort of the organism to maintain its osmotic equilibrium during the entry of water into the body. Different profiles between proteins with the same annotation could be due to their specific localization in the organism. It is known that in crustaceans the Na^+/K^+ ATPase activity is much higher in gills than in other tissues (Issartel et al., 2010; Li et al., 2015; Spencer et al., 1979), so we can hypothesize that the most intense proteins will be the ones localized in the gills (in particular the protein 32234).

Peptide ELFDFADAHR from the protein 213317 annotated as cellulase presented a different trend compared to that of the latter, as shown in Figure 4. Despite not being statistical significant, we can observe a trend of increasing levels of cellulase during the inter-molt stage followed by a decrease in the pre-molting phase. The increase trend is likely related to the fact that females have finished the molting process and must accumulate, before forming the amplexus (guarding behavior by male), the necessary energy for the high-energetic processes of oogenesis and ecdysis. The decrease trend in pre-molt might be related to the fact that once the female couples with a male they have a more restricted access to food. Charron et al. (Charron et al., 2014) observed that the activities of three digestive enzymes diminished significantly during the pre-molt stage, along with an increase in their energy reserves (glycogen, lipids and proteins). Other studies also stated decreases in the activities of digestive enzymes during the pre-molt stages in other arthropods (Fernandez Gimenez et al., 2002; Fernández et al., 1997; Weidlich et al., 2013).

The multiplex quantitation was also applied with isolated male organisms during seven days of the spermatogenesis process. The concentrations obtained for the entire set of peptides for each male organism of the spermatogenesis experiment are shown in Table A.3. Data from the spermatogenesis experiment is represented via a color heat map (Figure A.1) with the same

structure as before. Globally, unlike the data obtained from the female reproductive cycle, there are no set of proteins that we could identify as having peculiar patterns during these two experiments. Along with some individual variability, almost every protein levels remained stable between the four different time-points studied (days 1, 3, 5 and 7).

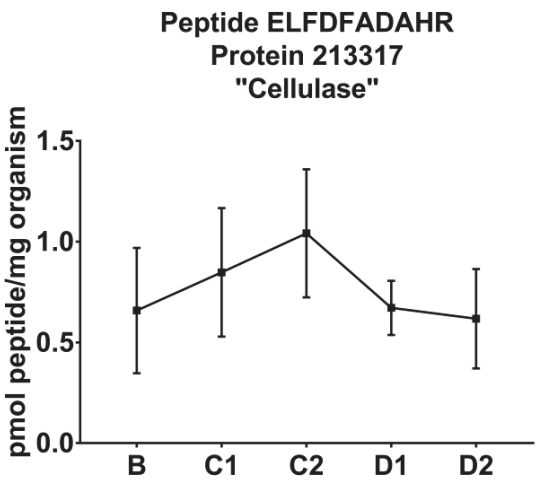


Figure 4: Quantification results in pmol peptide/mg organism for the cellulase signature peptide during the female reproductive cycle of *G. fossarum*.

Laboratory exposures to Cd and Pb

In order to check how our ensemble of biomarker candidates would respond to a chemical contamination stress, a laboratory-controlled exposure to two heavy metals (Cd and Pb) was conducted. The contaminants were chosen based on previous data obtained in crustacean species. With Cd exposures, Wang et al. 2009 (Wang and Wang, 2009) have observed modulations in some detoxification and antioxidant enzymes of *Tigriopus japonicus mori*. Issartel 2010 (Issartel et al., 2010) observed important changes in the haemolymph osmolality and Na^+K^+ pumps in *Gammarus fossarum*. Geffard 2010 (Geffard et al., 2010) and Yang 2015 (Yang et al., 2015) stated that Cd has an impact in vitellogenesis in the freshwater amphipod *Gammarus fossarum* and in the crab *Sinopotamon henanense,* respectively. Increased activities of Glutathion-S-transferase (GST) and catalase (CAT) were observed after dietary exposure to Cd and Pb in the green garden snail, *Cantareus apertus* (Mleiki et al., 2015). Hariharan et al. (Hariharan et al., 2012) demonstrated that alterations in antioxidant enzymes reflect the consequences of heavy metal exposure (Pb and Zn) in *P. monodon*.

In our experiment, we followed the protocol test proposed by Geffard 2010 for reproductive toxicity in *G. fossarum*: coupled male and female organisms (with females in the final D2 molting stage) were exposed to Cd 2µg/L and Pb 10µg/L during two different time points: 7 and 21 days. By choosing this methodology, we ensured that the contamination in females occurred during the new molt cycle, follicular maturation and embryonic development (Geffard et al., 2010). In males, the first sampling time (7d) occurred during spermatogenesis which had begun since 3 to 4 days (on a period of 8 to 10 days (Lacaze et al., 2011a)), whereas the second time covered the entire male spermatogenesis process, with mature organisms.

Molecular effects of contamination

Females

The concentrations obtained for the entire set of protein biomarkers in female organisms after the 7 and 21-day exposure are shown in Tables A.4 and A.5, respectively. In general, no significant differences from contaminated and control conditions were observed after the 7-day experiment, so the discussion is focused on the results of the 21-day experiment. Exceptions are cited during the text.

Figure 5A represents data collected from female organisms. This data confirms the profiles obtained for the vtg-like peptides during the female reproductive cycle, being evident the difference between the levels of these peptides in females at day-7 (B molting stage) and day-21 (C2 molting stage) of the experiment.

Comparing the day-21 controls and contaminated females, we can observe that both contaminants have modulated vtg-like peptides in some females: female F6 from the Cd condition and females F1, F3 and F14 from the Pb condition have much lower vtg levels compared to the other females. This could indicate that either the molting cycle of these females was delayed, or the oogenesis process was blocked by the contaminants. Previous studies had also observed that Cd inhibited molting (Luo et al., 2015; Rodríguez Moreno et al., 2003) and/or slow down vitellogenesis and ovarian growth in crabs (Revathi et al., 2011; Rodríguez et al., 2000; Yang et al., 2015). Pieces of information about physiological effects of Pb in crustacean species are scarce, but delays in the molting cycle were already observed in chironomids (Vermeulen et al., 2000), although the specific mechanism of action is not yet understood.

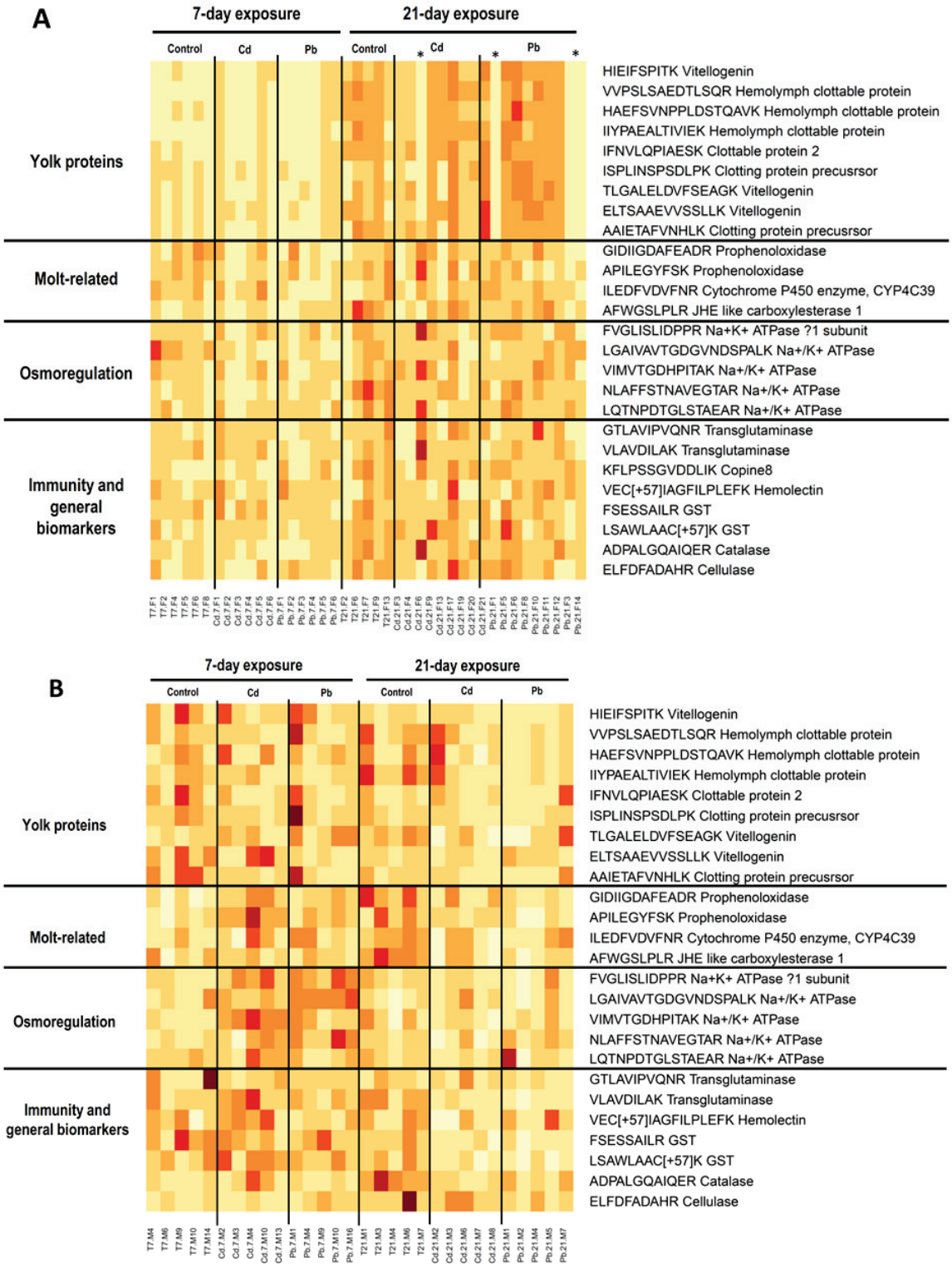


Figure 5: Heatmap of the quantification results of the entire set of peptides for the contamination experiment in **female (A) and male (B) organisms**. For each condition (T7, Cd7, Pb7, T21, Cd21, Pb21) there are 5 individuals, represented in the y axis. The color gradient ranges from yellow (lowest concentrations) to red (highest concentrations) and signature peptides are clustered in accordance with their putative function: yolk peptides, molt-related peptides, osmoregulation-related Na^+K^+ ATPases, and general ecotoxicological biomarkers. T7 – seven day control, Cd7 – seven day cadmium contamination, Pb7 – seven day lead contamination, T21 – twenty-one day control, Cd21 - twenty-one day cadmium contamination, Pb21 - twenty-one day lead contamination.

Males

The concentrations obtained for the entire set of protein biomarkers in male organisms after the 7 and 21-day exposure are shown in detail in Tables A.4 and A.5, and are represented as a heatmap in Figure 5B. Some peptides presented significant down-regulations at day 21 in contaminated organisms (Figure 6).

Peptide LSAWLAAC[+57]K (reporter for protein 142711; *p* value=0.0159 for both Cd and Pb) and peptide FSESSAILR (reporter for protein 40079; statistically not validated as different but with *p* values of 0.0510 and 0.0542 very near from the threshold) were down-regulated by both Cd and Pb. Both proteins are annotated as GST, an enzyme that catalyzes the conjugation of glutathione to xenobiotics and their metabolites, and commonly employed as a biomarker of oxidative stress and detoxification. Although normally we would expect an increase in the concentration of this enzyme due to the presence of metallic contaminants, which could induce an oxidative stress, our results show another trend for these two potential GSTs. In males exposed for seven days there was no difference in concentration between conditions for both proteins (Figure A.2). However, after the 21-day exposure, peptide LSAWLAAC[+57]K had significant lower concentrations in both contaminated conditions when compared to the control condition (Figure 6). Although not statistically different, peptide FSESSAILR also had a decrease in concentration in contaminated conditions. In the literature the majority of the studies observe increases in the activity (Lobato et al., 2013; Marinković et al., 2012; Mleiki et al., 2015) or gene expression (Arockiaraj et al., 2014; Chaurasia et al., 2016; Marinković et al., 2012) of this enzyme, with the same contaminants but in other crustacean species. However, some studies also observed different responses. Wang et al. (Wang and Wang, 2009), for example, exposed the copepod Tigriopus japonicus to different concentrations of Cd, and observed an inhibition of GST activity after one day of exposure, followed by an increase at day 7 and a return to the control level at day 12 of exposure. Jemec et al. (Jemec et al., 2007) found no differences in exposed *D. magna* to Cd. This biomarker is therefore probably very dependent of the species and nature of toxicants used. Moreover, there are numerous structurally diverse classes of GSTs whose involvement in immune defense could be different in the same or between species. Finally, the response dynamic is dependent of the level of exposure and duration of the species life and reproduction cycles. Lee et al. (Lee et al., 2008), for example, studied the responses from 10 different GSTs after Cd contamination and observed that these proteins responded differently to several Cd concentrations.

Peptide ADPALGQAIQER, reporter protein 110912 annotated as another oxidative stress related enzyme, CAT, also presented significant down-regulations in both exposed conditions. This enzyme is responsible for the decomposition of hydrogen peroxide to water and oxygen, an important process in cell defense against oxidative damage by reactive oxygen species. Regularly employed as a biomarker of oxidative stress, catalase activity and transcripts have shown to be modulated by a wide range of pollutants (Auguste et al., 2016; Barros et al., 2017; Ulm et al., 2015; Yuan et al., 2015). Peptides AFWGSLPLR, GIDIIGDAFEADR and VLAVDILAK (reporters for proteins annotated as JHE-like carboxylesterase, PPO and TG, respectively) were also modulated after contamination. Annotations of these peptides suggest a role in reproduction, molting and immunity processes. JHE-like carboxylesterase is related to the degradation of methyl farnesoate, the crustacean hormone implied in molting, control of oogenesis, and possibly in testicular development (Lee et al., 2011; Reddy et al., 2004). Although the metabolism of this hormone is still not clear in crustaceans, it has many similarities with the metabolism of the juvenile hormone in insects. This peptide was significantly down-regulated by Pb after the 21-day contamination 5 (*p* value=0.079). Although not statistically different (*p* value=0.0556), a decrease in the concentration of this peptide is also observed for the Cd-exposed condition.

JHE-like carboxylesterase is not commonly used as a biomarker in ecotoxicology, but its potential as a biomarker for reproductive impairments in crustaceans begins to gain strength as highlighted by several studies over these last years. Lee et al. (Lee et al., 2011) observed that a transcript of JHE-like carboxylesterase was up-regulated after eyestalk ablation of *Pandalopsis japonica* (technique for inducing molt in crustaceans) and Houde et al. (Houde et al., 2015) also observed an up-regulation of this biomarker after an exposure to vynil chloride in *Daphnia magna*. Another study (Xu et al., 2016) presented a down-regulation of this enzyme after transcriptomic analysis in precocious juveniles of the Chinese mitten crab *Eriocheir sinensis*.

In the 7-day contamination experiment, concentrations of the PPO-reporter peptides were slightly higher in both contaminated conditions (Figure A.2). Although not statistically different, it could mean an activation of the proPO defense mechanism. In contrast, after 21 days of

contamination, a clear decrease in the concentration of peptide GIDIIGDAFEADR is observed (Figure 6), especially in the Pb contaminated condition (*p* value=0.0159). This decrease is probably due either to an incapability of the stressed organism to fight contamination, or to a direct influence of Pb in molting-related molecules, as seen for JHE-like carboxylesterase peptide AFWGSLPLR, and that could lead to delays in the molting cycle.

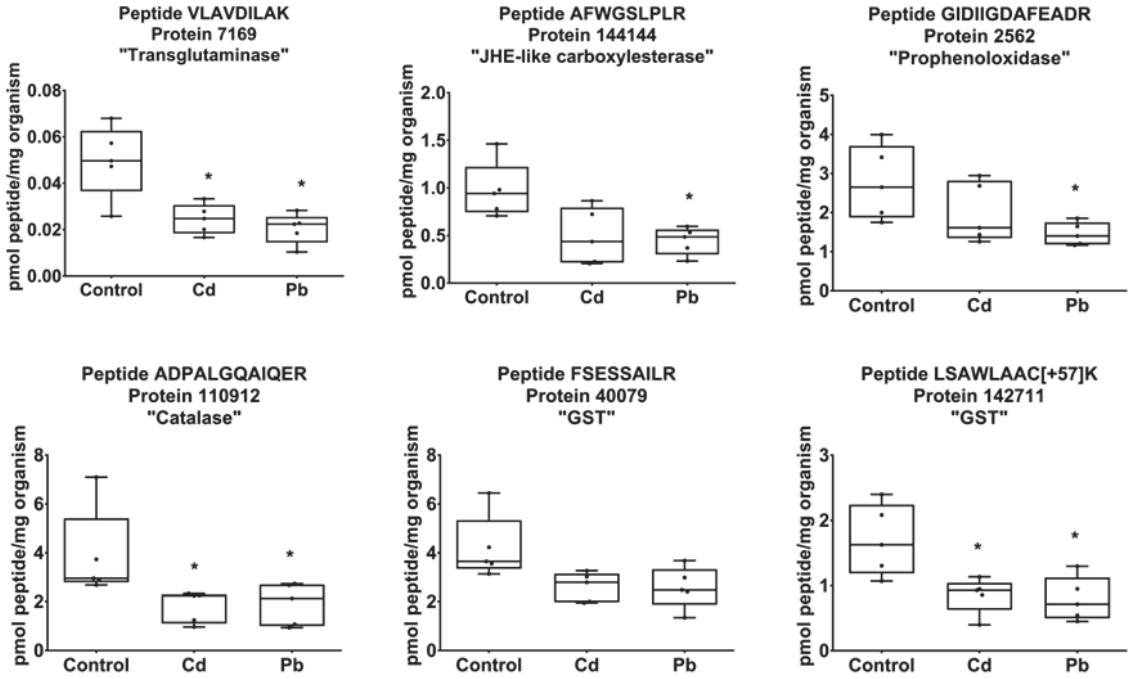


Figure 6: Down-regulated peptides (pmol peptide/mg of organism) after the 21-d contamination in **male organisms**.

Finally, peptide VLAVDILAK, annotated as transglutaminase, was down-regulated by both Cd and Pb (*p* values of 0.0317 and 0.0159). As the proPO, this TG-annotated peptide was also present at normal levels after 7 days of contamination, but down-regulated by both metals at day 21 (Figure 6). The fact that these two hemocyte-related proteins are down-regulated could be related to a possible decline in circulating hemocytes, and consequently in the immune capacities of the organism.

Conclusion

In this study, the combination of a mass spectrometry targeted SRM-based approach with the physiological knowledge of the sentinel species in ecotoxicology allowed to propose a multiplexed biomarker quantification strategy taking into account the biological relevance of the recorded response pattern. This high throughput and highly-sensitive methodology allowed for the detection of low amounts of peptide (in the order of 0.02 pmol peptide / mg organism) and the simultaneously monitoring of the concentrations of 26 peptides reporting for 20 biomarker proteins from the whole-body of a sentinel non-model species as starting biological material.

Hence, this methodology appears as a relevant tool for specific detection and quantification of a panel of protein biomarkers in order to assess the organism health status. Constraints as the lack of molecular information for non-model species and the multitude of protocols employed for the analysis of each individual biomarker were surpassed using our innovative strategy. With the alliance of proteogenomics for the construction of a large transcript and protein database, targeted proteomics for precise quantitation of biomarkers, and physiological knowledge of the species for finding and validating pertinent biomarkers, we validated the proof of concept of a multiplexed methodology for quantifying large numbers of biomarkers upon contaminant exposure.

However, these works were performed in laboratory-controlled conditions. One of the main goals of the use of these techniques in ecotoxicology is to make the transition from proof of concept to an operational monitoring tool for regulatory purposes. That is why ongoing

developments are currently conducted in an *in situ* application of such methodology to confront biomarker response pattern to field environmental variability.

Acknowledgements

We thank the "Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture" (France), the "Institut des Sciences Analytiques" (France), the "Commissariat de l'Energie Atomique et aux Energies Alternatives" (France), the Regional Water Agency Rhône-Méditerranée-Corse for financial support, and the ANR program "ProteoGam" (ANR-14-CE21-0006-02) for financial support. We also thank Adeline François and Renaud Tutundjian for technical support.

The authors declare no conflict of interest.

References

Amiard-Triquet, C., Berthet, B., 2015. Chapter 7 - Individual Biomarkers, Aquatic Ecotoxicology. Academic Press, pp. 153-182.

Anderson, L., Hunter, C.L., 2006. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol Cell Proteomics 5, 573-588.

Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., Hartmann, E.M., 2014. Non-model organisms, a species endangered by proteogenomics. J Proteomics 105, 5-18.

Arockiaraj, J., Gnanam, A.J., Palanisamy, R., Bhatt, P., Kumaresan, V., Chaurasia, M.K., Pasupuleti, M., Ramaswamy, H., Arasu, A., Sathyamoorthi, A., 2014. A cytosolic glutathione s-transferase, GST-theta from freshwater prawn Macrobrachium rosenbergii: Molecular and biochemical properties. Gene 546, 437-442.

Auguste, M., Mestre, N.C., Rocha, T.L., Cardoso, C., Cueff-Gauchard, V., Le Bloa, S., Cambon-Bonavita, M.A., Shillito, B., Zbinden, M., Ravaux, J., Bebianno, M.J., 2016. Development of an ecotoxicological protocol for the deep-sea fauna using the hydrothermal vent shrimp Rimicaris exoculata. Aquatic Toxicology 175, 277-285.

Barros, S., Montes, R., Quintana, J.B., Rodil, R., Oliveira, J.M.A., Santos, M.M., Neuparth, T., 2017. Chronic effects of triclocarban in the amphipod Gammarus locusta: Behavioural and biochemical impairment. Ecotoxicology and Environmental Safety 135, 276-283.

Benninghoff, A., 2007. Toxicoproteomics--the next step in the evolution of environmental biomarkers. Toxicol Sci 95, 1-4.

Besse, J.-P., Coquery, M., Lopes, C., Chaumot, A., Budzinski, H., Labadie, P., Geffard, O., 2013. Caged Gammarus fossarum (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: Towards the determination of threshold values. Water Research 47, 650-660.

Charmantier, G.U.Y., 1998. Ontogeny of osmoregulation in crustaceans: a review. Invertebrate Reproduction & Development 33, 177-190.

Charnot, A., Gouveia, D., Armengaud, J., Almunia, C., Chaumot, A., Lemoine, J., Geffard, O., Salvador, A., 2017. Multiplexed assay for protein quantitation in the invertebrate Gammarus fossarum by liquid chromatography coupled to tandem mass spectrometry. Anal Bioanal Chem 17, 017-0348.

Charron, L., Geffard, O., Chaumot, A., Coulaud, R., Jaffal, A., Gaillet, V., Dedourge-Geffard, O., Geffard, A., 2014. Influence of molting and starvation on digestive enzyme activities and energy storage in Gammarus fossarum. PLoS ONE 9.

Chaurasia, M.K., Ravichandran, G., Nizam, F., Arasu, M.V., Al-Dhabi, N.A., Arshad, A., Harikrishnan, R., Arockiaraj, J., 2016. In-silico analysis and mRNA modulation of detoxification enzymes GST delta and kappa against various biotic and abiotic oxidative stressors. Fish and Shellfish Immunology 54, 353-363.

Connon, R.E., Geist, J., Werner, I., 2012. Effect-Based Tools for Monitoring and Predicting the Ecotoxicological Effects of Chemicals in the Aquatic Environment. Sensors (Basel, Switzerland) 12, 12741-12771.

Coulaud, R., Geffard, O., Xuereb, B., Lacaze, E., Quéau, H., Garric, J., Charles, S., Chaumot, A., 2011. In situ feeding assay with Gammarus fossarum (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring. Water Research 45, 6417-6429.

Dagnino, A., Allen, J.I., Moore, M.N., Broeg, K., Canesi, L., Viarengo, A., 2007. Development of an expert system for the integration of biomarker responses in mussels into an animal health index. Biomarkers 12, 155-172.

Dayras, P., Charmantier, G., Chaumot, A., Vigneron, A., Coquery, M., Quéau, H., Artells, E., Lignot, J.H., Geffard, O., Issartel, J., 2017. Osmoregulatory responses to cadmium in reference and historically metal contaminated Gammarus fossarum (Crustacea, Amphipoda) populations. Chemosphere 180, 412-422.

Fan, L., Wang, A., Wu, Y., 2013. Comparative proteomic identification of the hemocyte response to cold stress in white shrimp, Litopenaeus vannamei. Journal of Proteomics 80, 196-206.

Felten, V., Charmantier, G., Mons, R., Geffard, A., Rousselle, P., Coquery, M., Garric, J., Geffard, O., 2008. Physiological and behavioural responses of Gammarus pulex (Crustacea: Amphipoda) exposed to cadmium. Aquatic Toxicology 86, 413-425.

Fernandez Gimenez, A.V., Garcia-Carreno, F.L., Navarrete del Toro, M.A., Fenucci, J.L., 2002. Digestive proteinases of Artemesia longinaris (Decapoda, Penaeidae) and relationship with molting. Comp Biochem Physiol B Biochem Mol Biol 132, 593-598.

Fernández, I., Oliva, M., Carrillo, O., Wormhoudt, A.V., 1997. Digestive enzyme activities of Penaeus notialis during reproduction and moulting cycle. Comparative Biochemistry and Physiology Part A: Physiology 118, 1267-1271.

Gallien, S., Duriez, E., Domon, B., 2011. Selected reaction monitoring applied to proteomics. J Mass Spectrom 46, 298-312.

Geffard, O., Xuereb, B., Chaumot, A., Geffard, A., Biagianti, S., Noël, C., Abbaci, K., Garric, J., Charmantier, G., Charmantier-Daures, M., 2010. Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment. Environmental toxicology and chemistry / SETAC 29, 2249-2259.

Hariharan, G., Suresh Kumar, C., Laxmi Priya, S., Paneer Selvam, A., Mohan, D., Purvaja, R., Ramesh, R., 2012. Acute and chronic toxic effect of lead (Pb) and zinc (Zn) on biomarker response in post larvae of Penaeus monodon (Fabricus, 1798). Toxicological & Environmental Chemistry 94, 1571-1582.

Houde, M., Douville, M., Gagnon, P., Sproull, J., Cloutier, F., 2015. Exposure of Daphnia magna to trichloroethylene (TCE) and vinyl chloride (VC): Evaluation of gene transcription, cellular activity, and life-history parameters. Ecotoxicology and Environmental Safety 116, 10-18.

Issartel, J., Boulo, V., Wallon, S., Geffard, O., Charmantier, G., 2010. Cellular and molecular osmoregulatory responses to cadmium exposure in Gammarus fossarum (Crustacea, Amphipoda). Chemosphere 81, 701-710.

Jemec, A., Drobne, D., Tisler, T., Sepcić, K., 2010. Biochemical biomarkers in environmental studies--lessons learnt from enzymes catalase, glutathione S-transferase and cholinesterase in two crustacean species. Environmental science and pollution research international 17, 571-581.

Jemec, A., Drobne, D., Tišler, T., Trebše, P., Roš, M., Sepčić, K., 2007. The applicability of acetylcholinesterase and glutathione S-transferase in Daphnia magna toxicity test. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology 144, 303-309.

Jubeaux, G., Audouard-Combe, F., Simon, R., Tutundjian, R., Salvador, A., Geffard, O., Chaumot, A., 2012a. Vitellogenin-like proteins among invertebrate species diversity: potential of proteomic mass spectrometry for biomarker development. Environmental science & technology 46, 6315-6323.

Jubeaux, G., Simon, R., Salvador, A., Lopes, C., Lacaze, E., Quéau, H., Chaumot, A., Geffard, O., 2012b. Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: Inconclusive experience. Aquatic Toxicology 122-123, 9-18.

Jubeaux, G., Simon, R., Salvador, A., Quéau, H., Chaumot, A., Geffard, O., 2012c. Vitellogenin-like proteins in the freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males. Aquatic toxicology (Amsterdam, Netherlands) 112-113, 72-82.

Lacaze, E., Devaux, A., Jubeaux, G., Mons, R., Gardette, M., Bony, S., Garric, J., Geffard, O., 2011a. DNA damage in Gammarus fossarum sperm as a biomarker of genotoxic pressure: intrinsic variability and reference level. Science of the Total Environment 409, 3230-3236.

Lacaze, E., Devaux, A., Mons, R., Bony, S., Garric, J., Geffard, A., Geffard, O., 2011b. DNA damage in caged Gammarus fossarum amphipods: a tool for freshwater genotoxicity assessment. Environ Pollut 159, 1682-1691.

Lee, K.W., Raisuddin, S., Rhee, J.S., Hwang, D.S., Yu, I.T., Lee, Y.M., Park, H.G., Lee, J.S., 2008. Expression of glutathione S-transferase (GST) genes in the marine copepod Tigriopus japonicus exposed to trace metals. Aquatic Toxicology 89, 158-166.

Lee, S.O., Jeon, J.M., Oh, C.W., Kim, Y.M., Kang, C.K., Lee, D.S., Mykles, D.L., Kim, H.W., 2011. Two juvenile hormone esterase-like carboxylesterase cDNAs from a Pandalus shrimp (Pandalopsis japonica): cloning, tissue expression, and effects of eyestalk ablation. Comp Biochem Physiol B Biochem Mol Biol 159, 148-156.

Li, J., Ma, P., Liu, P., Chen, P., Li, J., 2015. The roles of Na(+)/K(+)-ATPase alpha-subunit gene from the ridgetail white prawn Exopalaemon carinicauda in response to salinity stresses. Fish Shellfish Immunol 42, 264-271.

Liu, C.-H., Tseng, D.-Y., Lai, C.-Y., Cheng, W., Kuo, C.-M., 2006. Molecular cloning and characterisation of prophenoloxidase cDNA from haemocytes of the giant freshwater prawn, Macrobrachium rosenbergii, and its transcription in relation with the moult stage. Fish & Shellfish Immunology 21, 60-69.

Liu, C.H., Chang, C.C., Chiu, Y.C., Cheng, W., Yeh, M.S., 2011. Identification and cloning of a transglutaminase from giant freshwater prawn, Macrobrachium rosenbergii, and its transcription during pathogen infection and moulting. Fish and Shellfish Immunology 31, 871-880.

Lobato, R.O., Nunes, S.M., Wasielesky, W., Fattorini, D., Regoli, F., Monserrat, J.M., Ventura-Lima, J., 2013. The role of lipoic acid in the protection against of metallic pollutant effects in the shrimp Litopenaeus vannamei (Crustacea, Decapoda). Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology 165, 491-497.

Lucu, Č., Towle, D.W., 2003. Na++K+-ATPase in gills of aquatic crustacea. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 135, 195-214.

Luo, J., Pei, S., Jing, W., Zou, E., Wang, L., 2015. Cadmium inhibits molting of the freshwater crab Sinopotamon henanense by reducing the hemolymph ecdysteroid content and the activities of chitinase and N-acetyl-beta-glucosaminidase in the epidermis. Comp Biochem Physiol C Toxicol Pharmacol 169, 1-6.

Marinković, M., De Leeuw, W.C., Ensink, W.A., De Jong, M., Breit, T.M., Admiraal, W., Kraak, M.H.S., Jonker, M.J., 2012.

Gene expression patterns and life cycle responses of toxicant-exposed chironomids. Environmental Science and Technology 46, 12679-12686.

Martyniuk, C.J., Alvarez, S., Denslow, N.D., 2012. DIGE and iTRAQ as biomarker discovery tools in aquatic toxicology. Ecotoxicology and Environmental Safety 76, 3-10.

Martyniuk, C.J., Denslow, N.D., 2009. Towards functional genomics in fish using quantitative proteomics. Gen Comp Endocrinol 164, 135-141.

Mleiki, A., Marigómez, I., El Menif, N.T., 2015. Effects of dietary Pb and Cd and their combination on glutathion-S-transferase and catalase enzyme activities in digestive gland and foot of the green garden snail, Cantareus apertus (Born, 1778). Bulletin of Environmental Contamination and Toxicology 94, 738-743.

Pellet, B., Geffard, O., Lacour, C., Kermoal, T., Gourlay-france, C., Tusseau-vuillemin, M.H., 2009. A model predicting waterborne cadmium bioaccumulation in Gammarus pulex: the effects of dissolved organic ligands, calcium, and temperature. Environ Toxicol Chem 28, 2434-2442.

Reddy, P.R., Nagaraju, G.P.C., Reddy, P.S., 2004. Involvement of Methyl Farnesoate in the Regulation of Molting and Reproduction in the Freshwater Crab Oziotelphusa senex senex. Journal of Crustacean Biology 24, 511-515.

Renuse, S., Chaerkady, R., Pandey, A., 2011. Proteogenomics. Proteomics 11, 620-630.

Revathi, P., Arockia Vasanthi, L., Munuswamy, N., 2011. Effect of cadmium on the ovarian development in the freshwater prawn Macrobrachium rosenbergii (De Man). Ecotoxicology and Environmental Safety 74, 623-629.

Rodríguez, E.M., López Greco, L.S., Fingerman, M., 2000. Inhibition of Ovarian Growth by Cadmium in the Fiddler Crab, Uca pugilator (Decapoda, Ocypodidae). Ecotoxicology and Environmental Safety 46, 202-206.

Rodríguez Moreno, P.A., Medesani, D.A., Rodríguez, E.M., 2003. Inhibition of molting by cadmium in the crab Chasmagnathus granulata (Decapoda Brachyura). Aquatic Toxicology 64, 155-164.

Sanchez, B.C., Ralston-Hooper, K., Sepulveda, M.S., 2011. Review of recent proteomic applications in aquatic toxicology. Environ Toxicol Chem 30, 274-282.

Sharma, R., Komatsu, S., Noda, H., 2004. Proteomic analysis of brown planthopper: application to the study of carbamate toxicity. Insect Biochem Mol Biol 34, 425-432.

Silvestre, F., Dierick, J.F., Dumont, V., Dieu, M., Raes, M., Devos, P., 2006. Differential protein expression profiles in anterior gills of Eriocheir sinensis during acclimation to cadmium. Aquat Toxicol 76, 46-58.

Simon, R., Jubeaux, G., Chaumot, A., Lemoine, J., Geffard, O., Salvador, A., 2010a. Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: Application to vitellogenin in Crustacea (Gammarus fossarum). Journal of Chromatography A 1217, 5109-5115.

Simon, R., Jubeaux, G., Chaumot, A., Lemoine, J.r.m., Geffard, O., Salvador, A., 2010b. Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: Application to vitellogenin in Crustacea (Gammarus fossarum). Journal of Chromatography A 1217, 5109-5115.

Spencer, A.M., Fielding, A.H., Kamemoto, F.I., 1979. The Relationship between Gill NaK-ATPase Activity and Osmoregulatory Capacity in Various Crabs. Physiological Zoology 52, 1-10.

Sritunyalucksana, K., Soderhall, K., 2000. The proPO and clotting system in crustaceans. Aquaculture 191, 53-69.

Trapp, J., Armengaud, J., Gaillard, J.C., Pible, O., Chaumot, A., Geffard, O., 2016. High-throughput proteome dynamics for discovery of key proteins in sentinel species: Unsuspected vitellogenins diversity in the crustacean Gammarus fossarum. J Proteomics 146, 207-214.

Trapp, J., Armengaud, J., Pible, O., Gaillard, J.C., Abbaci, K., Habtoul, Y., Chaumot, A., Geffard, O., 2015. Proteomic investigation of male Gammarus fossarum, a freshwater crustacean, in response to endocrine disruptors. Journal of proteome research 14, 292-303.

Trapp, J., Armengaud, J., Salvador, A., Chaumot, A., Geffard, O., 2014a. Next-generation proteomics: Toward customized biomarkers for environmental biomonitoring. Environmental Science and Technology 48, 13560-13572.

Trapp, J., Geffard, O., Imbert, G., Gaillard, J.C., Davin, A.H., Chaumot, A., Armengaud, J., 2014b. Proteogenomics of gammarus fossarum to document the reproductive system of amphipods. Molecular and Cellular Proteomics 13, 3612-3625.

Ulm, L., Krivohlavek, A., Jurašin, D., Ljubojević, M., Šinko, G., Crnković, T., Žuntar, I., Šikić, S., Vinković Vrček, I., 2015. Response of biochemical biomarkers in the aquatic crustacean Daphnia magna exposed to silver nanoparticles. Environmental Science and Pollution Research 22, 19990-19999.

Vermeulen, A.C., Liberloo, G., Dumont, P., Ollevier, F., Goddeeris, B., 2000. Exposure of Chironomus riparius larvae (diptera) to lead, mercury and β-sitosterol: effects on mouthpart deformation and moulting. Chemosphere 41, 1581-1591.

Vigneron, A., Geffard, O., Coquery, M., François, A., Quéau, H., Chaumot, A., 2015. Evolution of cadmium tolerance and associated costs in a Gammarus fossarum population inhabiting a low-level contaminated stream. Ecotoxicology 24, 1239-1249.

Wang, M.H., Wang, G.Z., 2009. Biochemical response of the copepod tigriopus japonicus mori experimentally exposed to cadmium. Archives of Environmental Contamination and Toxicology 57, 707-717.

Weidlich, S., Müller, S., Hoffmann, K.H., Woodring, J., 2013. Regulation of amylase, cellulase and chitinase

secretion in the digestive tract of the two-spotted field cricket, gryllus bimaculatus. Archives of Insect Biochemistry and Physiology 83, 69-85.

Xu, Y., Li, X., Deng, Y., Lu, Q., Yang, Y., Pan, J., Ge, J., Xu, Z., 2016. Comparative transcriptome sequencing of the hepatopancreas reveals differentially expressed genes in the precocious juvenile Chinese mitten crab, Eriocheir sinensis (Crustacea: Decapoda). Aquaculture Research.

Xuereb, B.t., Chaumot, A., Mons, R., Garric, J., Geffard, O., 2009. Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda). Intrinsic variability, reference levels, and a reliable tool for field surveys. Aquatic Toxicology 93, 225-233.

Yang, J., Liu, D., Dahms, H.U., Wang, L., 2015. Cadmium inhibits the vitellogenesis of freshwater crab Sinopotamon henanense. Environ Toxicol Chem 34, 1609-1616.

Yeh, M.S., Lai, C.Y., Liu, C.H., Kuo, C.M., Cheng, W., 2009. A second proPO present in white shrimp Litopenaeus vannamei and expression of the proPOs during a Vibrio alginolyticus injection, molt stage, and oral sodium alginate ingestion. Fish and Shellfish Immunology 26, 49-55.

Yuan, J., Wang, X., Gu, Z., Zhang, Y., Wang, Z., 2015. Activity and Transcriptional Responses of Hepatopancreatic Biotransformation and Antioxidant Enzymes in the Oriental River Prawn Macrobrachium nipponense Exposed to Microcystin-LR. Toxins 7, 4006-4022.

Supplementary Figures

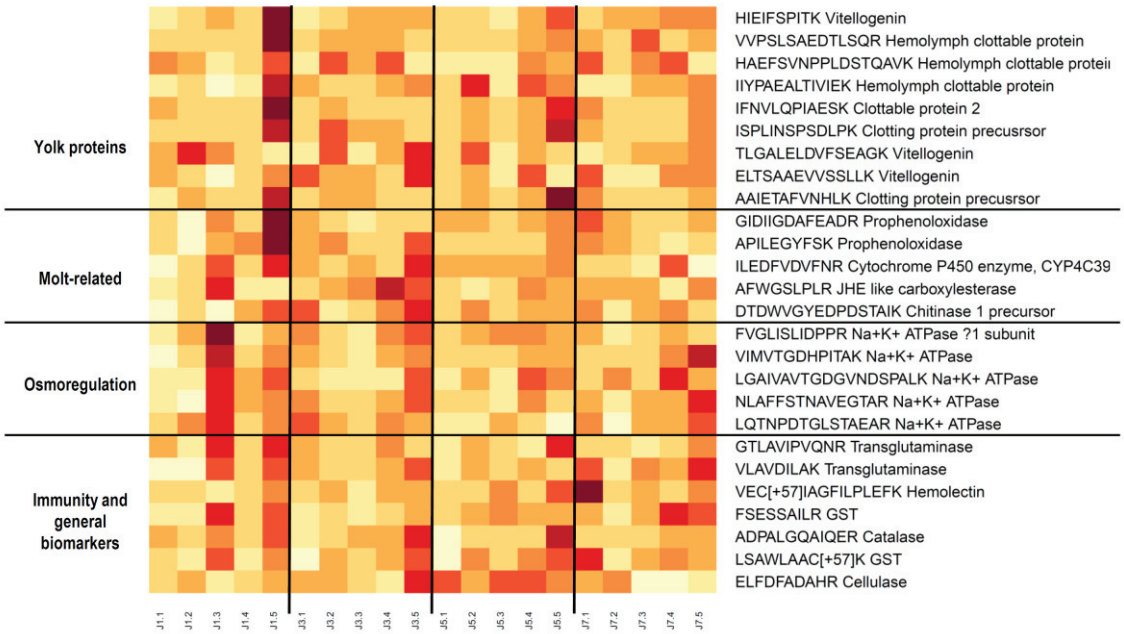


Figure A.1 - Heatmap of the quantification results in pmol peptide/mg organism of the entire set of peptides during four stages of spermatogenesis. For each stage (J1, J3, J5, J7) there are 5 individuals, represented in the y axis. The color gradient ranges from yellow (lowest concentrations) to red (highest concentrations) and signature peptides are clustered in accordance with their putative function: yolk peptides, molt-related peptides, osmoregulation-related Na^+K^+ ATPases, and general ecotoxicological biomarkers.

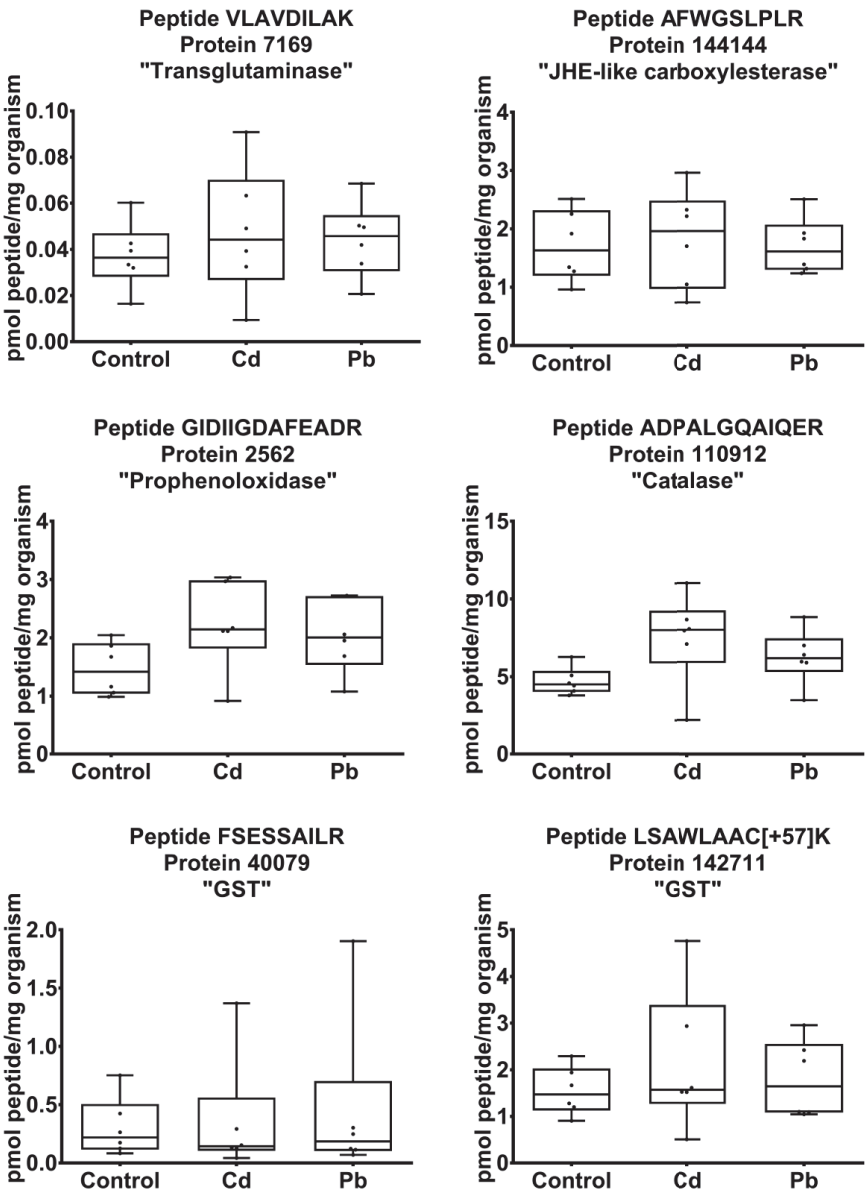


Figure A.2 - Levels of modulated proteins from the 21-day exposure in **male organisms** at day seven of exposure.

Publication nº3

Environmental Science & Technology

Article

Cite This: *Environ. Sci. Technol.* 2017, 51, 13417-13426

pubs.acs.org/est

Ecotoxico-Proteomics for Aquatic Environmental Monitoring: First in Situ Application of a New Proteomics-Based Multibiomarker Assay Using Caged Amphipods

Duarte Gouveia,[†,‡] Arnaud Chaumot,[†] Aurore Charnot,[§] Christine Almunia,[‡] Adeline François,[†] Lionel Navarro,[∥] Jean Armengaud,[‡] Arnaud Salvador,[§] and Olivier Geffard[*,†]

[†]Irstea, UR MALY, Laboratoire d'écotoxicologie, F-69625 Villeurbanne, France

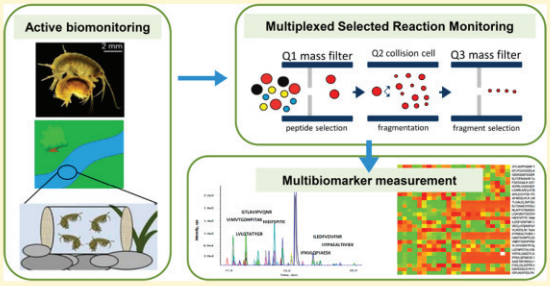
[‡]CEA-Marcoule, DRF/Joliot/DMTS/SPI/Li2D, Laboratory Innovative Technologies for Detection and Diagnostics, Bagnols-Sur-Cèze, F-30207, France

[§]Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENS de Lyon, Institut des Sciences Analytiques, UMR 5280, 5 rue de la Doua, F-69100 VILLEURBANNE, France

[∥]Agence De L'Eau Rhone Mediterranée Corse, F-69363 Lyon, France

Supporting Information

ABSTRACT: As a proof of principle, a selected reaction monitoring (SRM) mass spectrometry-based methodology was applied to the simultaneous quantification of dozens of protein biomarkers in caged amphipods (*Gammarus fossarum*). We evaluated the suitability of the methodology to assess complex field contaminations through its application in the framework of a regional river monitoring network. Thanks to the high throughput acquisition of biomarker levels in *G. fossarum* exposed in four reference and 13 contaminated sites, we analyzed the individual responses of 38 peptides reporting for 25 proteins of interest in 170 organisms. Responses obtained in contaminated sites included inductions of vitellogenin-like proteins in male organisms, inductions of Na^+K^+/ATPases, and strong inhibitions of molt-related proteins such as chitinase and JHE-carboxylesterase. Proteins from detoxification and immunity processes were also found modulated in abundance. Summarizing, the results presented here show that the SRM strategy developed for multibiomarker measurement paves a very promising way to define multiple indicators of the health status of sentinel organisms for environmental hazard assessment.



Introduction

In amphipod crustaceans, some well-established biomarkers were developed in the last decades, such as the protein biomarkers acetylcholinesterase (AChE) (Xuereb et al., 2009b), the detoxification/antioxidant enzymes glutathione *S*-transferase (GST), catalase (CAT) and superoxide dismutase (SOD) (Turja et al., 2014), or digestive enzymes(Charron et al., 2013). However, their use in routine biomonitoring is still limited by several technical factors like the lack of direct quantification methods in non-model sentinel species. The fact that each assay is specific for one biomarker of interest, mainly through the measurement of enzymatic activities, multiplies laboratory work, cost, and time necessary for the analysis of a large number of samples.

Hence, the development of multi-biomarker strategies that allow the simultaneous monitoring of a wide range of biological responses is required for improving environmental hazard assessment. In the sentinel amphipod *Gammarus fossarum*, a targeted proteomics method was recently developed for the simultaneous quantification of 40 proteins, using Selected Reaction Monitoring (SRM) mass spectrometry (Charnot et al., 2017). This methodology has been used to validate candidate proteins with key physiological roles in order to propose them as biomarkers of toxicity (Gouveia et al., 2017). These studies highlighted the great potential of this methodology in

ecotoxicology, due to its capacity to monitor simultaneously an ensemble of protein biomarkers from one single organism, and to its reduced technical variability, higher sensitivity, and lower detection limits when compared to shotgun methods (Simmons et al., 2015).

However, it is now mandatory to make proof of principle of the use of these techniques in field applications. One drawback is the lack of control of biological factors of sentinel organisms (genetic background, age, sex, reproductive status), which introduce variability in proteomic biomarker responses in the field, especially when comparing different populations (Bahamonde et al., 2016; Hidalgo-Galiana et al., 2014). For instance, while gender has no influence in some biomarker responses, like for the AChE activity in gammarids (Xuereb et al., 2009b), it has been reported as a source of confusion in several studies. For example, Sornom et al. (Sornom et al., 2010) and Gismondi et al. (Gismondi et al., 2013) observed that in *Gammarus roeseli*, gender influences the organisms responses to a salinity and a heavy metal stress, respectively. To overcome these difficulties of confounding biological factors, a robust active biomonitoring strategy based on caged organisms from the same reference population, was developed with *G. fossarum* and applied to several *in situ* studies (Besse et al., 2013; Coulaud et al., 2011; Jubeaux et al., 2012b; Lacaze et al., 2011b). *G. fossarum* is an ecologically relevant test species commonly used for freshwater monitoring. They are widespread in rivers around Europe, occurring often in high densities, and are easily identifiable to the species level. Data are available for a wide range of stressors and lethal and sublethal responses for this genus. Moreover, a reproduction toxicity test is available for *G. fossarum* (Geffard et al., 2010).

The objective of the present study was to assess the suitability of the SRM mass-spectrometry-based multibiomarker methodology developed in *G. fossarum*. For this, protein biomarker quantitation was carried out during a caging deployment operated in South-Eastern France throughout the regional river monitoring network from the "Rhône Méditerranée Corse" (RMC) French Water Agency. We analyzed caged male and female organisms in 4 reference sites and 13 sites subjected to chemical contamination, previously monitored and prioritized by the French Water Agency. After exposure, we applied the SRM strategy to analyze 38 peptides that report for 25 protein biomarkers of interest (involved in reproduction, immunity, homeostasis, detoxification, and defense mechanisms). Firstly, our aim was to evaluate the robustness of the methodology and its capacity to detect and describe the possible response patterns of biomarkers in male and female organisms, when facing a complex and realistic field exposure. Thirdly, we aimed at identifying proteins whose modulations were physiologically relevant, in order to propose robust biomarkers of toxic exposure in *G. fossarum*.

Materials and Methods

Sampling, selection and field exposure of caged organisms

Gammarids were collected from the Bourbre River in France from a source population commonly used in our laboratory (Besse et al., 2013; Ciliberti et al., 2017; Coulaud et al., 2011; Jubeaux et al., 2012b) and acclimatized to laboratory conditions as previously described (Jubeaux et al., 2012c). *G. fossarum* were collected by kick sampling using a net, and quickly transported to the laboratory. In the laboratory, organisms were kept in 30 L tanks continuously supplied with drilled groundwater and under constant aeration for at least 10 days. The drilled groundwater water from the tanks was adjusted to a conductivity of 600±20 µS/cm, pH of 7.4±0.2, with a flow rate of 2.25 L/h. A 16/8h light/dark photoperiod was maintained and the temperature was kept at 12±1°C. Organisms were fed *ad libitum* with alder leaves (*Alnus glutinosa*), previously conditioned for 6±1 days using the same drilled groundwater water.

Male and female gammarids in amplexus (female at final D2 molting stage) were caged and fed *ad libitum* with alder leaves during exposure. Organisms were placed in punctured polypropylene cylinders to allow free circulation of water. One polypropylene cylinder containing 10 couples was placed in each site (a total of 10 males and 10 females in a common cylinder per site). Each exposure was conducted from the female final D2 molting stage (day 0) until they reach the C2/D1 stage. Since the gammarid molting cycle is temperature-dependent, the duration of exposure was different among sites (Table 1). This ensures that the female organisms were exposed during new starting reproductive cycle, up to a determined stage of follicular maturation and marsupial embryonic development (see Geffard et al. 2010 for details (Geffard et al., 2010)), and that males are in a common reproductive status (no copulation after the initial fertilization at the beginning of the exposure). At the end of exposure, organisms were counted (for survival rate assessment), weighed, and directly frozen in liquid nitrogen and stored at -80°C until further analysis.

Study sites

The 17 study sites belong to two river monitoring networks implemented by the French Water Agency Rhône-Méditerranée-Corse (South Eastern France), in an attempt to meet the Water Framework Directive (WFD) requirements. The surveillance-monitoring network aims to allow the assessment of long-term changes in natural conditions, and to provide an assessment of the overall surface water status within each catchment and sub-catchment of the river basin district. This network is complemented by the operational monitoring network, which aims to establish the status of water bodies identified as being at risk of failing to meet the WFD environmental objectives and their evolution in response to remediation actions. Seventeen sites, which have been intensively monitored since more than 10 years by the Agency, were selected among these observation networks on a large geographical scale (**Figure S1**). Hence, the study considered four reference sites *versus* thirteen contaminated sites. These sites are in fact situated on water bodies, for which the Agency previously assessed the risk of failing to achieve good chemical and ecological quality in 2021, due to specific adverse anthropic influences (**Table 1**). The risk related to micro-pollutant contamination was estimated through the expert system developed by the regulatory Water Agency, which consists in a standardized procedure allowing the integration of heterogeneous quantitative data from different sources (experimental observations, model outputs) and expert judgment. More precisely, this expert system mainly considers data from multi-year chemical monitoring of toxic substances in water and sediments (including WFD priority substances), data from the detected and predicted levels of contaminants in connected industrial and wastewater treatment plant effluents, and records from land cover analysis notably to quantify pesticide pressure.

Biomarker peptides/proteins used in the multiplexed SRM assay

Quantified proteins are listed in Table S1. Proteins and peptides were selected based on their putative physiological role, given by their bioinformatic functional annotations provided in the GFOSS database (see Trapp et al. 2014 (Trapp et al., 2014b)). The protein list sought to cover different important physiological functions potentially subject to disruption by exposure to contaminants: sex-specific and reproduction-related proteins, and proteins whose annotations suggested roles in immunity, homeostasis, detoxification and defence mechanisms. The reporter peptides used for each protein were validated in previous studies (Charnot et al., 2017; Gouveia et al., 2017).

Each peptide is assigned to a specific protein ID (some peptides belong to the same protein ID), and each protein ID corresponds to an assembled contig (=one putative protein) from the RNAseq GFOSS database (Trapp et al., 2014b). Different protein IDs share sometimes a common functional annotation. Because the GFOSS database is a *de novo* assembled RNAseq, we cannot ascertain whether distinct protein IDs with the same annotation correspond to duplicated paralog genes, to alternatively spliced isoforms of a unique gene, or to redundant contigs in the *de novo* assembled RNAseq. The targeted functions were considered as crucial for the reproduction, homeostasis, and defense mechanisms of the organisms. Therefore, the chosen biomarkers could give important insights into toxic disruption of these key physiological processes.

Protein analysis by LC-MS/MS

Five male and five female organisms were analyzed per site. The procedures used for total protein extraction and digestion were the same as in our previous studies (Charnot et al., 2017). Mass spectrometry analysis was performed using the targeted mass spectrometry mode SRM. This technique uses a two-stage mass filtering to analyze previous selected pairs of precursor/product ions in a triple quadrupole mass spectrometer. Compared to the untargeted full scan MS/MS analyses, the SRM approach is much more sensitive (in the order of 100 fold) and selective regarding quantitative analyses. The scheduled SRM algorithm from Analyst 1.5 software was used for creating short time windows around the peptides expected retention time, therefore improving the maximum number of transitions and diminishing analysis time. Three transitions per peptide were monitored. Heavy labeled isotopologs of target peptides of known concentration were spiked into the samples, and concomitantly monitored to ensure precise and accurate measurements of endogenous peptide concentrations.

LC–MS/MS analysis was performed with an HP1200 series HPLC device (Agilent Technologies, Waldbronn, Germany) coupled to a QTRAP® 5500 LC/MS/MS System hybrid triple quadrupole/linear ion trap mass spectrometer (Applied Biosystems/ MDS Analytical Technologies, Foster City, CA, USA) equipped with a Turbo V™ ion source.

Table 1: The seventeen study sites, their quality indicators as established by the RMC Water Agency, conditions during the caging experiment and some individual effects recorded during the assay.

Study site (river, city)	Code	Water agency indicators*						Conditions during the *in situ* assay					Individual effects recorded during the *in situ* assay		
		National grid reference	Status in national surveillance network	Detection of toxic substances in the waterbody	Presence of toxic substances in discharged industrial and urban effluents	Risk of degraded river quality due to the presence of chemicals (without pesticides)	Risk of degraded river quality due to the presence of pesticides	Deployment date	Average water temperature (°C)	Average conductivity (µs/cm)	pH	Exposure duration (days)	Survival (%)	Feeding inhibition	Molt delay
Guiers, Saint-Laurent-du-Pont	R1	6078200	reference	low	none	none	none	04/22/2015	8.0	242.5	8,6	31	95		
Ain , Saint-Maurice-de-Gourdans	R2	6092000	reference	low	low	weak	weak	04/23/2015	12.1	340.0	8,4	27	80		
Ardèche, Saint-Julien-de-Peyrolas	R3	6115700	reference	low	low	weak	weak	04/22/2015	16.1	210.5	7,9	21	85		
Albarine, Argis	R4	6090600	reference	low	none	weak	none	04/23/2015	10.6	300.0	8,5	27	95		
Saône, Saint-Bernard	C1	6053800		medium	high	weak	high	04/20/2015	16.3	470.0	8,2	20	100		
Durance, Les Mées	C2	6159000		high	medium	high	weak	04/23/2015	13.8	522.0	8,4	22	90		
Eygoutier, Toulon-Le Bosquet	C3	6710200		high	low	high	high	04/23/2015	16.4	875.0	8,1	18	90	×	
Ognon, Chassey-lès-Montbozon	C4	6425800		low	low	high	none	04/20/2015	13.8	218.0	8,2	26	100		
Allan, Bart	C5	6026000		high	medium	high	weak	04/21/2015	13.5	364.0	8,1	26	100	×	
Cance, Sarras	C6	6103500		high	medium	high	high	04/22/2015	14.1	146.0	7,9	22	90	×	
Azergues, Lucenay	C7	6057700		high	medium	high	high	04/20/2015	14.0	375.5	8,1	20	90		×
Durgeon, Pontcey	C8	6003500		medium	medium	weak	high	04/20/2015	14.1	591.0	8,3	26	100		
Fure, Tullins-Le Verney	C9	6147140		high	medium	high	high	04/22/2015	13.2	462.0	8,4	23	90		
Fier, Poisy	C10	6070100		medium	high	high	none	04/21/2015	11.6	304.0	8,4	28	45		
Argens, Le Thoronet	C11	6300121		low	medium	none	weak	04/23/2015	15.6	1098.0	8,2	18	85		×
Turdine, L'Arbresle	C12	6057200		high	low	high	high	04/20/2015	14.0	316.0	7,8	20	85	×	
Vistre, Aubord	C13	6193500		low	high	weak	high	04/22/2015	16.3	744.0	8,0	18	85	×	

* These indicators qualify the risk to fail to achieve good chemical and ecological quality of water bodies due to specific adverse anthropic influences.

The LC separation of the 20 µL injected sample was carried out on an Xbridge C_{18} column (100mm×2.1mm, particle size 3.5 µm) from Waters (Milford, MA, USA). Details concerning sample preparation, chromatographic separation of peptides and mass spectrometric analysis can be found in Supporting Information.

Quantification and statistical analysis

Peptide absolute concentrations were calculated using the endogenous/labelled peptide peak area ratio (peak analysis with Skyline v3.1, MacCoss Lab Software, USA), since the amount of labelled peptide was known (1000 ng/mL).

Results are presented either as heatmaps (heatmap.2 function in R software), or as min max boxplots of five biological replicate samples. Data are reported in pmol of peptide per milligram of organism wet-weight. All graphs and statistical analysis for testing significant differences in peptide concentrations between conditions were performed with the GraphPad Prism Version 7.02 software. Non-parametric Kruskal-Wallis tests with Dunn's corrections for multiple comparisons analysis were used for comparison between reference and contaminated sites. Significant differences were accepted at p-values <0.05.

Results and Discussion

In this study, we applied a mass spectrometry-based multibiomarker strategy in organisms after an *in situ* exposure to several aquatic environments. The survival rates obtained for the sites chosen as references were in accord with the U.S. Environmental Protection Agency's survival rate recommended for routine amphipod bioassays (≥80%). In contaminated sites, only site C9 presented significant mortality (survival rate of 45%). The simultaneous analysis of 25 proteins of interest in each individual allowed gaining important insights into toxicity of potential environmental contaminations.

Identification of peptide response patterns after exposure

The concentrations obtained after peptide quantification for the entire set of peptides in each exposed male and female organisms are available in **Tables S2 and S3**, respectively. **Figure 1** shows two color heatmaps of the male and female peptide quantification data, giving a general overview of biomarker levels and their modulation among study sites. Following our previous study (Gouveia et al.), reporter peptides were separated into 4 main groups, according to the annotated function of their corresponding proteins: yolk-related, molt-related, osmoregulation-related, immune-related and general ecotoxicological biomarkers.

Globally, the *in situ* application of the SRM method allowed for the quantitation of a set of 38 peptides reporting for 25 proteins of interest, in 170 male and female organisms that were exposed in 17 different rivers. This resulted in the acquisition of 6460 peptide concentration values, roughly assessed within a total of seven days for laboratory sample preparation and analytical LC-MS/MS runs. The methodology allowed for the simultaneous detection, in one whole organism, of peptides with concentrations ranging from 0.0016 to 10.7 pmol protein per mg of organism. This demonstrates the competitive time and cost efficiency of this approach when compared to the multibiomarker approaches currently used in ecotoxicology. These approaches involve the use of a wide range of biological responses that are integrated to create a global monitoring index. Despite simplifying the analysis by integrating the several biomarkers in one single index, there is still a need of performing a specific analytical procedure for each biomarker (Chalghmi et al., 2016; Kerambrun et al., 2012; Minier et al., 2000; Serafim et al., 2011), leading to extremely resource-consuming protocols (time, cost, organisms). Moreover, direct protein quantification methods are scarce, the exception being the vitellogenin ELISA measurement. Since this test is specific for a certain protein, it can rarely be applied in other species due to low cross-reactivity across species.

Analyzing the heatmaps represented in **Figure 1** makes possible to identify some clear patterns in response to the exposure. Because of the simultaneous measurement of several proteins with different physiological functions, a quick assessment of the affected pathways can indeed be made. It is clear, for example, that yolk peptides were modulated in both male (**Figure 1A**) and female organisms (**Figure 1B**). As for the other protein classes, there was a strong decrease in the concentration of some molt-related proteins (especially proteins 181833 and 144144), verified in both sexes but more pronounced in male organisms in sites C1, C2, C3, C4 and C5. Interestingly, in these same sites there is an increase of the concentrations of osmoregulatory sodium pumps in both male and female organisms, which suggests disturbances in the molting processes. Some biomarkers were not significantly modulated throughout exposure as exemplified in **Figure S2**. These non-responsive biomarkers show the repeatability of the measurement.

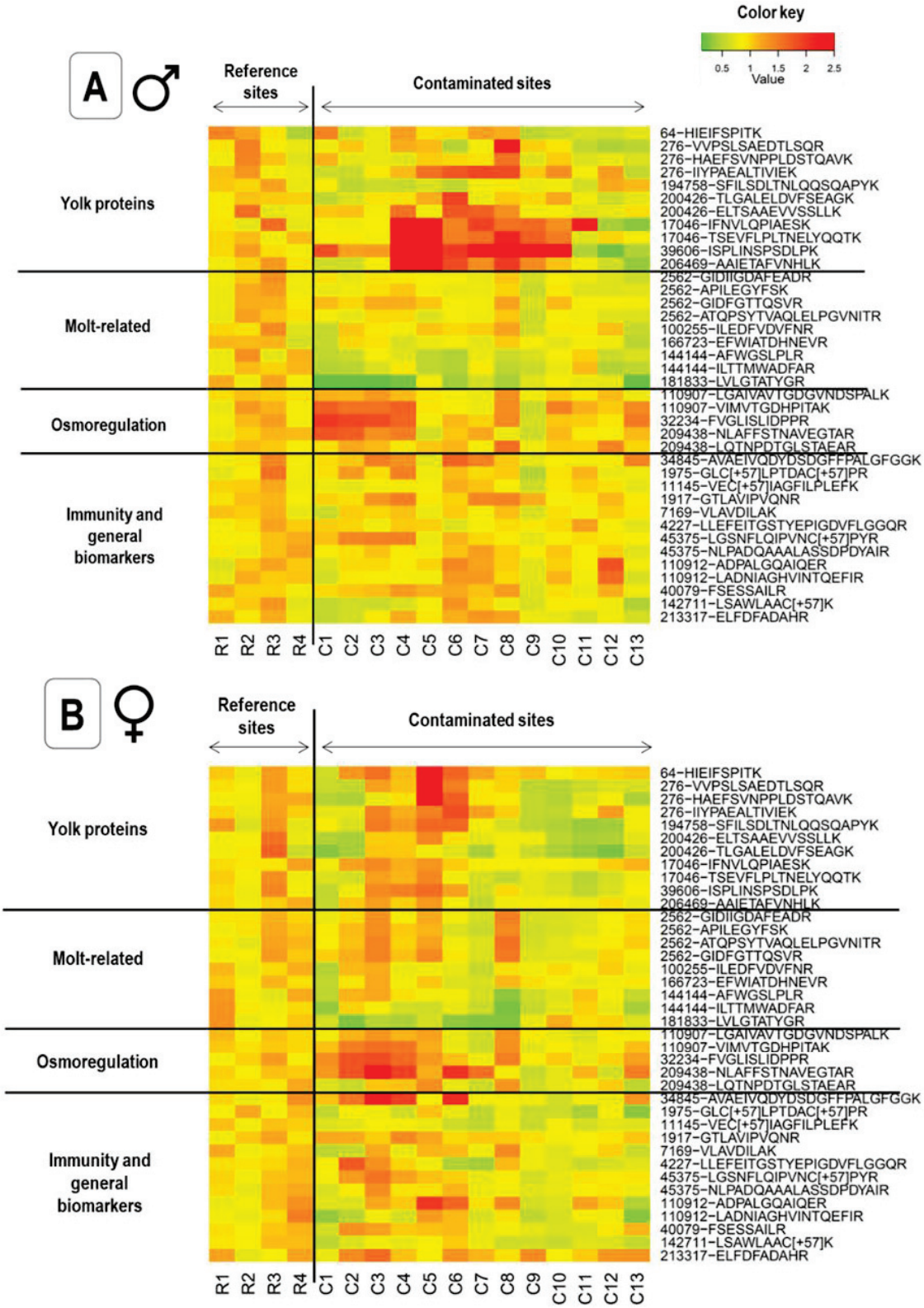


Figure 1: Heatmap of the quantification results in male (**A**) and female (**B**) organisms for the entire set of peptides in the seventeen studied sites. Each site is represented in columns, and each signature peptide in lines. For the heatmap construction, the fold change values relative to the mean of the four reference sites were used. The color gradient ranges from green (lowest concentrations) to red (highest concentrations) and signature peptides are clustered in accordance with their putative function: yolk peptides, molt-related peptides, osmoregulation-related Na^+K^+ ATPases, and general ecotoxicological biomarkers.

Of note, for the site C9 there was a decrease in the concentration of the majority of biomarkers in both males and females. This general inhibition could be due to an acute toxicity that translated in significant mortality. After having identified the most contrasted patterns in the overall targeted proteomic response through the heatmaps, the next sections examine in more details the modulations of peptides supported by comparative statistical tests between reference and contaminated sites.

Yolk proteins

As highlighted in a recent study (Trapp et al., 2016), there is a large variety of proteins with vitellogenin-like function in *G. fossarum*, with at least eight proteins originating from different families of the large lipid transfer protein superfamily (notably proteins closely related to clotting proteins or apolipocrustaceins identified in decapod crustaceans). These proteins were previously validated as being involved in yolk formation in *G. fossarum* (Gouveia et al., 2017; Trapp et al., 2016). In this study, seven of these eight proteins were analyzed.

Average peptide concentrations for yolk proteins in female organisms ranged from 0.37 to 4.16 pmol/mg of organism, being the most intense together with peptides from proteins 32234 (Na^+K^+ ATPase), 45375 (catalase) and 166723 (Farnesoic acid-O-methyl transferase). Comparing females from reference and contaminated sites, the signature peptide TLGALELDVFSEAGK from protein 200426 is inhibited in site C12 (adjusted *p* value of 0.0097), and present trends for inhibition in sites C1 and C11 (adjusted *p* values of 0.0516 and 0.0776 respectively). Concentration decreases could be related to an inhibition in the vitellogenesis processes, seen that other vtg-like peptides presented the same trend for decreasing concentrations in the same sites. **Figure 2** reveals potential site-specific inductions for vtg-like proteins in exposed males, along with some inter-individual variability, especially in sites C4 and C5. Of the five biological replicates, only some individuals present higher levels of vtg-like proteins. However, these inductions are in the order of 5-18 fold change, much lower than those observed in endocrine disrupted male fishes (Jobling and Tyler, 2003; Kirby et al., 2004; Scott et al., 2007). Similar fold changes and inter-variability were obtained by Xuereb et al (Xuereb et al., 2011) and Jubeaux et al (Jubeaux et al., 2012b) in assays that measured vtg transcripts by qPCR and a vtg peptide by mass spectrometry in *G. fossarum* male organisms, following laboratory exposures and caging field studies.

Despite being commonly used as biomarker of reproductive impairments in fish exposed to estrogenic compounds, the use of this biomarker in crustaceans has led to contradictory and controversial results that question its usefulness as a biomarker of feminization in these species (Matozzo et al., 2008; Short et al., 2014b). Our results advocate for further studies regarding the unusual diversity of vitellogenins in *G. fossarum* previously identified (Trapp et al., 2016), since the different yolk proteins are easily followed simultaneously and discriminatively by mass spectrometry. Challenging organisms to several stressors in laboratory-controlled conditions and monitoring these proteins would allow withdrawing more conclusions about their pertinence as biomarkers of reproductive toxicity (endocrine disruption) or other types of toxicity such as immunity and/or oxidative stress.

Molt-related proteins

Proteins with putative molt-related functions are annotated as prophenoloxidase, cythochrome P450, chitinase, juvenile hormone esterase-like carboxylesterase, and farnesoic acid methyl transferase. Among these, the chitinase peptide LVLGTATYGR (reporter for protein 181833) had the most significant modulations in response to contamination, as shown in **Figure 3A**.

Significant decreases in protein concentration were found for sites C1, C2, C3, C4, C6, and C13 in male organisms, and sites C2, C6, C7, and C8 in female organisms. In arthropods, chitinase plays an important role in the molting process, being responsible for chitin decomposition of the old cuticle into a more soluble form. This soluble cuticle can then be partially reabsorbed by the organism and reused for the synthesis of the new exoskeleton. Several studies in the literature show the implication of this enzyme in the molting cycle of crustaceans (Pesch et al., 2017; Rocha et al., 2012; Yang et al., 2016) and as a sensitive biomarker to different types of contamination. Upon pesticide contamination, Guan et al (Guan et al., 2016) reported a down-regulation of the chitinase1 transcript and enzymatic activity in the decapod *Macrobrachium nipponense* after treatment with the insecticide KK-42. Olsvik et al (Olsvik et al., 2015) observed an up-regulation of the chitinase gene expression by qPCR after exposure to high doses of teflubenzuron.

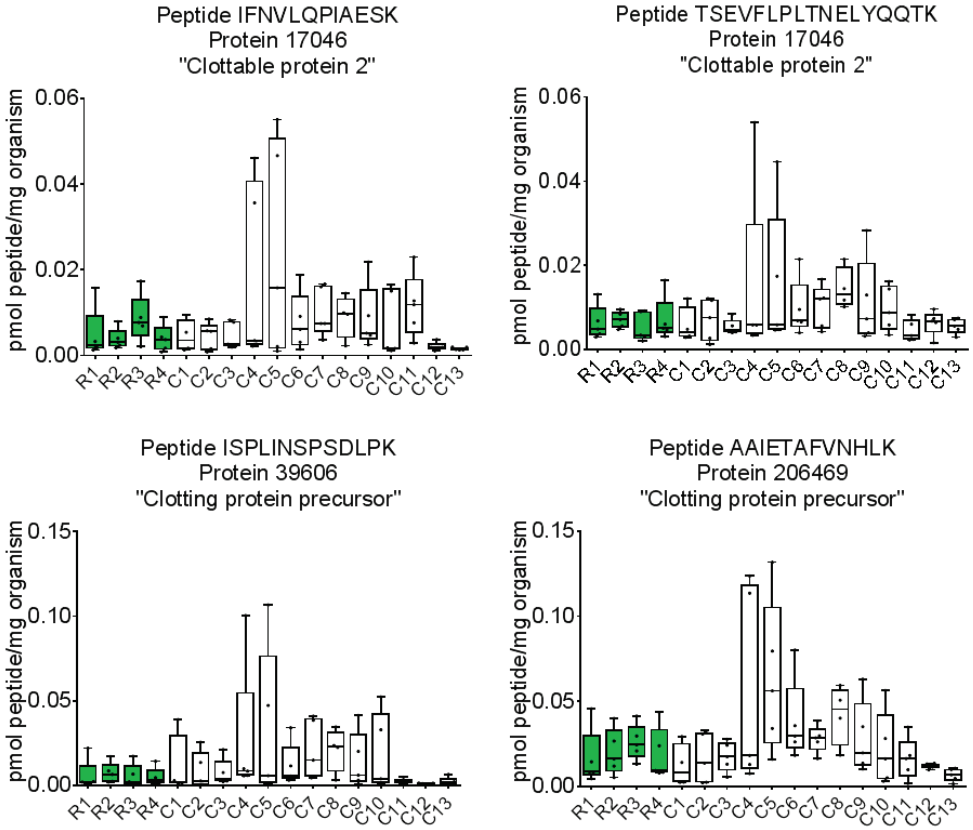


Figure 2: Quantification results in pmol peptide/mg organism of four vtg-signature peptides in exposed male organisms.

After zinc and cadmium exposures, Poynton et al (Poynton et al., 2007) and Luo et al (Luo et al., 2015) reported decreases in the chitinase enzymatic activity. In invertebrates, the majority of the studies found in the literature have studied either the enzymatic activity or the abundance of chitinase transcripts, but little information is available regarding chitinase in terms of direct protein analysis. However, a recent shotgun proteomic study (Gismondi et al., 2017) was performed in *G. fossarum* after a long-term exposure to three heavy metals. Using a label-free quantification methodology, a chitinase precursor was found overexpressed among the Cu modulated proteins in Cu-exposed organisms. Another study by Fan et al (Fan et al., 2016) searched for altered proteins after a cold-stress in the shrimp *Litopenaeus vannamei* by gel-based proteomics and mass spectrometry protein identification, but only as a screening method, since quantitative real-time PCR was used later for quantification of chitinase transcripts.

Peptide ILTTMWADFAR (reporter for protein 144144, Juvenile Hormone Esterase-like carboxylesterase 1) also presented significant modulations in response to contamination, as shown in **Figure 3B**. Significant decreases were found for sites C1, C8, and C13 in female organisms, with fold changes of 2.96, 4.2, and 2.5, respectively. We can also observe a trend for down-regulation in the same sites for male organisms. However, these modulations were not statistically significant. JHE-like carboxylesterase is related to the degradation of methyl farnesoate, a crustacean hormone implied in molting, control of oogenesis, and possibly in testicular development (Lee et al., 2011; Reddy et al., 2004). Although the metabolism of this hormone is still not clear, it has many similarities with the metabolism of the juvenile hormone in insects. This protein was also down regulated in our previous study, following a 10 µg Pb/L contamination in laboratory-controlled conditions, which concomitantly inhibited the molting process of females (Gouveia et al.). In that work, we had already underlined its potential as a biomarker of reproductive impairments in this species.

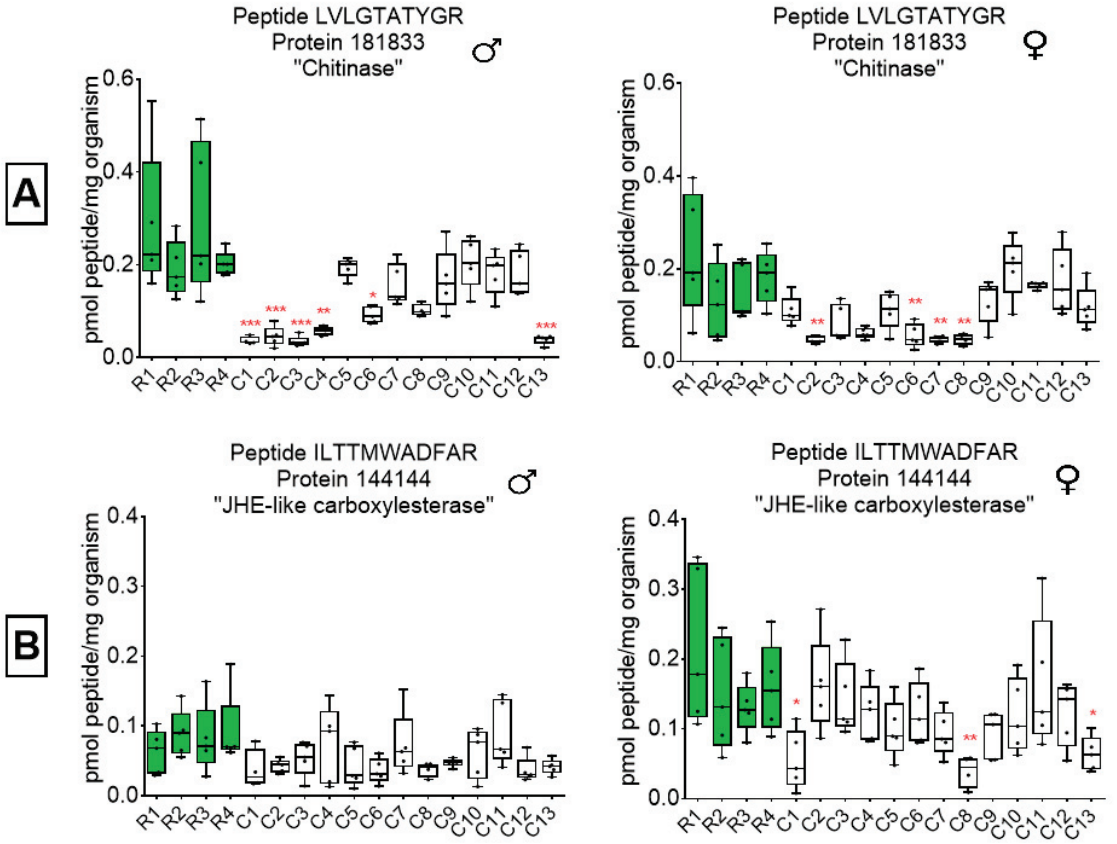


Figure 3: Quantification results in pmol peptide/mg organism of the signature peptide from (**A**) chitinase protein 181833 and (**B**) JHE carboxylesterase protein 144144 in exposed male and female organisms. * $p<0.05$; ** $p<0.01$; *** $p<0.001$

Molt-related biomarkers were the most responsive amongst the different peptides analyzed, with modulations in eleven of the thirteen contaminated sites. This was probably due to the high risk of pesticide contamination identified in six impacted sites as stated by the regulatory water agency (C1, C3, C8, C6, C7, and C13) (**Table 1**). Interestingly, three of these sites presented significant inhibitions of both chitinase and JHE-like carboxylesterase: C1, C8, and C13. Hence, we could hypothesize that these inhibitions were the consequence of the action of molt-interfering contaminants. This fact is additionally supported by the fact that these two proteins are related to ecdysone (Zhang and Zheng, 2017) and methyl farnesoate (Tao et al., 2017), two major targets of pesticide contaminants. The molting stage of female *G. fossarum*, parameter commonly used as a biomarker of reproductive impairments in this species (Geffard et al., 2010), was also determined for females simultaneously exposed in the same sites. However, delays in the molting cycle were only observed in two sites: C7 and C11 (**Table 1**). The inhibitions observed for sites C1, C8, and C13 could therefore represent precocious molecular responses that cannot be assessed by the physiological biomarkers, and will maybe cause delays in the next molting cycle of the organisms. Such clear inhibitions show the high sensitivity of these biomarkers to contamination.

Ionic pumps

Peptide FVGLISLIDPPR, annotated as Na^+/K^+ ATPase α1 subunit was modulated in some contaminated sites, as shown in **Figure 4**. Statistically significant differences were found in sites C2, C3, and C4 for females, and C1, C2, and C3 for males (adjusted *p* values between 0.006 and 0.0322). Na^+/K^+ ATPase proteins are present in the plasma membranes of animal cells and are responsible for providing energy through ATP hydrolysis during the translocation process of Na^+ and K^+ ions in the cell. The activity of these enzymes is essential for crustaceans and plays a crucial role in the mechanisms of osmoregulation by which these freshwater organisms regulate osmotic and ionic concentrations (Lucu and Towle, 2003). This means that osmoregulation processes were disturbed in sites C1, C2, C3 and C4. Since no marginal differences were observed in the physico-

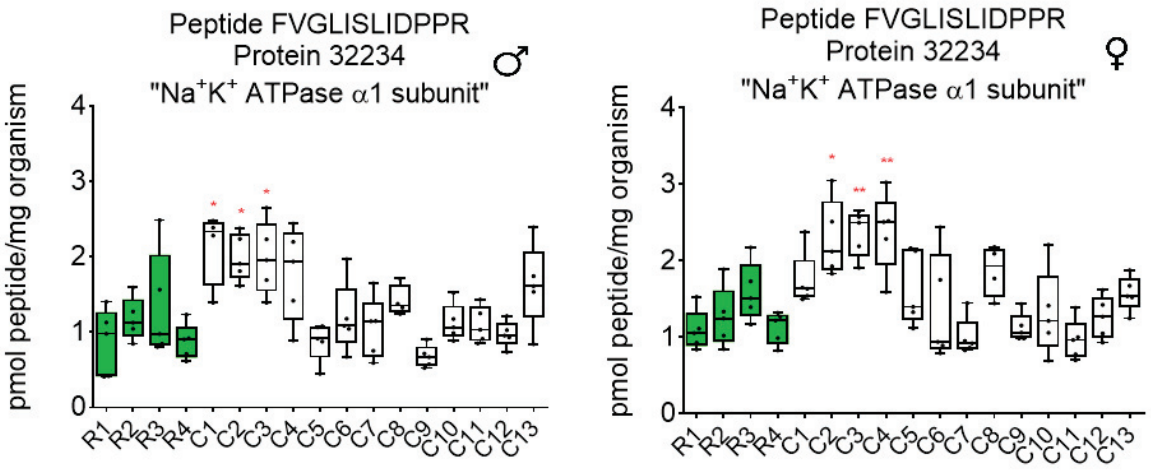


Figure 4: Quantification results in pmol peptide/mg organism of Na^+K^+/ATPase signature peptide (reporters for protein 32234) in exposed male and female organisms. * $p<0.05$; ** $p<0.01$; *** $p<0.001$

chemical parameters of the water, especially in conductivity, we can suggest that these changes were due to the presence of contaminants (**Table 1**).

Na^+/K^+ ATPase activity is often used as an indicator of heavy metal exposure and some studies also reported modulations in the activity of these enzymes after pesticide exposure. In a previous study (Gouveia et al., 2017), male and female *G. fossarum* were exposed to environmental concentrations of cadmium and lead. After seven days of contamination, a small induction was observed for lead-exposed males, but after twenty-one day contamination, no significant modulations of these pumps were observed. Despite being a very sensitive marker, the response is very dependent on the types of exposure. In *G. fossarum*, Issartel et al. (Dayras et al., 2017) observed an overexpression of gill Na^+/K^+ ATPase in Cd impacted animals, while Gismondi et al. (Gismondi et al., 2017) showed inductions of these enzymes after long-term exposure to Cu of *G. fossarum*. Harris and Santos (Harris and Santos, 2000) observed significant higher Na^+/K^+ ATPase activities in posterior gills of crabs *Ucides cordatus* and *Callinectes danae* after contamination with three different heavy metals, and Reddy et al (Reddy et al., 1992) and Comoglio et al (Comoglio et al., 2005) observed inhibitions (the latter concentration-dependent) in Na^+/K^+ ATPase enzymatic activity after pesticide exposures in the crab *Barytelphusa guerini* and the shrimp *Litopenaeus vannamei*, respectively. Using shotgun proteomics and label-free relative quantification, Tkatcheva et al (Tkatcheva et al., 2015) observed inhibitions of the Na^+/K^+ ATPase protein after lithium exposure in the rainbow trout. Our results, along with data from the literature, reinforce the pertinence of using Na^+/K^+ ATPase as a biomarker of toxic impairments in *G. fossarum*. However, the different modulations observed between studies in the literature call for further studies in *G. fossarum* for 1/ the characterization of different isoforms and/or subunits of the protein, and 2/ eliminating the possible effects of confounding factors in the modulations.

Immunity-related and general biomarkers

From the biomarkers related to immunity and general physiology, the peptide LSAWLAAC[+57]K (reporter for protein 142711, annotated as Gluthathion-S-Transferase) presented an interesting profile, as shown in **Figure S3**. Significant differences were found in sites C1, C10 and C13 in female organisms, and in site C13 for male organisms.

GST is an enzyme that catalyzes the conjugation of glutathione to xenobiotics and their metabolites, and is commonly employed as a biomarker of oxidative stress and detoxification. This protein was also modulated in our previous study (Gouveia et al., 2017) by two metals upon a twenty-one day contamination, cadmium and lead. As discussed in that study, this enzyme is normally found up-regulated and/or with a higher enzymatic activity in contaminated situations, and is very dependent on the species and types of toxicants used. Thus, these data confirm the sensitivity of this enzyme in *G. fossarum* towards contamination, after both a two-week metal contamination and field exposure.

Gender effect in biomarker responses

Despite being a source of confusion in biomarker analysis, the effect of xenobiotics on both male and female organisms is rarely assessed. In this study, we used both sexes in the *in situ* exposures, which allows assessing the possible differences in biomarker responses between genders.

Among the biomarkers discussed in this paper, Vtg-like yolk proteins presented the most significant differences in the responses obtained between sexes. As expected, these yolk proteins are much more abundant in female organisms. Considering the average concentrations in all sites, female/male concentration ratios range from 30 for peptide TLGALELDVFSEAGK (reporter for protein 200426) to 351 for peptide ISPLINSPSDLPK (reporter for protein 39606). In terms of response to contamination, some signature peptides from yolk proteins presented increased concentrations in males in some contaminated sites (especially proteins 17046, 39606, and 206469 in sites C4, C5, C6, C7, C8, C9, C10). In contrast, other yolk proteins presented modulated levels in females, notably inhibition for proteins 194758 and 200426 in sites C1, C2, C10, C11, and C12, or increased concentrations for proteins 64, 276 in site C5. These divergent response patterns between males and females clearly underline that the modulation of signature peptides from Vtg-like proteins is obviously related to different physiological processes in the two sexes, and that they cannot be substituted for each other for contaminant toxic action assessment.

Protein 144144, annotated as JHE-like carboxylesterase, was two times more abundant in females than in males, as represented in **Figure 3B** (average of 0.12 pmol/mg vs 0.059 pmol/mg in all sites). This is likely because the exposure was synchronized on the molt cycle of females and not males. Considering the putative role of the JHE-like carboxylesterase in the hormonal control of molting, it is not surprising that males and females, which were not sampled in the same molt stages at the end of exposure, presented different levels of signature peptides for this enzyme. Nevertheless, we can find similar patterns of modulation among sites for both genders, with low peptide concentration in C8.

Peptides GLCLPTDACPR (reporter for protein 1975 "hemocytin-like") and peptide VECIAGFILPLEFK (reporter for protein 11145 "hemolectin"), also had some strong inhibitions only in female organisms (**Figure S4**). The two proteins are immunity-related and their inhibition could reflect an impact of contaminants in the immunity capacities of female organisms, which appear then more vulnerable than males to this impact.

For other relevant biomarkers such as the enzymes chitinase (**Figure 3A**) and Na^+/K^+ ATPase ionic pump (**Figure 4**), both basal levels and modulation patterns were similar for both genders.

Relevance of targeted proteomics for aquatic Biomonitoring using sentinel species in Ecotoxicology

Summarizing, this high throughput methodology for the measurement of new generation proteomic biomarkers allows one to assess the simultaneous responses of 38 peptides that report for 25 protein biomarkers of interest in ecotoxicology in a large number of sites. In order to try to control biotic confounding factors and exposure conditions, a caging strategy of calibrated organisms from the same reference population was used. With this strategy, factor such as genetic background, age, sex, reproductive status, parasitism, food resource, and exposure duration were controlled.

Despite this, inter-individual variability was always present in our results, but this variability did not hinder to reveal clear responses to exposure in contaminated environments. Since only five biological replicates (i.e., five organisms) were used per site, this variability clearly appeals for an increase of the number of measurements per site to enhance the statistical power for toxicity detection when using these protein biomarkers in further field studies. This reinforces the importance to have high throughput assays for such biomarker data acquisition. **Table S4** summarizes the responses obtained in the different contaminated sites for the different protein "classes" analyzed in the multibiomarker assay. Inversely, sites C1and C4 modulated a greater number of biomarkers, symbolizing a wider impact in the organisms' molecular machinery. Besides, a substantial part of inter-site modulations could be due to environmental factors. However, the possibility of a high throughput data acquisition with this technique offers the perspective to gain insights in the natural variability of biomarkers to improve toxicity assessment. As we previously exemplified for several physiological and biochemical markers in *Gammarus* (Besse et al., 2013; Charron et al., 2013; Coulaud et al., 2011; Lacaze et al., 2011a; Xuereb et al., 2009b) one can thus seek to establish reference values integrating this natural variability, allowing an improved assessment of contamination-derived biological effects in the field. With this study, we showed that

the application of targeted proteomics for measuring dozens of biomarkers simultaneously in one single non-model organism, *in situ*, is a very promising strategy regarding one of the main objectives nowadays in ecotoxicology: developing functional monitoring tools for regulatory purposes.

Acknowledgements

We thank the "Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture" (France), the "Institut des Sciences Analytiques" (France), the "Commissariat de l'Energie Atomique et aux Energies Alternatives" (France), the Regional Water Agency Rhône-Méditerranée-Corse, and the ANR program "ProteoGam" (ANR-14-CE21-0006-02) for financial support. We also thank Hervé Quéau and Alexandre Ciliberti for technical support.

The authors declare no conflict of interest.

References

1. Xuereb, B. t.; Chaumot, A.; Mons, R.; Garric, J.; Geffard, O., Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda). Intrinsic variability, reference levels, and a reliable tool for field surveys. Aquatic Toxicology 2009, 93, 225-233.

2. Turja, R.; Guimarães, L.; Nevala, A.; Kankaanpää, H.; Korpinen, S.; Lehtonen, K. K., Cumulative effects of exposure to cyanobacteria bloom extracts and benzo[a]pyrene on antioxidant defence biomarkers in Gammarus oceanicus (Crustacea: Amphipoda). Toxicon : official journal of the International Society on Toxinology 2014, 78, 68-77.

3. Charron, L.; Geffard, O.; Chaumot, A.; Coulaud, R.; Queau, H.; Geffard, A.; Dedourge-Geffard, O., Effect of water quality and confounding factors on digestive enzyme activities in Gammarus fossarum. Environmental Science and Pollution Research 2013, 20, (12), 9044-9056.

4. Sornom, P.; Felten, V.; Médoc, V.; Sroda, S.; Rousselle, P.; Beisel, J.-N., Effect of gender on physiological and behavioural responses of Gammarus roeseli (Crustacea Amphipoda) to salinity and temperature. Environmental Pollution 2010, 158, (5), 1288-1295.

5. Gismondi, E.; Cossu-Leguille, C.; Beisel, J. N., Do male and female gammarids defend themselves differently during chemical stress? Aquatic Toxicology 2013, 140, (Supplement C), 432-438.

6. Charnot, A.; Gouveia, D.; Armengaud, J.; Almunia, C.; Chaumot, A.; Lemoine, J.; Geffard, O.; Salvador, A., Multiplexed assay for protein quantitation in the invertebrate Gammarus fossarum by liquid chromatography coupled to tandem mass spectrometry. Anal Bioanal Chem 2017, 17, (10), 017-0348.

7. Gouveia, D.; Chaumot, A.; Charnot, A.; Quéau, H.; Armengaud, J.; Almunia, C.; Salvador, A.; Geffard, O., Assessing the relevance of a multiplexed methodology for proteomic biomarker measurement in the invertebrate species Gammarus fossarum: a physiological and ecotoxicological study. Aquatic Toxicology 2017, 190, 199-209.

8. Simmons, D. B. D.; Benskin, J. P.; Cosgrove, J. R.; Duncker, B. P.; Ekman, D. R.; Martyniuk, C. J.; Sherry, J. P., Omics for aquatic ecotoxicology: Control of extraneous variability to enhance the analysis of environmental effects. Environmental Toxicology And Chemistry 2015, 34, (8), 1693-1704.

9. Hidalgo-Galiana, A.; Monge, M.; Biron, D. G.; Canals, F.; Ribera, I.; Cieslak, A., Reproducibility and Consistency of Proteomic Experiments on Natural Populations of a Non-Model Aquatic Insect. PLoS ONE 2014, 9, (8), e104734.

10. Bahamonde, P. A.; Feswick, A.; Isaacs, M. A.; Munkittrick, K. R.; Martyniuk, C. J., Defining the role of omics in assessing ecosystem health: Perspectives from the Canadian environmental monitoring program. Environ Toxicol Chem 2016, 35, (1), 20-35.

11. Jubeaux, G.; Simon, R.; Salvador, A.; Lopes, C.; Lacaze, E.; Quéau, H.; Chaumot, A.; Geffard, O., Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: Inconclusive experience. Aquatic Toxicology 2012, 122-123, 9-18.

12. Besse, J.-P.; Coquery, M.; Lopes, C.; Chaumot, A.; Budzinski, H.; Labadie, P.; Geffard, O., Caged Gammarus fossarum (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: Towards the determination of threshold values. Water Research 2013, 47, (2), 650-660.

13. Lacaze, E.; Devaux, A.; Mons, R.; Bony, S.; Garric, J.; Geffard, A.; Geffard, O., DNA damage in caged Gammarus fossarum amphipods: a tool for freshwater genotoxicity assessment. Environ Pollut 2011, 159, (6), 1682-91.

14. Coulaud, R.; Geffard, O.; Xuereb, B.; Lacaze, E.; Quéau, H.; Garric, J.; Charles, S.; Chaumot, A., In situ feeding assay with Gammarus fossarum (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring. Water Research 2011, 45, (19), 6417-6429.

15. Geffard, O.; Xuereb, B.; Chaumot, A.; Geffard, A.; Biagianti, S.; Noël, C.; Abbaci, K.; Garric, J.; Charmantier, G.; Charmantier-Daures, M., Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment. Environmental toxicology and chemistry / SETAC 2010, 29, (10), 2249-59.

16. Ciliberti, A.; Chaumot, A.; Recoura-Massaquant, R.; Chandesris, A.; François, A.; Coquery, M.; Ferréol, M.; Geffard, O., Caged Gammarus as biomonitors identifying thresholds of toxic metal bioavailability that affect gammarid densities at the French national scale. Water Research 2017, 118, 131-140.

17. Jubeaux, G.; Simon, R.; Salvador, A.; Quéau, H.; Chaumot, A.; Geffard, O., Vitellogenin-like proteins in the freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males. Aquatic toxicology (Amsterdam, Netherlands) 2012, 112-113, 72-82.

18. Trapp, J.; Geffard, O.; Imbert, G.; Gaillard, J. C.; Davin, A. H.; Chaumot, A.; Armengaud, J., Proteogenomics of gammarus fossarum to document the reproductive system of amphipods. Molecular and Cellular Proteomics 2014, 13, (12), 3612-3625.

19. Minier, C.; Levy, F.; Rabel, D.; Bocquené, G.; Godefroy, D.; Burgeot, T.; Leboulenger, F., Flounder health status in the Seine Bay. A multibiomarker study. Marine environmental research 2000, 50, (1-5), 373-377.

20. Kerambrun, E.; Henry, F.; Marechal, A.; Sanchez, W.; Minier, C.; Filipuci, I.; Amara, R., A multibiomarker approach in juvenile turbot, Scophthalmus maximus, exposed to contaminated sediments. Ecotoxicology and Environmental Safety 2012, 80, 45-53.

21. Chalghmi, H.; Zrafi, I.; Gourves, P.-Y.; Bourdineaud, J.-P.; Saidane-Mosbahi, D., Combined effects of metal contamination and abiotic parameters on biomarker responses in clam Ruditapes decussatus gills: an integrated approach in biomonitoring of Tunis lagoon. Environmental Science: Processes & Impacts 2016, 18, (7), 895-907.

22. Serafim, A.; Lopes, B.; Company, R.; Cravo, A.; Gomes, T.; Sousa, V.; Bebianno, M. J., A multi-biomarker approach in cross-transplanted mussels Mytilus galloprovincialis. Ecotoxicology 2011, 20, (8), 1959-1974.

23. Trapp, J.; Armengaud, J.; Gaillard, J. C.; Pible, O.; Chaumot, A.; Geffard, O., High-throughput proteome dynamics for discovery of key proteins in sentinel species: Unsuspected vitellogenins diversity in the crustacean Gammarus fossarum. J Proteomics 2016, 146, 207-14.

24. Scott, A. P.; Sanders, M.; Stentiford, G. D.; Reese, R. A.; Katsiadaki, I., Evidence for estrogenic endocrine disruption in an offshore flatfish, the dab (Limanda limanda L.). Mar Environ Res 2007, 64, (2), 128-48.

25. Jobling, S.; Tyler, C. R., Endocrine disruption in wild freshwater fish. In Pure and Applied Chemistry, 2003; Vol. 75, p 2219.

26. Kirby, M. F.; Allen, Y. T.; Dyer, R. A.; Feist, S. W.; Katsiadaki, I.; Matthiessen, P.; Scott, A. P.; Smith, A.; Stentiford, G. D.; Thain, J. E.; Thomas, K. V.; Tolhurst, L.; Waldock, M. J., Surveys of plasma vitellogenin and intersex in male flounder (Platichthys flesus) as measures of endocrine disruption by estrogenic contamination in United Kingdom estuaries: temporal trends, 1996 to 2001. Environ Toxicol Chem 2004, 23, (3), 748-58.

27. Xuereb, B.; Bezin, L.; Chaumot, A.; Budzinski, H.; Augagneur, S.; Tutundjian, R.; Garric, J.; Geffard, O., Vitellogenin-like gene expression in freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization in females and potential for use as an endocrine disruption biomarker in males. Ecotoxicology (London, England) 2011, 20, (6), 1286-99.

28. Matozzo, V.; Gagne, F.; Marin, M. G.; Ricciardi, F.; Blaise, C., Vitellogenin as a biomarker of exposure to estrogenic compounds in aquatic invertebrates: a review. Environ Int 2008, 34, (4), 531-45.

29. Short, S.; Yang, G.; Kille, P.; Ford, A. T., Vitellogenin is not an appropriate biomarker of feminisation in a crustacean. Aquatic toxicology (Amsterdam, Netherlands) 2014, 153, 89-97.

30. Yang, M.; Wang, Y.; Jiang, F.; Song, T.; Wang, H.; Liu, Q.; Zhang, J.; Zhang, J.; Kang, L., miR-71 and miR-263 Jointly Regulate Target Genes Chitin synthase and Chitinase to Control Locust Molting. PLoS Genetics 2016, 12, (8).

31. Rocha, J.; Garcia-Carreño, F. L.; Muhlia-Almazán, A.; Peregrino-Uriarte, A. B.; Yépiz-Plascencia, G.; Córdova-Murueta, J. H., Cuticular chitin synthase and chitinase mRNA of whiteleg shrimp Litopenaeus vannamei during the molting cycle. Aquaculture 2012, 330–333, 111-115.

32. Pesch, Y. Y.; Riedel, D.; Behr, M., Drosophila Chitinase 2 is expressed in chitin producing organs for cuticle formation. Arthropod Structure and Development 2017, 46, (1), 4-12.

33. Guan, J.; Lv, Y.; Zhang, Y.; Ning, Q., A shortening effect of KK-42 on the moult cycle of juvenile (De Haan, 1849) (Decapoda, Palaemonidae). Crustaceana 2016, 89, (1), 85-95.

34. Olsvik, P. A.; Samuelsen, O. B.; Agnalt, A. L.; Lunestad, B. T., Transcriptional responses to

teflubenzuron exposure in European lobster (Homarus gammarus). Aquatic Toxicology 2015, 167, 143-156.

35. Poynton, H. C.; Varshavsky, J. R.; Chang, B.; Cavigiolio, G.; Chan, S.; Holman, P. S.; Loguinov, A. V.; Bauer, D. J.; Komachi, K.; Theil, E. C.; Perkins, E. J.; Hughes, O.; Vulpe, C. D., Daphnia magna ecotoxicogenomics provides mechanistic insights into metal toxicity. Environmental Science and Technology 2007, 41, (3), 1044-1050.

36. Luo, J.; Pei, S.; Jing, W.; Zou, E.; Wang, L., Cadmium inhibits molting of the freshwater crab Sinopotamon henanense by reducing the hemolymph ecdysteroid content and the activities of chitinase and N-acetyl-beta-glucosaminidase in the epidermis. Comp Biochem Physiol C Toxicol Pharmacol 2015, 169, 1-6.

37. Gismondi, E.; Thomé, J. P.; Urien, N.; Uher, E.; Baiwir, D.; Mazzucchelli, G.; De Pauw, E.; Fechner, L. C.; Lebrun, J. D., Ecotoxicoproteomic assessment of the functional alterations caused by chronic metallic exposures in gammarids. Environmental Pollution 2017, 225, 428-438.

38. Fan, L.; Wang, A.; Miao, Y.; Liao, S.; Ye, C.; Lin, Q., Comparative proteomic identification of the hepatopancreas response to cold stress in white shrimp, Litopenaeus vannamei. Aquaculture 2016, 454, 27-34.

39. Lee, S. O.; Jeon, J. M.; Oh, C. W.; Kim, Y. M.; Kang, C. K.; Lee, D. S.; Mykles, D. L.; Kim, H. W., Two juvenile hormone esterase-like carboxylesterase cDNAs from a Pandalus shrimp (Pandalopsis japonica): cloning, tissue expression, and effects of eyestalk ablation. Comp Biochem Physiol B Biochem Mol Biol 2011, 159, (3), 148-56.

40. Reddy, P. R.; Nagaraju, G. P. C.; Reddy, P. S., Involvement of Methyl Farnesoate in the Regulation of Molting and Reproduction in the Freshwater Crab Oziotelphusa senex senex. Journal of Crustacean Biology 2004, 24, (3), 511-515.

41. Zhang, X.; Zheng, S., 20-hydroxyecdysone enhances the expression of the chitinase 5 via Broad-Complex Zinc-Finger 4 during metamorphosis in silkworm, Bombyx mori. Insect Molecular Biology 2017, 26, (2), 243-253.

42. Tao, T.; Xie, X.; Liu, M.; Jiang, Q.; Zhu, D., Cloning of two carboxylesterase cDNAs from the swimming crab Portunus trituberculatus: Molecular evidences for their putative roles in methyl farnesotae degradation. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 2017, 203, 100-107.

43. Lucu, Č.; Towle, D. W., Na++K+-ATPase in gills of aquatic crustacea. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 2003, 135, (2), 195-214.

44. Dayras, P.; Charmantier, G.; Chaumot, A.; Vigneron, A.; Coquery, M.; Quéau, H.; Artells, E.; Lignot, J. H.; Geffard, O.; Issartel, J., Osmoregulatory responses to cadmium in reference and historically metal contaminated Gammarus fossarum (Crustacea, Amphipoda) populations. Chemosphere 2017, 180, 412-422.

45. Harris, R. R.; Santos, M. C. F., Heavy metal contamination and physiological variability in the Brazilian mangrove crabs Ucides cordatus and Callinectes danae (Crustacea: Decapoda). Marine Biology 2000, 137, (4), 691-703.

46. Reddy, A. N.; Venugopal, N. B.; Reddy, S. L., Effect of endosulfan 35 EC on ATPases in the tissues of a freshwater field crab Barytelphusa guerini. Bull Environ Contam Toxicol 1992, 48, (2), 216-22.

47. Comoglio, L.; Amin, O.; Roque, A.; Betancourt-Lozano, M.; Anguas, D.; Haro, B. M., Evaluation of sublethal biomarkers in Litopenaeus vannamei on foodborne exposure to methyl parathion. Ecotoxicol Environ Saf 2005, 62, (1), 66-74.

48. Tkatcheva, V.; Poirier, D.; Chong-Kit, R.; Furdui, V. I.; Burr, C.; Leger, R.; Parmar, J.; Switzer, T.; Maedler, S.; Reiner, E. J.; Sherry, J. P.; Simmons, D. B. D., Lithium an emerging contaminant: Bioavailability, effects on protein expression, and homeostasis disruption in short-term exposure of rainbow trout. Aquatic Toxicology 2015, 161, 85-93.

49. Lacaze, E.; Devaux, A.; Jubeaux, G.; Mons, R.; Gardette, M.; Bony, S.; Garric, J.; Geffard, O., DNA damage in Gammarus fossarum sperm as a biomarker of genotoxic pressure: intrinsic variability and reference level. Science of the Total Environment 2011, 409, (17), 3230-3236.

Supporting Information

Protein extraction and trypsin digestion

Reagents and chemicals - Acetonitrile, methanol and water (LC-MS grade) were obtained from Fisher Scientist (Strasbourg, France). Dithiothreitol (DTT), iodoacetamide (IAM), formic acid (FA) (LC–MS grade), trypsin (treated TCPK from bovine Pancreas and not treated type 1X-S from porcine pancreas), urea and TRIS, EDTA, Triton X, sodium chloride, leupeptin and aprotinin were purchased from Sigma–Aldrich (St Quentin-Fallavier, France). Isotopically labelled peptides containing either a C terminal [$^{13}C_6$, $^{15}N_2$] lysine or arginine were synthesized by Fisher Scientist (Strasbourg, France). Absolute ethanol and ethylic ether were obtained from Carlo Erba (Val de Reuil, France).

The procedure used for total protein extraction was adapted from Simon et al. (Simon et al., 2010). Whole-body organisms were homogenized in Tris Buffer (Tris 50mM, 100mM NaCl, 0.01 mM EDTA, 0.1% v/v Triton X-100) pH 7.8 plus 10 µg/L of each leupeptin and aprotinin, with a bead mill homogenizer. For each organism, the volume of buffer was adapted according to their weight (25 µL buffer/mg male; 50 µL buffer/mg female). The homogenates were centrifuged at 10,000 g at 4 °C for 15 min, and 250 µL of clear supernatant collected to new tubes. 750 µL of an ethanol/diethyl ether delipidating solution (1:1, v/v) were added to the supernatant, and the mixture was vortexed and incubated on ice for 10 minutes. After a 10 minute centrifugation at 10,000 g, the resulting supernatants were removed, and the pellets resuspended in 250 µL Tris Buffer.

The protein extracts were then treated with 3 mL of ammonium bicarbonate (AMBIC) 50mM and 362 µL of dithiothreitol (DTT) 150 mM for 40 minutes at 60 °C. After cooling to room temperature, samples were incubated in the dark with iodoacetamide (final concentration of 15 mM), during 40 minutes at room temperature. 150 µL of a 2 mg/mL TPCK treated trypsin solution prepared in 50 mM of AMBIC were added, and the samples incubated for 60 minutes at 37 °C. 20µL of formic acid (FA) were added to stop the trypsin reaction. 3 mL of the obtained peptide solution (plus 10 µL of the isotopically labelled peptides solution at 1 µg/mL) were then purified in Oasis HLB 3 cc solid phase extraction columns (60 mg). After peptide elution with 1 mL of methanol/0.5 % FA, 100 µL of a glycerol 10%/methanol were added, and the samples evaporated under a flow of nitrogen. The pellet was then resuspended in 90 µL of a H_2O/ACN (90/10) + 0.1% FA solution.

Liquid chromatography and mass spectrometry

LC–MS/MS analysis was performed on an HP1200 series HPLC device (Agilent Technologies, Waldbronn, Germany) coupled to a QTRAP® 5500 LC/MS/MS System hybrid triple quadrupole/linear ion trap mass spectrometer (Applied Biosystems/ MDS Analytical Technologies, Foster City, CA, USA) equipped with a Turbo VTM ion source. The LC separation of the 20µL injected sample was carried out on an Xbridge $C6_8$ column (100mm×2.1mm, particle size 3.5 µm) from Waters (Milford, MA, USA). Elution was performed at a flow rate of 300µL/min with water containing 0.1% (v/v) formic acid as eluent A and acetonitrile containing 0.1% (v/v) formic acid as eluent B, employing a linear gradient from 2% B to 33% B in 19 min, followed by a second linear gradient from 33% B to 64% B in 6 min. Then, column washing and re-equilibration was performed for 6 min. The injection duty cycle was 35 min, taking into account the column equilibration time. Instrument control, data acquisition and processing were performed using the Analyst 1.5 software. The mass spectrometer was initially tuned and calibrated using polypropylene glycol, reserpine and Agilent Tuning Mix (all Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. MS analysis was carried out in positive ionization mode using an ion spray voltage of 5500 V. The nebulizer and the curtain gas flows were set at 50 psi using nitrogen. The Turbo V™ ion source was set at 550 °C with the auxiliary gas flow (nitrogen) set at 40 psi. The SRM transitions were monitored and acquired at unit resolution in Q1 and Q3, with a dwell time of 10 ms used for each transition, to obtain 10 data points per chromatographic peak minimum.

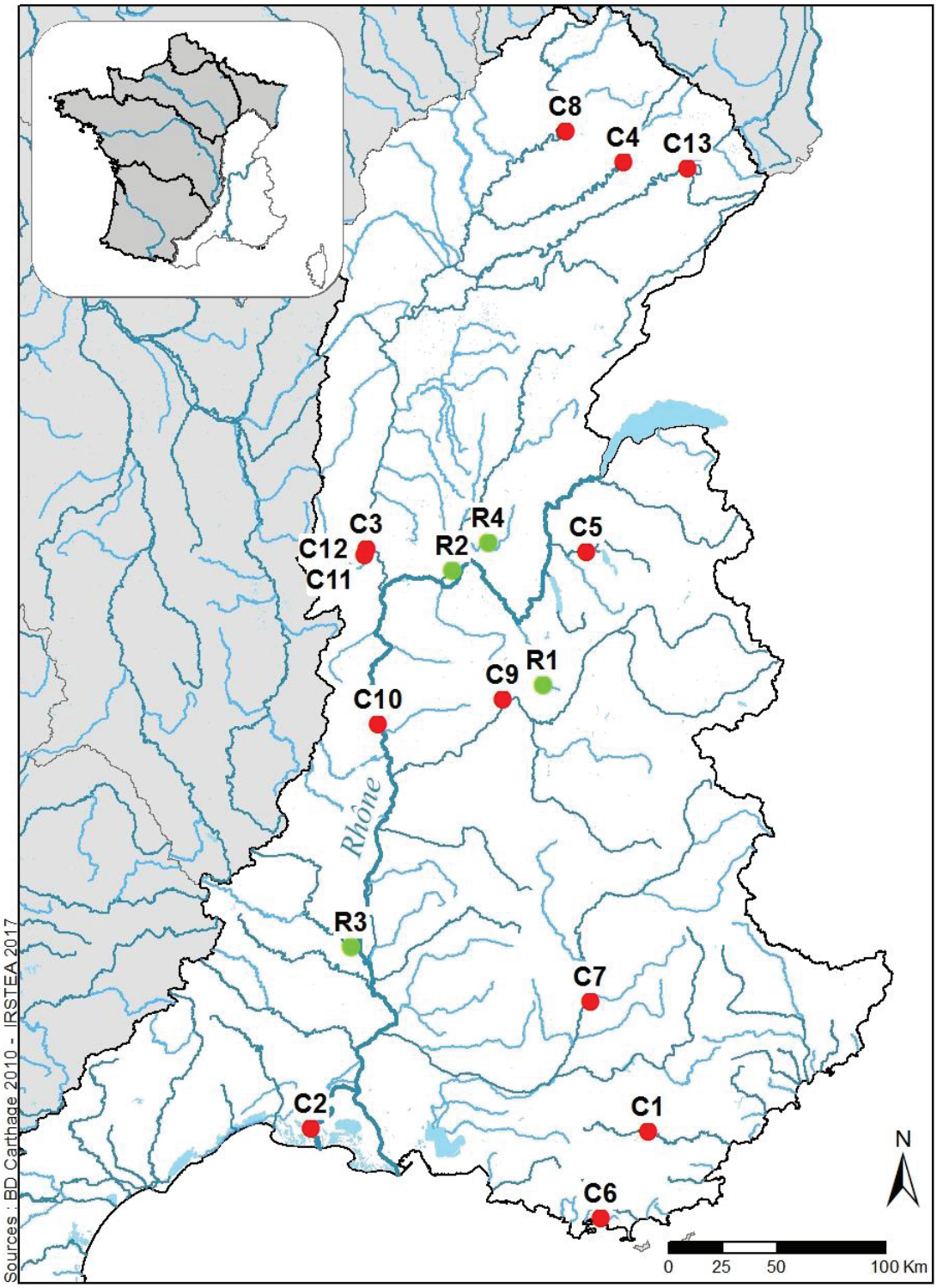


Figure S1 - Map representation of the study sites in which the gammarids were exposed.

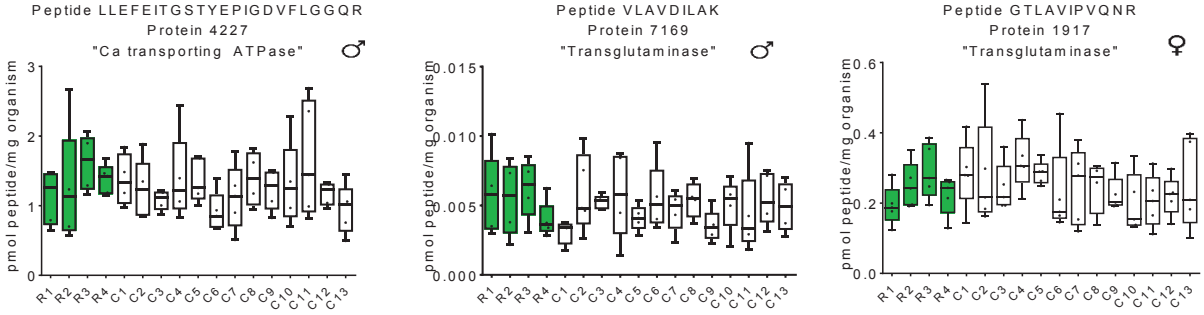


Figure S2 - Quantification results in pmol peptide/mg organism of three non-responsive peptide biomarkers.

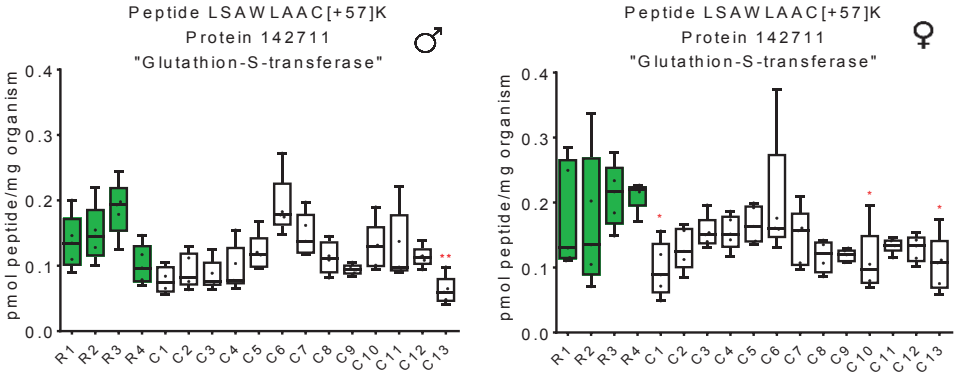


Figure S3 - Quantification results in pmol peptide/mg organism of one GST signature peptide (from protein 142711) in exposed male and female organisms.

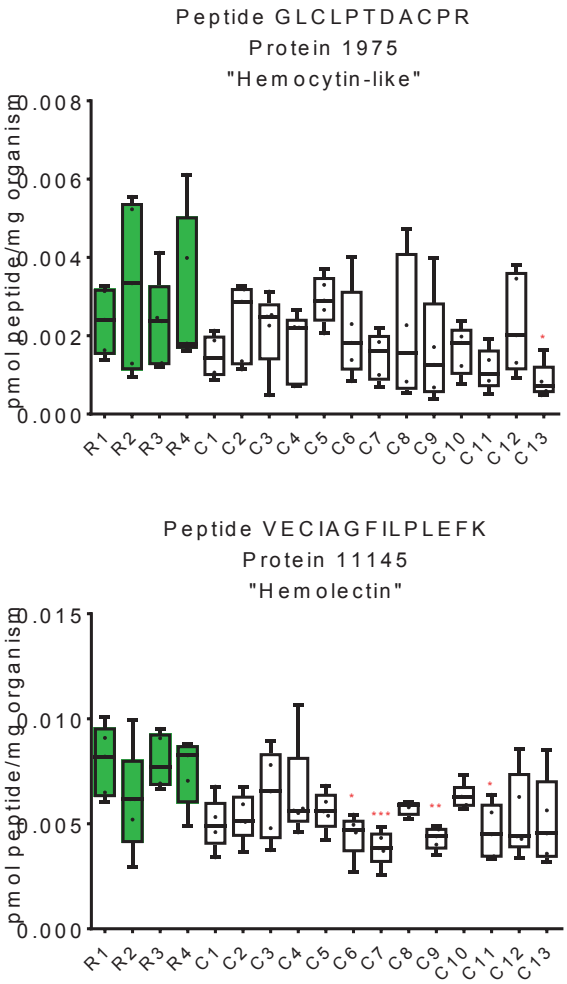


Figure S4 - Quantification results in pmol peptide/mg organism of two immune-related proteins in exposed female organisms.

Table S1

Peptide Modified Sequence	Protein ID	Functional Annotation	Peptide MW (Da)	Functional Group
HIEIFSPITK	64	Vitellogenin	1184,38	Female yolk proteins
VVPSLSAEDTLSQR	276	Hemolymph clottable protein	1501,64	
HAEFSVNPPLDSTQAVK	276		1839,92	
IIYPAEALTIVIEK	276		1572,92	
IFNVLQPIAESK	17046	Clottable protein 2	1358,77	
TSEVFLPLTNELYQQTK	17046		2011,24	
ISPLINSPSDLPK	39606	Clotting protein precursor	1380,77	
TLGALELDVFSEAGK	200426	Vitellogenin	1549,81	
ELTSAAEVVSSLLK	200426		1446,81	
AAIETAFVNHLK	206469	Clotting protein precursor	1313,72	
SFILSDLTNLQQSQAPYK	194758	Apolipocrustacein	2053,28	
FVGLISLIDPPR	32234	Na+K+ ATPase α1 subunit	1326,78	Osmoregulation
VIMVTGDHPITAK	110907	Na+/K+ ATPase	1796,98	
LGAIVAVTGDGVNDSPALK	110907		1381,75	
NLAFFSTNAVEGTAR	209438	Na+/K+ ATPase	1597,80	
LQTNPDTGLSTAEAR	209438		1573,78	
GIDIIGDAFEADR	2562	Prophenoloxidase	1391,68	Molt-related
APILEGYFSK	2562		1124,60	
GIDFGTTQSVR	2562		1180,26	
ATQPSYTVAQLELPGVNITR	2562		2158,42	
ILEDFVDVFNR	100255	Cytochrome P450 enzyme, CYP4C39	1366,70	
AFWGSLPLR	144144	JHE-like carboxylesterase 1	1046,58	
ILTTMWADFAR	144144		1324,55	
EFWIATDHNEVR	166723	Farnesoic acid methyltransferase	1516,61	
LVLGTATYGR	181833	Chitinase	1050,21	
GTLAVIPVQNR	1917	Transglutaminase	1167,68	General biomarkers
GLC[+57]LPTDAC[+57]PR	1975	Hemocytin-like	1143,35	
LLEFEITGSTYEPIGDVFLGGQR	4227	Calcium-transporting ATPase sacoplasmic/endoplasmic reticulum type-like isoform 1	1503,76	
VLAVDILAK	7169	Transglutaminase	941,60	
VEC[+57]IAGFILPLEFK	11145	Hemolectin	1578,86	
AVAEIVQDYDSDGFFPALGFGGK	34845	Copine-8	2403,61	
FSESSAILR	40079	Gluthation-S-transferase	1009,53	
LGSNFLQIPVNC[+57]PYR	45375	Catalase	1720,99	
NLPADQAAALASSDPDYAIR	45375		2059,20	
ADPALGQAIQER	110912	Catalase	1268,66	
LADNIAGHVINTQEFIR	110912		1911,13	
LSAWLAAC[+57]K	142711	Gluthation-S-transferase	962,51	
ELFDFADAHR	213317	Cellulase	1220,57	

Table S2

Table S2 is provided as a separate excel file.

Table S3

Table S3 is provided as a separate excel file.

Table S4 - Absence or presence of modulations for the different protein classes in each study site.

Putative functions	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Yolk proteins in males *				x	x								
Molt-related proteins	x	x	x	x		x	x	x				x	x
Ionic pumps	x	x	x	x									
Immunity and general biomarkers	x					x	x		x	x	x		x

* All of the "x" correspond to significant modulations seen for at least one protein associated with the function. Only for the "Yolk proteins in males", inductions seen at C4 and C5 were individual site-specific inductions, and therefore not significant.

CHAPTER IV. DEVELOPMENT OF SPECIFIC BIOMARKERS FOR ENDOCRINE DISRUPTION ASSESSMENT IN *GAMMARUS FOSSARUM*

This chapter assembles the results about the development of biomarkers of specific stressors of great environmental interest, endocrine disorders.

The first section (Publication nº4) describes the use of a new-generation of mass spectrometers for a comparative shotgun proteomic study with control and pyriproxyfen-exposed males. This approach allows mining the proteome and differentially modulated proteins without any *a priori* knowledge.

The second Section (Note 1, Note2, and Publication nº5) proposes an alternative strategy, based on a gene-candidate approach, for discovering key endocrine-related genes. Note 1 describes the literature-based candidate search strategy and consequent identification of the molecular players in crustacean endocrine regulation systems. Note 2 describes the application of a reprotoxicity test carried on in order to select model chemical compounds based on their physiological effects in female gammarids for the posterior comparative gene expression analyses.

Afterwards, publication n°5 presents the results obtained from the study of the identification and relevance of three ecdysone-related genes as biomarkers of endocrine disturbances in female gammarids. By combining sequence homology searches, phylogenetic analysis, and molecular biology techniques, we analyzed the involvement of the three candidate genes in the reproductive physiology of gammarids. Their potential use as biomarkers of endocrine disruption was assessed through their change after exposure to suspected endocrine disruptors compounds.

Publication nº 4

Digging deeper the pyriproxyfen-response of the amphipod *Gammarus fossarum* with the next-generation ultra-high-field Orbitrap analyzer: new perspectives for environmental toxicoproteomics.

Judith Trapp[1,2*], Duarte Gouveia[1,2*] Christine Almunia[2], Olivier Pible[2], Davide Degli Esposti[1], Jean-Charles Gaillard[2], Arnaud Chaumot[1], Olivier Geffard[1#], Jean Armengaud[2#]

[1]Irstea, UR MALY Laboratoire d'écotoxicologie, centre de Lyon-Villeurbanne, F-69625 Villeurbanne, France..
[2]Laboratoire Innovations technologiques pour la Détection et le Diagnostic (Li2D), Service de Pharmacologie et Immunoanalyse (SPI), CEA, INRA, F-30207 Bagnols sur Cèze, France.

*contributed equally and should be considered both as first co-author.

Submitted to "**Frontiers in Environmental Science**".

Abstract: Taxon-specific proteins from sentinel species have been shown to be amongst the most interesting proteins for monitoring biological effects from anthropogenic contaminations. Although shotgun proteomics allows high-throughput analysis of animal proteomes, their comprehensive characterization is difficult to reach due to their high dynamic range. Here, we document the new perspectives of a new generation of mass analyser for environmental toxicoproteomics in terms of discovery of candidate biomarkers in a sentinel animal species of great interest for environmental monitoring but not yet genome sequenced. We performed a label-free comparative proteomic study of testis from the freshwater amphipod *Gammarus fossarum* challenged with 0.5 µg or 50 µg/L pyriproxyfen, a juvenile hormone analogue used as insecticide. While 935 proteins were detected and quantified using a LTQ Orbitrap XL instrument, 4,031 proteins were identified and monitored by the most recent instrument incorporating an ultra-high-field Orbitrap analyzer, namely Q Exactive HF mass spectrometer, with a strictly equivalent mass spectrometry time, proteogenomic methodology and validation pipeline. A comparison of protein abundance between control and exposed organisms to 0.5 µg or 50 µg/L of pyriproxyfen showed that a total of 17 proteins were significantly modulated between each of the two conditions and the control in LTQ, while 53 could be uncovered in QEx. The functional annotation of the novel candidate biomarkers and their specificities are discussed in the light of their possible use for monitoring freshwater quality. The perspectives of next-generation mass spectrometers for environmental toxicoproteomics are debated.

Keywords: *Proteomics, high-resolution tandem mass spectrometer, environmental toxicoproteomics, Gammarus fossarum*

INTRODUCTION

Ecotoxicological assessment is based on expert knowledge of the physiological response of sentinel species and the use of biomarkers to monitor toxic effects of contaminants on these organisms. Clearly, there is today a need for better understanding the molecular mechanisms of homeostasis and response to environmental chemical stress in numerous models, and global omics approaches are contributing to this aim (Monsinjon and Knigge 2007; Garcia-Reyero and Perkins 2011). Proteomics use in environmental toxicology is more and more frequent. Compared to gene expression studies, protein modulation is more likely to be linked to effective impact on the organism phenotype. Proteomics allows exhaustive comparison of proteins from controls and environmentally challenged organisms during laboratory or field exposure. From such comparison, it is possible to identify the modes of action of chemical contaminants and propose new biomarkers for environmental risk assessment. By

now, the majority of proteomics studies in ecotoxicology report alterations of housekeeping and abundant proteins such as the ones involved in ATP supply and maintenance of the cytoskeletal structure, and therefore tends to list déjà-vu proteins as differentially expressed biomarker candidates (Petrak et al., 2008). These observations result principally of i) the use of 2DE-PAGE electrophoresis that favours the detection of highly abundant proteins, and ii) the lack of species-specific protein sequence database that restraint interpretation of mass spectra data in species currently used in environmental monitoring (Trapp, Armengaud et al. 2014). Compared to 2D-PAGE, shotgun proteomics allows a deeper proteome characterization with an increased number of identified proteins, therefore allowing identification of under-represented proteins (Armengaud, 2016). This technique consists in the identification of proteins from complex mixtures by high-throughput tandem mass spectrometry without prior protein separation on 2D-PAGE. For this, the whole protein content is proteolyzed with trypsin; the resulting peptides are resolved by high-performance liquid chromatography (HPLC) and identified with high-throughput tandem mass spectrometry. Once identified, proteins can be easily quantified by counting the number of MS/MS spectra. This strategy does not give the absolute amount of each protein, but allows comparing the relative quantities of each detected protein between conditions if equivalent samples are analyzed.

Approaches combining genomics and proteomics data (*i.e.* proteogenomics) constitute a promising alternative to homology-driven proteomics when working with species belonging to taxa poorly documented in genomic databases (Armengaud, Trapp et al., 2014). Indeed proteogenomics leads to higher rate of protein identification than a decade ago (Nesvizhskii A, 2014). For the rapid proteogenomics identification of protein sequences in eukaryotes, mRNAs are sequenced by next-generation sequencing technology and the resulting sequences are translated into the different possible reading frames, giving a specific protein database where most sequences are erroneous but comprising the exact protein sequences. This customized draft predicted protein sequence database is then used for the interpretation of MS/MS data, allowing discriminating sequences corresponding to real proteins (Armengaud, Trapp et al. 2014).

Recently, we applied such strategy to discover key reproductive proteins from the freshwater crustacean *Gammarus fossarum*, a sentinel species in ecotoxicology. The transcriptome from the reproductive organs was established by RNA-Seq and the proteome was studied using shotgun proteomics (Trapp, Geffard et al. 2014). We first reported the existence of 1,873 proteins certified by mass-spectrometry, including 218 orphans. For uncovering new candidates into *G. fossarum* reproductive process, we performed a comparative analysis of the abundance of proteins in male and female reproductive tissues. We pinpointed 204 sexually-dimorphic proteins. In addition, comparative proteomics of the testis at seven different stages during spermatogenesis revealed proteins potentially involved in male reproduction. We also conducted a proteomic analysis in testis of male *G. fossarum* subjected to three different xenobiotics in order to propose candidates as specific biomarkers for toxicity assays (Trapp, Armengaud et al. 2015). Interestingly, a substantial proportion of the candidates uncovered in these experiments were orphans or highly divergent proteins, thus not accessible by traditional homology-driven proteomics as till now few genome sequences have been established and annotated for crustaceans. Reproductive proteins are highly diversified among the tree of life as a result of sexual selection and are consequently highly species-specific (Swanson and Vacquier 2002). Furthermore, as first evidenced by transcriptomic analysis on *Daphnia pulex*, orphan genes are among the most eco-responsive to environmental challenges (Colbourne, Pfrender et al. 2011). We stressed that studying highly divergent proteins is of utmost importance in environmental toxicology (Trapp, Armengaud et al. 2014), and thus proteogenomics approaches with environmental organisms should flourished in the near future (Armengaud, Trapp et al., 2014).

Without adequate proteome fractionation, shotgun strategies detect preferentially abundant proteins, while species-specific proteins are generally present in low abundance. Increasing the mass spectrometry efforts by multiplying fractions and nanoLC-MS/MS runs is time-consuming and thus limits the number of samples to be treated and compared. To get more data from the same sample the use of next-generation mass spectrometer is an attractive alternative. Novel generations of high-resolution mass spectrometers are periodically released, with improved performances such as resolution and scan speed that improve the detection of low-abundance ions in complex matrices. Deeper coverage of proteomes is impressive with the latest released instrument, exemplified with the detection of a set of 8,526 proteins from human tissue (Wei et al., 2016).

The objective of this study is to document the performance of the latest Orbitrap detector, named ultra-high-field Orbitrap, in the context of environmental ecotoxicology and proteogenomics approaches. Two mass spectrometers currently comprise such detector: the Fusion/Lumos tribrid mass spectrometer and the Q Exactive HF/HF-X hybrid tandem mass spectrometer, both from the Thermo company (Scheltema et al., 2014). Here, we analysed the testis proteome of the amphipod *G. fossarum* for deciphering a larger set of candidate biomarkers of male reproductive impairments potentially induced by endocrine disruptor contamination. For this, we analyzed the proteomes of control organisms and organisms exposed to the insecticide pyriproxyfen with two instruments: the LTQ Orbitrap XL mass spectrometer (first generation of Orbitrap analyzer) and the new Q Exactive HF tandem mass spectrometer (ultra-high-field Orbitrap analyzer). While 935 proteins were validated using the LTQ Orbitrap XL, 4,031 proteins were validated by means of the Q Exactive HF. A comparison of protein abundance between control and exposed organisms showed that 17 and 53 proteins were significantly modulated between the contaminated condition and the control when using the LTQ Orbitrap XL and Q Exactive HF, respectively. These novel identified proteins found in higher abundance in pyriproxyfen-exposed animals could be relevant for environmental toxicology monitoring and may be novel interesting biomarkers of reproductive disorders in *G. fossarum*.

Materials & Methods

Exposure of gammarids to pyriproxyfen and protein extraction

Amphipods *G. fossarum* were sampled from the Bourbre river in mid-eastern France (45°34'10''N; 5°27'33''E). They were collected by kick sampling, and acclimatized to laboratory conditions in 30 L tanks. Male organisms were exposed to pyriproxyfen at 0.5 µg/L and 50 µg/L during two consecutive spermatogenesis cycles (15 days), as previously described (Trapp, Armengaud et al. 2015). For this and for each condition, 35 males were placed into five 500 mL polyethylene beakers (seven animals per beaker), under constant oxygenation. A piece of net (6 x 5 cm) was added as a resting surface into each beaker. Organisms were fed *ad libitum* with conditioned alder leaves (*Alnus glutinosa*). Controls consisted in exposure to 0.005% acetone (solvent) or without solvent. After one week of exposure, new mature females were introduced with males, allowing a second fertilisation by males and thus the beginning of a second spermatogenesis cycle. Over the experiment, exposure media and nutriment leaves were renewed every two days. Survival was checked daily. Seven days after the second fertilisation, the experiment was ended. Male gonads were sampled under stereomicroscopic magnification using fine forceps. The reproductive organs of five animals per condition were analyzed by shotgun proteomics. Each pair of testis was directly dissolved in 40 µL of LDS sample buffer (Invitrogen). After sonication for 1 min (transonic 780H sonicator), samples were boiled for 5 min at 95 °C. A volume of 35 µL of each sample was then subjected to SDS-PAGE for a short electrophoretic migration as described previously (Trapp et al., 2015). The whole protein content from each well was extracted as a single polyacrylamide band, processed, and proteolyzed with trypsin (Roche) using 0.01% ProteaseMAX surfactant (Promega) as described (Hartmann, Allain et al. 2014).

Q Exactive HF and LTQ Orbitrap XL mass spectrometry settings

The peptide mixtures were analyzed in data-dependent mode with two tandem mass spectrometers. The parameters used with the LTQ Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings) have been previously described (Trapp, Almunia et al. 2015). The Q Exactive HF mass spectrometer (Thermofisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings) was operated essentially as described (Klein et al., 2016). In this later case, peptides were first desalted on line on a reverse phase precolumn C18 PepMap 100 column, and then, were resolved with a 90 min gradient of CH_3CN, 0.1% formic acid, at a flow rate of 0.2 µL/min. They were analyzed with a data-dependent Top15 method consisting in a scan cycle initiated with a full scan of peptide ions in the Orbitrap analyzer, followed by high energy collisional dissociation and MS/MS scans on the 15 most abundant precursor ions after. Full scan mass spectra were acquired from *m/z* 350 to 1800 with an Automatic Gain Control (AGC) Target set at $3x10^6$ ions and a resolution of 60,000. MS/MS scan was initiated when the ACG target reached 10^5 ions with a threshold intensity of 83,000 and potential charge states of 2^+ and 3^+ after ion selection performed with a dynamic exclusion of 10 sec. The sample analysis with the LTQ Orbitrap XL used the following tune method including the full scan spectra range from m/z 350 to 1800, the AGC target value set at $5x10^5$, and the mass resolving power set at 30000. The MS2 were

analysed with a minimum signal required set at 15000 and a Top5 method.

Protein identification and spectral count quantitation

MS/MS spectra were assigned to peptide sequences with the Mascot Daemon 2.3.2 search engine (Matrix Science) against the customized RNA-seq-derived database GFOSS described in (Trapp, Geffard et al. 2014). This database contains 1,311,444 putative protein sequences totalling 289,084,257 amino acids. The following parameters were used for MS/MS spectra assignation: full-trypsin specificity, maximum of two missed cleavages, mass tolerances of 5 ppm on the parent ion and 0.5 Da on the MS/MS analysed with the LTQ Obitrap XL and 0.02 Da on the MS/MS analysed with the Q Exactive HF, static modification of carboxyamidomethylated cysteine (+57.0215), and oxidized methionine (+15.9949) and deamidation of asparagine and glutamine as dynamic modifications. All peptide matches with a MASCOT peptide score below a *p* value of 0.05 were filtered and assigned to a protein according to the principle of parsimony. A protein was validated when at least two different peptide sequences were detected. The false discovery rate for protein identification was estimated by employing the decoy search option of MASCOT (Matrix Science) to be <0.1%. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaino, Cote et al. 2013) via the PRIDE partner repository with the dataset identifier PXD002XY?XX

Differential shotgun proteomics data analysis

Four protein datasets were recorded: Control without solvent (T); Control with solvent (TS); pyriproxyfen treatment at 0.5 µg/L and pyriproxyfen treatment at 50 µg/L. Each of these conditions were analysed in quintuplets as separate biological replicates. The Tfold module of the PatternLab program for proteomics version 4.0.0.59 (Carvalho, Lima et al. 2016) was used for normalizing selecting the option "Total Signal" and comparing proteomics datasets based on spectral count, as previously described (Hartmann & Armengaud, 2013). Based on the sum spectral count calculation, selection of proteins included in the analysis was based on the L-stringency parameter set at 0.6 and taking into account proteins detected in at least three of the different biological replicates. The F-stringency parameter was optimized as recommended (Carvalho, Lima et al. 2016). To define the set of proteins whose expression was modulated, the fold change was calculated by comparing spectral counts of each protein dataset versus the control. Proteins satisfying both the fold change and statistical criteria were considered as potential protein signatures of the pyriproxyfen treatments.

Protein annotation and gene ontology classification

Functional annotation was carried out with the open source program DIAMOND (Buchfink, Xie et al. 2015) targeting the database SWISSPROT downloaded on 2016/09/30 to extract proteins homologs. Only alignments with an expected value of ≤0.001 were considered and alignments with a bit score <40 were not reported. The GO annotation was then performed from the data provided by the GO Consortium at the Gene Ontology (GO) website with the web application Amigo. Two levels of GO annotation were extracted, level 1 and level 2.

RESULTS AND DISCUSSION

Strategy for comparing testis proteomes and evaluating pyriproxyfen effects

Two major distinct hormone classes are involved in the endocrine system of arthropods: ecdysteroids known as the molting hormone, and juvenoid hormones. Here, we exposed *G. fossarum* males to the juvenoid analogue pyriproxyfen, which is employed in agriculture as a growth regulator for inhibiting the larval development and maturation of many pest arachnids and insects: ticks, mosquitos, ants, houseflies, and fleas (Sullivan and Goh 2008). In crustaceans, this molecule induces morphological abnormalities and limits growth rate in decapods (LeBlanc 2007), decreases female fecundity and alters sex ratio in branchiopods (Olmstead and LeBlanc 2003; Ginjupalli and Baldwin 2013), and reduces sperm production in amphipods (Trapp, Armengaud et al. 2015). Organisms were exposed during two consecutive spermatogenesis cycles (*i.e.* 2 weeks) to 0.5 µg/L and 50 µg/L of pyriproxyfen and compared with the solvent control. No significant mortality was observed during the exposure, with survival values higher than 85%. For evaluation of the effects of pyriproxyfen in terms of protein content, the testes of five organisms were dissected per condition. Shotgun label-free proteomics was carried out on the testes of individual animals with

two high-resolution instruments: the LTQ Orbitrap XL and the Q Exactive HF hybrid tandem mass spectrometers. The former instrument combines a linear trap (LTQ), an intermediate C-trap and an Orbitrap analyser with extended mass resolution (100,000). In MS mode, the LTQ collects the ion population that are passed on to the C-trap for injection and analyzed in the Orbitrap analyzer. For MS/MS analysis with a top5 strategy, the five most intense precursor ions are selected, isolated and fragmented within the LTQ. The second instrument, the Q Exactive HF, combines a quadrupole mass filter with an intermediate C-trap, and HCD fragmentation cell, and an ultra-high-field Orbitrap analyser. The quadrupole offers high-performance precursor ion selection, switching among different narrow mass ranges that are analysed in a single composite MS spectrum. The quadrupole allows instantaneous isolation and fragmentation of different precursor masses followed by fragment ions analysis by the Orbitrap (Michalski, Damoc et al. 2011). The latest version of such configuration is equipped with an ultra-high-field Orbitrap allowing extended resolution (240,000) or quicker scan cycles at lower resolution. When comparing the global performances of both instruments, a MS/MS cycle can be theoretically achieved within 200 msec for the former while 55 msec are sufficient for the later. Such performances can be only achieved if the reverse phase column is able to deliver enough peptide ions along the whole applied gradient, thus performances should be measured in the light of real proteomic samples.

Evaluation of the number of proteins identified by each mass spectrometer

A total of 1,106,968 MS/MS spectra were recorded with the Q Exactive HF instrument when analysing the 20 samples while a total of 222,436 MS/MS spectra were recorded with the LTQ Orbitrap XL instrument, representing a global increase of 5 fold with the most recent tandem mass spectrometer compared to the previous generation. The proteins identified by means of the LTQ Orbitrap XL and the Q Exactive HF instruments are listed in Supplementary data Table S1 and Table S2, respectively. Table 1 presents the mean values of MS/MS spectra recorded, assigned spectra, unique peptide sequences and proteins identified for the four conditions and both instruments. The number of MS/MS spectra per run either with the LTQ Orbitrap XL instrument or with the Q Exactive HF is relatively identical throughout the analysed samples (relative standard deviation below 20%), thus demonstrating the relative stability of these analytical platforms. The spectra recorded with the two instruments were assigned to peptide sequence with the same proteogenomics database (GFOSS) derived from RNAseq information (Trapp, Geffard et al. 2014). The MS/MS spectrum attribution rate was notably improved with the Q Exactive HF analysis allowing 22 % of the total spectra assigned whereas only 13% matched with a peptide sequence with the LTQ Orbitrap XL data. A total of 241,135 MS/MS spectra were assigned with the former while 28,275 MS/MS spectra were assigned with the latter. This higher MS/MS assigning rate is directly correlated with higher MS/MS spectrum quality. The quicker acquisition rate and quadripole filtering properties of the Q Exactive HF resulted in less spectrum contamination. Because MS/MS quality and acquired MS/MS number are increased, the number of assigned MS/MS spectra with the Q Exactive HF is thus 8.5 fold the one obtained with the LTQ Orbitrap XL instrument. Regarding the number of unique peptide sequences, a higher difference was observed with a total of 3,103 non-redundant peptide sequences identified with the LTQ Orbitrap XL, whereas 14,818 items were identified by means of the Q Exactive HF. As shown in Figure 1 (Panel A), a total of 2,961 peptides were in common with both datasets.

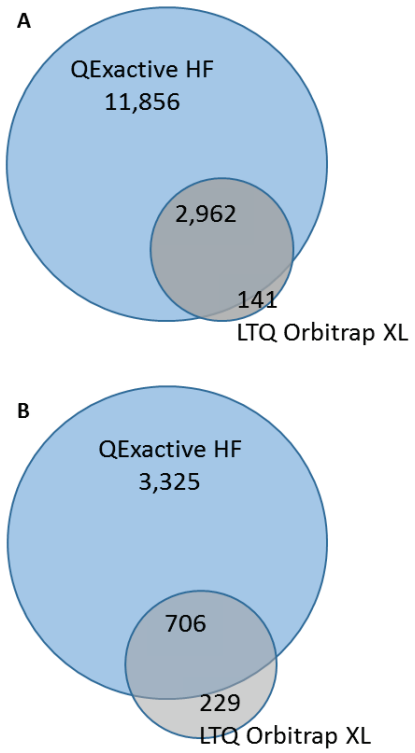


Figure 1: **Venn Diagrams of identified peptides and proteins by both mass spectrometry platforms.** Panel A. Numbers of peptide sequence-spectra matches. Panel B. Numbers of proteins identified with at least 2 peptides.

The proteome sequence coverage resulted much higher with the Q Exactive HF instrument compared to the LTQ Orbitrap XL platform (x4.8). In terms of proteins identified, a total of 935 proteins were validated with at least two unique peptides using the LTQ Orbitrap XL while 4,031 were confirmed using the Q-Exactive HF.

Table 1. Summary and statistics of *G. fossarum* proteome datasets.

		LTQ Orbitrap XL	Q Exactive HF						
		Control 1	Control 2	Treatment 1	Treatment 2	Control 1	Control 2	Treatment 1	Treatment 2
		No-Treatment	Solvant Treatment	Exposed 0.5 µg/L	Exposed 50 µg/L	No-Treatment	Solvant Treatment	Exposed 0.5 µg/L	Exposed 50 µg/L
MS/MS spectra	Mean[a]	11045	11804	11708	11522	59405	55769	56442	55333
	RSD[b]	4%	3%	8%	7%	6%	7%	17%	12%
Assigned spectra	Mean[a]	1446	1510	1566	1538	12861	12265	12176	12298
	RSD[b]	13%	15%	21%	20%	8%	10%	19%	17%
Unique peptides	Mean[a]	940	986	1000	1001	5428	5453	4998	5328
	RSD[b]	4%	11%	20%	23%	5%	7%	17%	23%
Identified proteins	Mean[a]	225	315	307	299	1440	1515	1309	1305
	RSD[b]	2%	7%	13%	18%	9%	6%	14%	21%

[a]Mean: average value taking into account the five biological replicates
[a]RSD: relative standard deviation

At the protein level, the Q Exactive HF enabled the discovery of 3,325 additional proteins, as 706 were in common with the LTQ Orbitrap XL analysis and 229 were detected by the LTQ Orbitrap XL only (Figure 1, Panel B). Thus, the Q Exactive HF allows recording almost five fold more peptides and validating five fold more proteins than the LTQ Orbitrap XL operated with the same gradient length. These increased performances can be explained by the faster scan rate (20Hz with the Q Exactive HF as compared to 4Hz with the LTQ Orbitrap XL) allowed by the ultra-high-field Orbitrap analyzer and the configuration of the instrument incorporating a quadrupole rather than a linear trap (Michalski, Damoc et al. 2011).

Proteome features imprinted by each mass spectrometer

Table 2 lists for the two mass spectrometers the most abundant proteins among the *G. fossarum* testis proteome when merging the twenty individual proteomes. These fourteen proteins represent roughly the fourth of the total proteome in terms of abundance as estimated based on their normalized spectral abundance factor (NSAF). Most of these proteins exhibit general housekeeping functions such as cytoskeletal structure maintenance (actin and myosin) and energy metabolism homeostasis (arginine kinases for maintenance of ATP level). Preponderance of histones, arginine kinase, and immunity proteins (hemocyanin like) in the *G. fossarum* testis proteome was already reported (Trapp, Armengaud et al. 2015). Strikingly, all the components of this short-list of most abundant proteins were identified with both analytical systems. Moreover, their relative quantities were estimated roughly in the same range as shown in Table 2. However, the fourteen most abundant proteins represent altogether almost 29% of the total proteome in terms of abundance when measured with the LTQ Orbitrap XL, but only 13% when proteins were quantified by the Q Exactive HF instrument. This difference is directly linked to the higher dynamic range of the later instrument, and thus its proteome coverage. When accounting only the fourteen most abundant proteins for calculating the NSAF factors (NSAF-14), the ratio of each protein is relatively equivalent with both analytical systems (Table 2).

Figure 2 shows the distribution of the number of non-redundant peptides according to protein abundance. For this, all detected proteins were ranked by quintile according to their abundance as measured by their NSAF. As expected, the most abundant proteins are detected through many different peptide sequences, while low abundant proteins are mostly detected with a few peptides. This classification was done for both datasets. As indicated in Figure 2, the mean ratio of peptide sequences is roughly comparable for each of the five classes of proteins for both analytical platforms. However, the numbers of proteins detected by each analytical system are quite different. We noted that the distribution of the number of peptides per proteins is slightly more

variable for the Q exactive HF, a phenomenon also illustrated by outliers represented by empty circles in Figure 2. This observation is especially true for low abundant proteins and can explain the increased variability of data recorded with this instrument.

Table 2. List of the ten most abundant proteins in global proteome identified by the LTQ Orbitrap XL or by the Q Exactive HF instruments.

Translated contig ID	BLAST E-value	Probable function	Origin of the closest homologue	NSAF (%) (total proteome)		NSAF-14 (%) (14 most abundant proteins)	
				LTQ Orbitrap XL	Q Exactive HF	LTQ Orbitrap XL	Q Exactive HF
49159_fr2	1.2E-62	actin-5, muscle-specific-like	*Hyalella azteca*	4.01%	1.35%	14.1%	10.52%
34079_fr5	3.8E-66	myosin heavy chain, muscle-like isoform X15	*Hyalella azteca*	3.85%	0.98%	13.5%	7.59%
14_fr3	9.1E-124	myosin heavy chain, muscle-like isoform X3	*Hyalella azteca*	3.52%	1.75%	12.4%	13.62%
14404_fr1	3.1E-106	arginine kinase	*Hyalella azteca*	2.95%	1.62%	10.3%	12.54%
37276_fr6	1.4E-70	myosin heavy chain, muscle-like isoform X17	*Hyalella azteca*	2.32%	1.15%	8.1%	8.91%
20975_fr4	1.8E-36	myosin light chain 1	*Procambarus clarkii*	2.00%	1.07%	7.0%	8.30%
3149_fr2	2.1E-69	myosin heavy chain type b	*Marsupenaeus japonicus*	1.73%	0.87%	6.1%	6.76%
105020_fr5	3.4E-122	myosin heavy chain, muscle-like isoform X16	*Hyalella azteca*	1.51%	1.22%	5.3%	9.46%
195102_fr4	5.6E-71	myosin heavy chain, muscle-like isoform X18	*Hyalella azteca*	1.35%	0.37%	4.7%	2.89%
100349_fr3	7E-53	histone core	*Cynara cardunculus*	1.24%	0.97%	4.3%	7.52%
49126_fr2	4.8E-62	actin, cytoplasmic 2 isoform X3	*Nomascus leucogenys*	1.06%	0.19%	3.7%	1.49%
8190_fr1	3.5E-120	myosin heavy chain, muscle-like isoform X19	*Hyalella azteca*	1.02%	0.31%	3.6%	2.43%
170702_fr2	1.1E-69	sarcoplasmic calcium-binding protein 1- isoform X1	*Hyalella azteca*	0.99%	0.54%	3.5%	4.17%
202279_fr1	4.1E-167	hemocyanin-like	*Hyalella azteca*	0.97%	0.49%	3.4%	3.80%

Pereira-Medrano et al. (2007) already evidenced some years ago that proteins at low concentration produce more variable results in terms of protein confidence and coverage. Indeed, the MS-based proteomics discovery experiments are exhaustive, but proteins are inconsistently observed across samples because of the stochastic sampling of complex proteomes (Gillette and Carr 2013). This effect is exacerbated with low abundant and/or small proteins since the detection of an additional peptide considerably increases the protein validation threshold and coverage, but also its quantitation by spectral count.

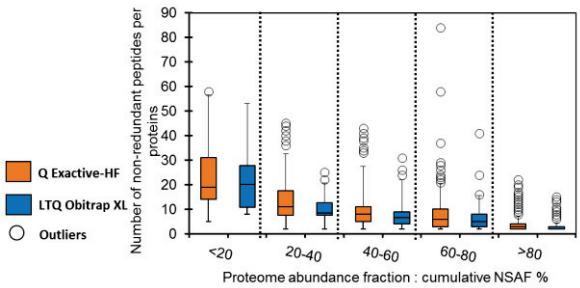


Figure 2: Distribution of the number of non-redundant peptides per protein according to their abundance in the whole proteome. Boxes extend from the first to the third quartile with a bold segment for the median value; the whiskers extend to the most extreme data points, no more than 1.5 times the interquartile range. Data points of more than 1.5 times the interquartile range are represented by empty circles (outliers). The Q-Exactive HF and the LTQ-Orbitrap XL datasets are indicated in orange and blue, respectively.

A total of 4,261 different translated contigs/polypeptides were identified when merging all the data. Figure 3 shows the saturation curve in terms of detection of proteins when proteins are ranked by their relative abundance. The cumulated percentage of NSAF is calculated for all proteins ranked from the most abundant to the least abundant. Here, 50 % of the proteome content was reached with the 53 most abundant proteins with the LTQ Orbitrap XL data, and 159 with the Q Exactive HF data. For 80 % of the proteome, 225 proteins were merged for the previous analytical platform while 692 proteins should be cumulated for the latter. As shown in Figure 3, the Q Exactive HF platform outperforms the LTQ Orbitrap XL platform for detecting the low abundant proteins. The 20 % less abundant proteome represents only 281 with the LTQ Orbitrap instrument, while 3,341 proteins could be identified with the Q Exactive HF. It is thus clear that the new generation of mass spectrometer offers a better identification of low abundant proteins with enhanced discovery capacities, without major detrimental changes in terms of protein sequence coverage.

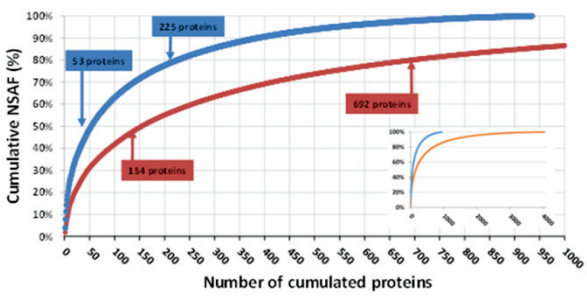


Figure 3: Saturation curve in terms of number of identified proteins. Proteins detected by the Q Exactive HF (orange) or LTQ Orbitrap XL (Blue) were ranked from the most to the least abundant (x axis). Their cumulated abundance (as assessed by their NSAF) is represented in terms of percentage of the global proteome (y axis). The insert presents the same figure with the fully extended x axis.

Systematic functional annotation of identified proteins.

Functional annotation of the whole dataset of 4,261 proteins identified by the shotgun approach was performed by means of cross-species prediction obtained with sequence similarity searches. A total of 3,606 proteins could be functionally annotated because relatively similar sequences were found (threshold E-value set at $1E^{-10}$). Table S1 and Table S2 list the 809 and 3,430 annotated proteins that were recorded with the LTQ Orbitrap XL and the Q Exactive HF instruments, respectively, and their GO term assignation. Although the proteomic analysis was carried out on testis, the GO analysis evidenced few biological processes or molecular functions more specifically related to the reproductive function, highlighting a lack of molecular knowledge regarding crustacean proteins. Only four proteins of the testis proteome were identified by GO term as involved in the biological process of reproduction: Contig_Gammarus_90_85071_fr3 which encodes a Laminin-like protein epi-1 with strong similarities with the *Caenorhabditis elegans* Q21313.1 protein homologs, the Contig_Gammarus_90_179325_fr3 (serine protease inhibitor 42Dd), the Contig_Gammarus_90_12305_fr5 (also a serine protease) and the Contig_Gammarus_90_98198_fr6 (Papilin homolog). These four proteins were estimated as low abundant. Automatic functional prediction based on sequence similarities remains challenging for non-model species distantly related from well characterized genomes. This is especially true for amphipods because of the lack of genomic sequences in public databases for these animals and above all, because of scarce knowledge on molecular functions carried out by most gene products in arthropods while this ancient branch of Life is really diverse and abundant. Recently, a first amphipod annotated RefSeq tentative genome has been released based on yet unpublished *Hyalella azteca* transcriptomics data (GCF_000764305.1). Unfortunately, this tentative genome comprises 18,000 scaffolds leading to a partial and dispersed view of *Hyalella* patrimony. Despite of this novel information, we noted that 20 % of the proteins identified in the reproductive organs from *G. fossarum* males could not be annotated by GO term. This lack of knowledge is especially true for reproductive proteins for which a trend to be species-specific as a result of sexual selection has been reported (Swanson and Vacquier 2002). In our first proteogenomics survey of *G. fossarum*, we already evidenced that there were more orphan proteins in reproductive organs than in other body parts (Trapp, Geffard et al. 2014).

Identification of molecular markers triggered by pyriproxyfen exposure

First, we compared the different proteome datasets acquired with the Q Exactive HF instrument. Supplementary Table S2 presents the protein abundance fold changes and their associated p-value confidence when comparing i) the control with solvent against the control without solvent, ii) the pyriproxyfen 0.5 µg/L versus the control with solvent, and iii) the pyriproxyfen 50 µg/L versus the control with solvent.

Taking a stringent TFold parameter of at least two (corresponding to a fold change equal 2), we noted a significant effect of the solvent which induced the expression increased for 28 proteins and decreased for 4 proteins. Among the proteins up regulated, most have a direct (Myosin, Collagen, Papilin, Gelsolin, Plastin) or indirect role (Glucosylceramidase by its hydrolase activity) in cell structure maintenance. A significant number of proteins are involved in oxidative stress response (Chromate reductase, Aldose reductase, Catalase, Glutathione S-transferase 1, Hydroxysteroid dehydrogenase-like protein 2). Here, proteins in lower abundance are known to be circulating proteins in haemolymph like Hemocyanins and Haemolymph clottable proteins. Altogether, these data highlighted an effect of the solvent (acetone) on protein profiles in the *G. fossarum* testis, even at a concentration usually considered as very low (0.005% v/v).

In order to assess the effects of Pyr on protein expression, taking into account the presence of the solvent, the low dose and the high dose groups were compared with the solvent control using the same fold change cut-off. In comparison with the solvent control, a set of 6 proteins were found more abundant and 26 were found less abundant at low pyriproxyfen dose (0.5 µg/L). At high pyriproxyfen dose (50 µg/L), 6 proteins were found in higher quantity and 15 in lower abundance. Figure 4 shows a clustering analysis of these results highlighting the differences between the two pyriproxyfen doses and the control. Notably, in our experimental conditions pyriproxyfen exposure is associated with a molecular signature characterized by few modifications of the proteomic profile in *Gammarus* testes. However, amongst these proteins, five were found modulated with the same variation at both doses (one upregulated, annotated as haemocyanin and the other four downregulated in pyriproxifen treated samples, two of which are homologous of not characterized Hyallela proteins). (Supplementary Table S2). Globally, proteins over- expressed in the control with solvent were under-expressed in sample

treated with pyriproxyfen and vice versa. For example, Hemocyanin, a protein circulating in haemolymph in charge of oxygen transport but also an important immune effector in invertebrates, was found in lower amounts in the control with solvent, but highly induced at the two pyriproxyfen doses, highlighting pyriproxifen may induce a stress response at both the tested doses.

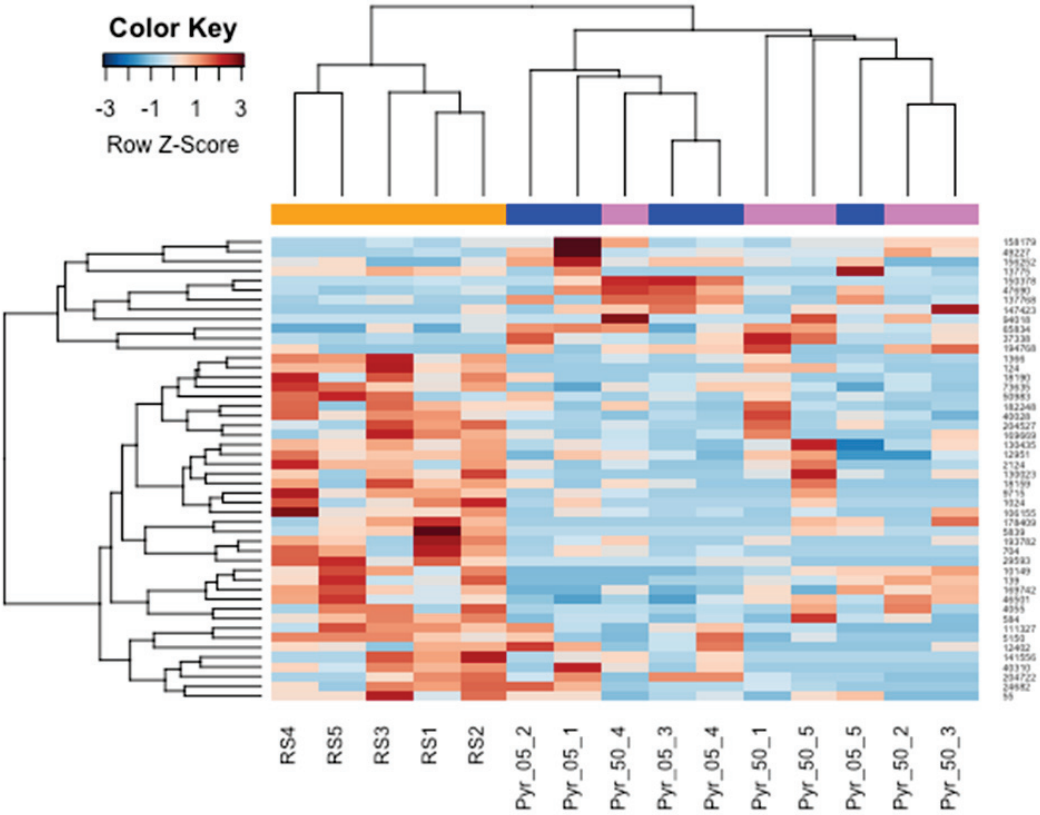


Figure 4: Hierarchical clustering of modulated proteins upon pyriproxyfen stress. Controls with solvent are labelled in orange, samples treated with PYR 0.5 in blue and PYR 50 in pink.

With the treatment at 0.5 µg/L, a set of 32 proteins were modulated with 6 proteins more abundant and 26 less abundant. Several metabolic enzymes were found as induced (Glycogen phosphorylase, Succinate dehydrogenase, L-lactate dehydrogenase) while closely related in term of function, *i.e.* glycogenesis and cellular respiration. Glycogen phosphorylase (Contig_Gammarus_90_147423_fr4) is found up-modulated (from 5.3 to 7.5 fold change, q-value< 0.05) after treatment with 0.5 µg/L of pyriproxyfen.Glycogen has been suggested as the main fuel oxidized in G. fossarum and other crustaceans (Hervant, 1996). Interestingly, in another crustacean species (*Artemica sinica*), glycogen phosphorylase expression was found to be stress-related (Zhao et al. 2014) . In addition, decrease of the pyruvate carboxylase abundance revealed reduction of pyruvate cycling in accordance with the higher abundances of the previously discussed metabolic enzymes. Furthermore, large cellular and organ cytoskeleton re-organization could be suspected through strong decrease in abundance of several proteins involved in such process (Ubiquitin-like modifier-activating enzyme, Myosin, Ezrin, Zasp-like isoform, Collagen alpha-2 chain, and Basement membrane-specific heparan sulfate proteoglycan core protein) or potentially involved in remodelling (Gelsolin, Carboxypeptidase B, Metallocarboxypeptidase ECM14). Gelsolin is an actin binding protein which is a key regulator of actin filament assembly (Sun, Yamamoto et al. 1999), helps actin nucleation before polymerization, and cleaves actin filament, resulting in actin mobilization (Laine, Phaneuf et al. 1998). Its decrease is a signature of cytoskeleton disturbance, reflecting loss of actin polymerisation. Ezrin is also an actin binding protein involved in cell proliferation, survival and migration whose regulation depends on its phosphorylation states (Bosk, Braunger et al. 2011). The treatment with high dose of pyriproxyfen showed, through the 21 proteins modulated, the same proteomic signatures (metabolism decrease and cytoskeleton disturbance) with a marked effect on mitochondria shown by the significant decrease (8.5 fold, q-value<0.05) of the Contig_Gammarus_90_111327_fr6 annotated as

MICOS complex subunit MIC60. This large protein complex of the mitochondrial inner membrane is crucial in maintenance of the crista junction, the inner membrane architecture and formation of contact sites to the outer membrane which is a prerequisite for an efficient respiration and ATP generation (Eydt et al., 2017). Mitochondrial dysfunction associated with lower abundance of the Contig_Gammarus_90_204722 annotated as a mitochondrial dihydrolipoyl dehydrogenase may suggest spermatogenesis impairment, as this enzyme is involved in hyperactivation of spermatozoa during capacitation and in spermatazoal acrosome reaction (Mitra et al. 2005). We also noted amongst these proteins with high dynamics a set of uncharacterized proteins (LOC108676008, LOC108679669, LOC108664518, LOC108679669, hypothetical protein MGE_01014), mostly observed at high pyriproxyfen concentration.

CONCLUSION

With the case study documenting the testes proteome of *G. fossarum* organisms exposed to pyriproxyfen, we have compared the performance of two different generations of tandem mass spectrometers based on the Orbitrap analyser technology: the LTQ Orbitrap XL and the Q Exactive HF. Due to its faster scan rate, the quadrupole filtering properties and the higher resolution on MS/MS secondary ions, the Q Exactive HF instrument is able to record almost 5 fold more MS/MS spectra on the same *G. fossarum* proteome samples, but more importantly increases the intrinsic quality of these spectra. It allows identification of a larger number of peptides and proteins (4 fold). The coverage of proteins was also enhanced with a slight increase of the average number of peptides recorded for each protein, thus improving the sequence coverage even for less abundant proteins. It also increases the quality in terms of quantitation because higher spectral counts are evaluated for each protein. Importantly, the list of potential biomarkers detected with the LTQ Orbitrap XL that was previously reported (Trapp, Armengaud et al. 2015) was further amended with the Q Exactive HF data. Here, we report as potential protein signature of pyriproxyfen treatment some metabolic proteins, confirming the trends of a disturbance of the cell homeostasis upon pyriproxyfen stress. While we recently found that pyriproxyfen induced a decrease in spermatozoa count, this toxic effect is reflected by a limited number of testes proteins found modulated upon exposure. This may suggest that a sub-population of testes cells, such as presursors of mature spermatozoa, could be the main target of pyriproxyfen exposure. Within a proteogenomics framework for non-model animals used in ecotoxicology, we thus recommend the use of tandem mass spectrometers with performances such as the Q Exactive HF instrument, which provides a more comprehensive view of the proteome dynamic upon stress, and so allows analysing with a great depth of field the effects of a given xenobiotic. Our experimental data speak in favour of the use of the Q Exactive HF instruments to better cover the proteome of any biological sample, although previous generation of tandem mass spectrometers are also able to provide enough information for obtaining a first insight into stress response. While large number of biological replicates and technical replicates are usually recommended to better assessing biological mechanisms or validating biomarkers for routine ecotoxicological risk assessment (Trapp, Armengaud et al. 2014), the next-generation instruments based on the ultra-high-field Orbitrap analyser outperform previous generations in achieving more quickly the same comparative proteomics overview. However, despite these advances in mass-spec technology, the high dynamic range of the proteome of complex organisms still hinders the analysis of very low-abundant proteins. These proteins may be present permanently or transiently in the proteome, and generally comprise proteins with a high diagnostic potential, normally associated with transcription and translation mechanisms, membrane receptors, allergens, polypeptides, and post-translationally modified proteins. Their detection by mass spectrometry is still a key technical challenge in shotgun proteomics and biomarker discovery (da Costa 2017). Nevertheless, as demonstrated in this study, the continuous progresses in mass-spec technology will lead us towards bigger proteome coverages without the need to use time-consuming depletion/equalization strategies to reduce sample complexity.

The results presented here show that proteins involved in cytoskeleton organisation are perturbed by acetone at 0.005% concentration. Several of these proteins are subjected to post-translational modifications. Examining the phosphorylation pattern of these cytoskeletal proteins by means of targeted proteomics approaches could be pertinent for increasing the understanding of pyriproxyfen mode of action and better disentangle the differences between the solvent and the pyriproxyfen effects. On the other hand, among the proteins whose expression was dramatically increased by the pyriproxyfen, the Contig_Gammarus_90_150378 annotated as a glycogen phosphorylase could be a promising biomarker of pyriproxyfen-induced stress.

Moreover, the increase of the protein expression was directly linked to the dose treatment. The Contig_Gammarus_90_204722 whose protein homolog was identified as the mitochondrial Dihydrolipoyl dehydrogenase, would be interesting to follow due to its expression modulation by the pyriproxyfen and its implication in the spermatogenesis.

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), the Agence Nationale de la Recherche program "ProteoGam" (ANR-14-CE21-0006-02) for financial support. The authors also thank Laurent Guillet-Revol, Khedidja Abbaci, Hervé Quéau and Renaud Tutundjian for technical assistance during laboratory exposure.

REFERENCES

Armengaud, J., J. Trapp, et al. (2014). "Non-model organisms, a species endangered by proteogenomics." J Proteomics 105: 5-18.

Bosk, S., J. A. Braunger, et al. (2011). "Activation of F-actin binding capacity of ezrin: synergism of PIP(2) interaction and phosphorylation." Biophys J 100(7): 1708-1717.

Buchfink, B., C. Xie, et al. (2015). "Fast and sensitive protein alignment using DIAMOND." Nat Methods 12(1): 59-60.

Carvalho, P. C., D. B. Lima, et al. (2016). "Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0." Nat Protoc 11(1): 102-117.

Carvalho, P. C., J. R. Yates, 3rd, et al. (2012). "Improving the TFold test for differential shotgun proteomics." Bioinformatics 28(12): 1652-1654.

Colbourne, J. K., M. E. Pfrender, et al. (2011). "The ecoresponsive genome of Daphnia pulex." Science 331(6017): 555-561.

Drummond, D. A., J. D. Bloom, et al. (2005). "Why highly expressed proteins evolve slowly." Proc Natl Acad Sci U S A 102(40): 14338-14343.

Garcia-Reyero, N. and E. J. Perkins (2011). "Systems biology: Leading the revolution in ecotoxicology." Environmental Toxicology and Chemistry 30(2): 265-273.

Gillette, M. A. and S. A. Carr (2013). "Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry." Nature methods 10(1): 28-34.

Ginjupalli, G. K. and W. S. Baldwin (2013). "The time- and age-dependent effects of the juvenile hormone analog pesticide, pyriproxyfen on Daphnia magna reproduction." Chemosphere 92(9): 1260-1266.

Hartmann, E. M., F. Allain, et al. (2014). "Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria." Methods Mol Biol 1197: 275-285.

i, K. C. (2013). "The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment." J Hered 104(5): 595-600.

Laine, R. O., K. L. Phaneuf, et al. (1998). "Gelsolin, a protein that caps the barbed ends and severs actin filaments, enhances the actin-based motility of Listeria monocytogenes in host cells." Infect Immun 66(8): 3775-3782.

LeBlanc, G. A. (2007). "Crustacean endocrine toxicology: a review." Ecotoxicology 16(1): 61-81.

Loy, C. J., K. S. Sim, et al. (2003). "Filamin-A fragment localizes to the nucleus to regulate androgen receptor and coactivator functions." Proc Natl Acad Sci U S A 100(8): 4562-4567.

Michalski, A., E. Damoc, et al. (2011). "Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer." Mol Cell Proteomics 10(9): M111.011015.

Monsinjon, T. and T. Knigge (2007). "Proteomic applications in ecotoxicology." Proteomics 7(16): 2997-3009.

Nikinmaa, M. and K. T. Rytkonen (2011). "Functional genomics in aquatic toxicology-do not forget the function." Aquat Toxicol 105(3-4 Suppl): 16-24.

Olmstead, A. W. and G. A. LeBlanc (2003). "Insecticidal juvenile hormone analogs stimulate the production of male offspring in the crustacean Daphnia magna." Environ Health Perspect 111(7): 919-924.

Percy, A. J., R. Simon, et al. (2014). "Enhanced sensitivity and multiplexing with 2D LC/MRM-MS and labeled standards for deeper and more comprehensive protein quantitation." J Proteomics 106: 113-124.

Pereira-Medrano, A. G., A. Sterling, et al. (2007). "A Systematic Evaluation of Chip-Based Nanoelectrospray Parameters for Rapid Identification of Proteins from a Complex Mixture." Journal of the American Society for Mass Spectrometry 18(9): 1714-1725.

Rehm, E., R. Hannibal, et al. (2008). "The Crustacean Parhyale hawaiensis, A New Model for Arthropod Development." Emerging Model Organisms: A Laboratory Manual 1: 373 - 404.

Schnier, J. B., K. Nishi, et al. (2005). "Glycogen synthesis correlates with androgen-dependent growth arrest in prostate cancer." BMC Urol 5: 6.

Sullivan, J. J. and K. S. Goh (2008). "Environmental fate and properties of pyriproxyfen." J Pest Sci 33(4): 339-350.

Sun, H. Q., M. Yamamoto, et al. (1999). "Gelsolin, a multifunctional actin regulatory protein." J Biol Chem 274(47): 33179-33182.

Swanson, W. J. and V. D. Vacquier (2002). "The rapid evolution of reproductive proteins." Nat Rev Genet 3(2): 137-144.

Trapp, J., C. Almunia, et al. (2015). "Proteogenomic insights into the core-proteome of female reproductive tissues from crustacean amphipods." J Proteomics.

Trapp, J., J. Armengaud, et al. (2015). "Proteomic investigation of male Gammarus fossarum, a freshwater crustacean, in response to endocrine disruptors." J Proteome Res 14(1): 292-303.

Trapp, J., J. Armengaud, et al. (2014). "Next-generation proteomics: toward customized biomarkers for

environmental biomonitoring." Environ Sci Technol 48(23): 13560-13572.

Trapp, J., O. Geffard, et al. (2014). "Proteogenomics of Gammarus fossarum to document the reproductive system of amphipods." Mol Cell Proteomics 13(12): 3612-3625.

Vizcaino, J. A., R. G. Cote, et al. (2013). "The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013." Nucleic Acids Res 41(Database issue): D1063-1069.

Note 1: Identification of candidate genes via literature search

In the second axis of the thesis, we aimed at identifying proteins/genes of interest as biomarkers of endocrine disruption in *G. fossarum*. Since the majority of molecules implicated in the neuroendocrine regulation of reproduction and molt were absent from the protein sequence database, we mined the literature and the transcriptome GFOSS in order to develop biomarkers *de novo*. We defined hormones, hormonal receptors, enzymes from hormonal metabolism, and other genes regulated by or that regulate hormones as targets for the research.

Firstly, we focused the research in the enzymes responsible for the metabolism and the receptors that mediate the actions of both the sesquiterpenoid juvenile hormone (methyl farnesoate in crustaceans) and the steroidal ecdysone. Among arthropods, the knowledge of these hormonal regulators is well known in insects, but in other arthropods much less is understood. Due to some studies in arthropod phylogeny that placed the Insecta within the Crustacean (Regier et al., 2005), the metabolism of these hormones is thought to be shared among several species of arthropods. One study from 2014 (Sin et al., 2014) explored the sequenced genome of the decapod crustacean *Neocaridina denticulata*, and proposed for the first time evidence of the sesquiterpenoid and ecdysteroid pathways in a crustacean. These pathways are illustrated in Figures 1 and 2.

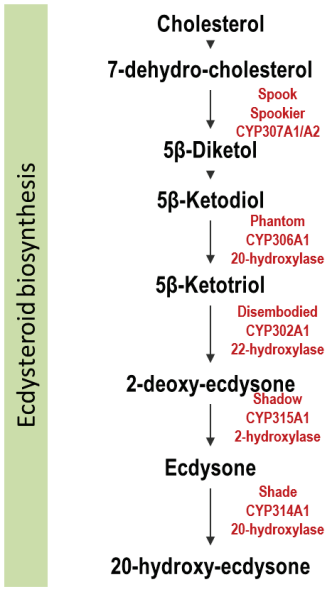


Figure 1 - Ecdysone hormone pathway overview (biosynthesis). (Sin et al., 2014)

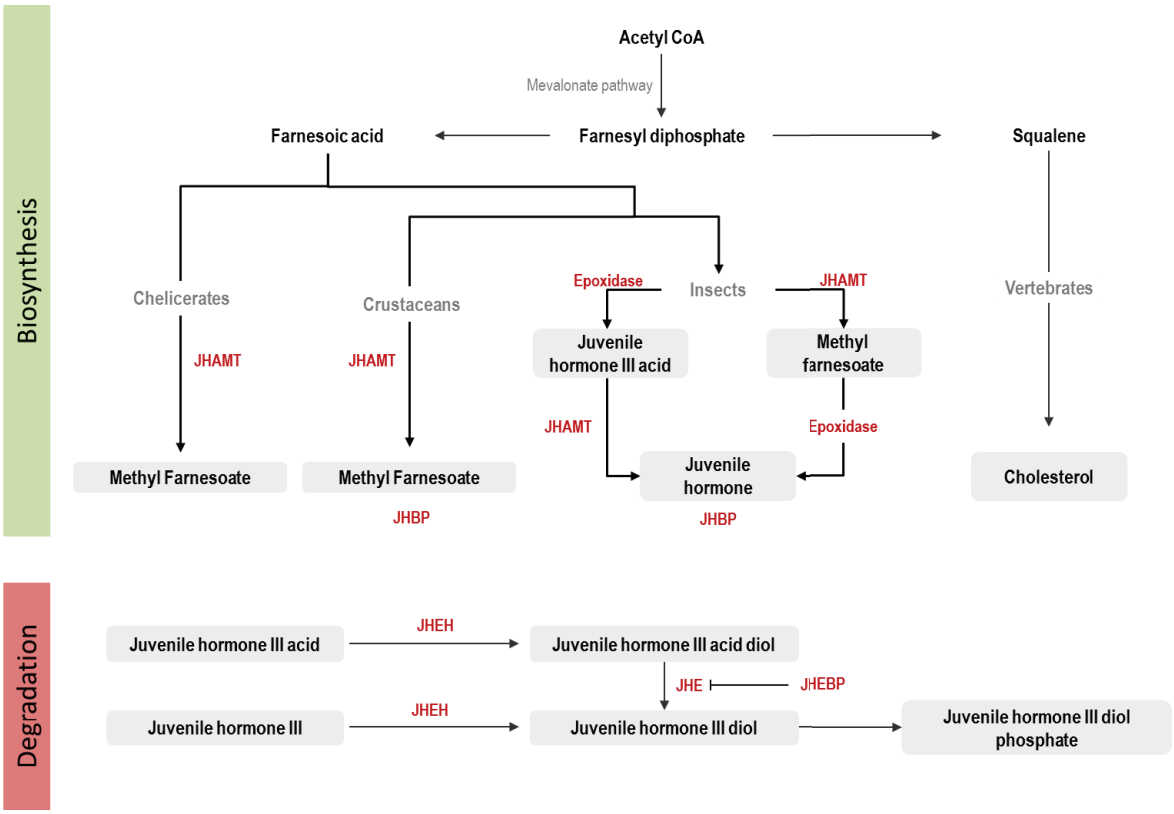


Figure 2 - Juvenile hormone pathway overview (biosynthesis and degradation). (Sin et al., 2014)

The majority of the genes known to be involved in these pathways in insects were also found in *N. denticulata* genome. The authors found the metabolic enzymes juvenile hormone acid methyltransferase (JHAMT), juvenile hormone binding protein (JHBP), juvenile hormone esterase (JHE), juvenile hormone esterase binding protein (JHEBP) and juvenile hormone epoxide hydrolase (JHEH). From the ecysteroid biosynthetic cascade, identified genes included *spook*, *phantom*, *disembodied*, *shadow* and *CYP18*. Potential hormonal regulators and signal transducers such as allatostatins (ASTs), Methoprene-tolerant (Met), Retinoid X receptor (RXR), Ecdysone receptor (EcR), calponin-like protein Chd64, FK509-binding protein 39KDa (FKBP39), Broad-complex (BR), and crustacean hyperglycemic hormone/molt-inhibiting hormone/gonad-inhibiting hormone (CHH/MIH/GIH) genes are all present in the shrimp *N. denticulata*. In the same year, the JH system was also found to be present in the myriapod *Strigamia maritima* (Chipman et al., 2014), which confirms the high conservation of the overall functioning of this system in the different arthropod subdivisions. The occurrence in chelicerates, insects, crustaceans, and myriapods suggests that the JH system has deep evolutionary roots, and so it is likely to find this system in the amphipods *Gammarus*.

Based on these datasets, we established a list of molecules implied in sesquiterpenoid and ecdysteroid related-processes. For each candidate, a BLASTP search was performed against our RNAseq-derived transcriptome database GFOSS, in order to find homolog *G. fossarum* specific sequences. Query sequences used for the blast analysis were from phylogenetically close species obtained from the NCBI protein database. Table 1 indicates the availability of these sequences either in the RNAseq-derived transcriptome database GFOSS and/or the protein database. From the 32 candidate genes in the list, 19 had homolog sequences in the GFOSS transcript database. Among these 19, only nine were identified in the proteomics experiments. Seen that all the enzymes involved in the synthesis and degradation of sesquiterpenoids are present in the *G. fossarum* transcriptome, one can hypothesize that the JH system is also present in this amphipod species. Moreover, one transcript annotated as being the "crustacean hyperglycemic hormone" CHH, known to be a potential regulator of MF activity, was also found. The neuropeptides Allatostatins and their respective receptors were not found. These molecules are responsible for the inhibition of the biosynthesis of JH in insects (Stay and Tobe, 2007).

No sequences were found for the putative JH receptor Methoprene-tolerant (Met). This receptor has a high affinity for JH, can act as an important transducer of JH signal (Konopova et al., 2011), and interacts with other regulatory elements such as Chd64 and FKBP39 (Li et al., 2007), whose sequences are also absent from our reference transcriptome. Homolog sequences were found for the ecdysone receptors RXR and EcR. The steroidogenic cytochrome P450 (CYP) enzymes responsible for ecdysone synthesis from cholesterol were disregarded due to the presence of too many CYP-annotated transcripts in the database. Because of the large spectrum of enzymes in this family, we also stated that it would be too difficult to use these enzymes as biomarkers of endocrine disruption. Genes listed as "other candidates" were found in the literature and classified as interesting genes due to either their implication in sex development or putative hormonal regulatory functions. All of these genes were found in the GFOSS transcriptome database.

Table 1– Candidate ED biomarker list constructed after literature mining. Functional information obtained from (Chipman et al., 2014; Sin et al., 2014), except when indicated otherwise. The two columns on the right indicate the presence or absence of sequences in the transcript and protein database. "?" represent a doubt whether the sequences are present or not due to too many contigs with CYP annotations.

Sesquiterpenoid synthesis and regulation				
Gene name	**Abbreviation**	**Function**	**GFOSS RNAseq**	**Proteomics experiment**
Juvenile hormone acid methyltransferase / Farnesoic acid methyltransferase	JHAMT	Synthesis of JH or MF from FA	✓	✓
Juvenile hormone binding protein	JHBP	Carrier of JH to transport it from the synthesis site to target tissues	✓	☒
Juvenile hormone esterase	JHE	JH degradation	✓	✓
Juvenile hormone esterase binding protein	JHEBP	Transport and degradation of JHE	✓	☒
Juvenile hormone epoxide hydrolase	JHEH	JH degradation	✓	✓
Allatostatin A, B, C	Ast-A	Inhibition of JH synthesis	☒	☒
Allatostatin A, B, C receptors	Ast-A-R	Receptors for allatostatins	☒	☒
Methoprene-tolerant	Met	Met and steroid receptor coactivator proteins form a heterodimer important for JH reception	☒	☒
Broad-complex	BR	Target of Met; Ecdysone responsive gene	✓	☒
Chd64	Chd64	JH response element binding protein	☒	☒
FKBP39	FKBPA39	JH response element binding protein	☒	☒
Hexamerin	Hex	Regulator of JH	☒	☒
Crustacean hyperglicemic hormone	CHH	Potential regulator of MF	✓	☒
Ecdysteroid synthesis and regulation				
Gene name	**Abbreviation**	**Function**	**GFOSS RNAseq**	**Proteomics experiment**
Retinoid X receptor (ultraspiracle)	RXR	Ecdysteroids regulate molting by activating a heterodimer formed by the EcR and RXR	✓	☒
Ecdysone receptor	EcR	Ecdysteroids regulate molting by activating a heterodimer formed by the EcR and RXR	✓	☒
CYP307A1 spook	spo	synthetize 20-hydroxyecdysone from cholesterol	?	☒
CYP306A1 phantom	phm	synthetize 20-hydroxyecdysone from cholesterol	?	☒
CYP302A1 disembodied	dib	synthetize 20-hydroxyecdysone from cholesterol	?	☒
CYP315A1 shadow	sad	synthetize 20-hydroxyecdysone from cholesterol	?	☒
CYP314A1 shade	shd	synthetize 20-hydroxyecdysone from cholesterol	?	☒
CYP18A1		Paralog of CYP306	?	☒
Other candidates				
Gene name	**Abbreviation**	**Function**	**GFOSS RNAseq**	**Proteomics experiment**
Calreticulin	CRP55, CaBP3, ERp60	Inhibits the binding of androgen receptor to its hormone responsive DNA element ; can also inhibit androgen receptor and retinoic acid receptor transcriptional activities in vivo (Dedhar et al., 1994)	✓	✓

Vasa	Vs	Protein essential for germ line development in *Melita plumulosa* (Hook et al., 2014b); specifically expressed in gonads and plays an important role in germ cell formation in *Scylla paramamosain* (Gao et al., 2014)	✓	✓
Dmrt (double sex)	Dmrt	Regulation of sex determination/differentiation pathways in most vertebrates and invertebrates (Gao et al., 2014)	✓	☒
FEM - 1	FEM-1	Required for all aspects of testis development (Gao et al., 2014); signal transducing regulator in the *C. elegans* sex determination pathway (Gaudet et al., 1996)	✓	☒
Tra-2	Tra-2	Involved in *M. nipponense* sex determination (Jin et al., 2013); sex determination in insects (Shen et al., 2014)	✓	☒
Fru	Fru	Sex determination in insects (Shen et al., 2014)	✓	☒
Chromobox protein	CBX	associated with sex determination (Jin et al., 2013)	✓	✓
Cathepsin D et L	CTS	D is necessary for the formation of the yolk ; L regulates the development of the ovary in many species (Jin et al., 2013)	✓	✓
Hemocyanin	Hc	Hemoglobin is known to be regulated by juvenile hormone in daphnia, which may explain the up-regulation of hemocyanin observed in amphipods exposed to bifenthrin (Hook et al., 2014b)	✓	✓

Note 2: Reprotoxicity assessment of model molecules in female gammarids

The objective was to evaluate the impact of several insecticides in the reproductive processes of female gammarids, particularly on the molt cycle, and oocyte/embryo development. These parameters are proposed as biomarkers of reproductive disorders and potential endocrine disruption in this species (Geffard et al., 2010). The molt cycle, oocyte growth, and embryonic development are perfectly synchronized in crustacean amphipods. Geffard *et al.* 2010 established that a female *G. fossarum* in the C2 intermolt stage possesses embryos in stage 3 and oocytes with an average surface of 110,000 µm2. Therefore, the impact of a pollutant on one of these processes and their consequent desynchronization regarding the others will highlight a specific endocrine disruption mode of action.

The exposure protocol and the physiological measurements performed in this experiment are described in chapter II, section 1.2.1 and 1.3.1 (Experiment 3). The molecules used were selected based on their toxicity on other crustacean species, especially in daphnids (Olmstead and LeBlanc, 2003), and are listed in Table 1. The molecules identified as having an endocrine disruption impact will be then proposed for the subsequent validation of the candidate ED biomarker genes.

Table 1 – List of molecules tested for toxicity reproductive assessment in females.

Molecule	CAS	Empirical formula	Description
Methoxyfenozide	161050-58-4	$C_{22}H_{28}N_2O_3$	Analog of 20-E
Tebufenozide	112410-23-8	$C_{22}H_{28}N_2O_2$	Analog of 20-E
Fenoxycarb	72490-01-8	$C_{17}H_{19}NO_4$	Analog of JH
Pyriproxifen	95737-68-1	$C_{20}H_{19}NO_3$	Analog of JH
Juvenile hormone III	24198-95-6	$C_{16}H_{26}O_3$	Natural hormone
E,E Methyl farnesoate	10485-70-8	$C_{16}H_{26}O_2$	Natural hormone

A summary of the results obtained in this experiment are given in Table 2. Briefly, all molecules exerted a toxic effect at least at one of the tested concentrations, except methoxyfenozide. Oocyte growth and development were the most sensitive parameters, since five test molecules affected them. The juvenile hormone III was the only molecule

that affected the molt cycle, while fenoxycarbe and juvenile hormone caused embryotoxicity, *i.e.* a significate lower number of embryos in the brood pouch.

Table 2 – General overview of the significant modulations obtained between control vs contaminated conditions, for the four physiological parameters analyzed.

		0.05 µg/L	0.5 µg/L	5 µg/L	50 µg/L
Fenoxycarbe	Retard de mue	NS	NS	NS	NS
	Nb embryons	**	NS	NS	NS
	Nb ovocytes	NS	NS	NS	*
	Surface ovocytaire	NS	NS	NS	NS
Tebufenozide	Retard de mue	NS	NS	NS	NS
	Nb embryons	NS	NS	NS	NS
	Nb ovocytes	NS	*	NS	NS
	Surface ovocytaire	NS	*	NS	NS
Hormone juvenile III	Retard de mue	NS	*	NS	NS
	Nb embryons	NS	NS	*	NS
	Nb ovocytes	NS	**	NS	*
	Surface ovocytaire	NS	**	NS	NS
Methyl-farnesoate	Retard de mue	NS	NS	NS	NS
	Nb embryons	NS	NS	NS	NS
	Nb ovocytes	NS	*	NS	NS
	Surface ovocytaire	NS	*	NS	NS
Pyriproxyfène	Retard de mue	NS	NS	NS	NS
	Nb embryons	NS	NS	NS	NS
	Nb ovocytes	NS	NS	*	*
	Surface ovocytaire	NS	NS	***	**
Methoxyfenozide	Retard de mue	NS	NS	NS	NS
	Nb embryons	NS	NS	NS	NS
	Nb ovocytes	NS	NS	NS	NS
	Surface ovocytaire	NS	NS	NS	NS

NS corresponds to not-significant; * p-value < 0,05 ; ** p-< 0,01 ; *** p-value < 0,001

In order to identify molecules that potentially exerted an ED-related effect, we analyzed the specificity of action of the molecules. Endocrine disruptors are also characterized by their non-monotonic dose response curves. Three molecules met these criteria and were selected as candidates for the subsequent biomarker validation experiments: tebufenozide, methyl farnesoate, and pyriproxyfen. These molecules presented a very specific toxicity, affecting only oocyte growth and number, at specific concentrations (0.5 µg/L for TEB and MF, 5 and 50 µg/L for PYR).

Publication nº5

Identification, expression, and endocrine-disruption of three ecdysone-responsive genes in the sentinel species *Gammarus fossarum*

Duarte Gouveia[1,2], François Bonneton[3], Jean Armengaud[2], Hérvé Quéau[1], Davide Degli-Esposti[1], Olivier Geffard[1], Christine Almunia[2], Arnaud Chaumot[1]

[1] Irstea, Unité de Recherche MALY, Laboratoire d'écotoxicologie, F-69626 Villeurbanne, France
[2] CEA-Marcoule, DSV-Laboratory « Innovative technologies for Detection and Diagnostics », Bagnols-sur-Ceze, F-30207, France
[3] IGFL, CNRS-UMR5242, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

Submitted to "**Scientific Reports**".

Abstract: Taking advantage of a large transcriptomic dataset recently obtained in the sentinel crustacean amphipod *Gammarus fossarum*, we developed an approach based on sequence similarity and phylogenetic reconstruction to identify key players involved in the endocrine regulation of *G. fossarum*. Our work identified three genes of interest: the nuclear receptors RXR and E75, and the regulator broad-complex (BR). Their involvement in the regulation of molting and reproduction, along with their sensitivity to chemical contamination were experimentally assessed by studying gene expression during the female reproductive cycle, and after laboratory exposure to model endocrine disrupting compounds chemical compounds (pyriproxyfen, tebufenozide and piperonyl butoxide). RXR and E75 transcripts were modulated during the female reproductive processes. RXR expression suggested a role of this gene in ecdysis and post-molting processes. E75 presented two peaks of expression: one first peak during the inter-molt stage corresponding to the beginning of secondary vitellogenesis, and a second peak at the final stage before molting. BR expression showed no variation during molting/reproductive cycle. After exposure to the three insecticides, a strong inhibition of the inter-molt E75 peak was observed with tebufenozide, and an induction of RXR after exposure to pyriproxyfen and piperonyl butoxide. These results confirm the implication of RXR and E75 in hormonal regulation of the female reproductive cycle in *G. fossarum* and their sensitivity towards insecticides opens the possibility of using them as specific endocrine disruption biomarkers.

Keywords: *Endocrine disruption, gene expression, ecdysone, nuclear receptors, Gammarus fossarum, biomarker*

Introduction

Endocrine disrupting compounds (EDCs) are exogenous substances that interfere with hormone-regulated physiological processes and provoke adverse health effects in exposed organisms and/or their progeny [1]. EDCs typically interfere with hormone signaling, acting as agonist or antagonist, or with hormone synthesis, through anti-hormonal effects for example [2,3]. These compounds constitute a worldwide concern for potential human health implications [4], especially due to the multiple developmental and reproductive disorders observed in wildlife [5-9].

In aquatic environments, occurrence and biological effects of EDCs have been reported, such as masculinization and feminization events. This has led to the need of specific monitoring and the development of several robust, specific and reliable biomarkers of EDC exposure in vertebrates, such as the vitellogenin (Vtg) induction in male fishes [10-12]. However, due to the molecular divergence acquired through animal evolution, vertebrate-derived biomarkers cannot be used in invertebrates, which represent the majority of animals and key species for ecosystem functioning. The need to develop relevant tools for identification and assessment of EDCs toxic effects

and mode of action in invertebrates has been already highlighted by the scientific community. The discovery of imposex in marine gastropods due to Tributyltin (TBT) [13], and intersex occurrence in natural populations of gammarids [14,15] are two examples of the importance of developing more studies of EDCs effects in reproductive function of invertebrate species. Nevertheless, due to a currently limited knowledge in invertebrate endocrinology, today there is still a scarce understanding of EDCs modes of action in the majority of invertebrate species.

Among crustacean species, the amphipod *Gammarus fossarum* is commonly used as a sentinel species in freshwater risk assessment [16]. Many toxicity markers are available in this species, mainly based on life-history traits such as reproductive features [17], locomotor behavior/feeding rate [18], neurotoxicity [19] and genotoxicity [20]. The recent advances in proteomics techniques led to the development of a specific methodology based on liquid chromatography coupled to tandem mass spectrometry for quantifying a Vtg protein in *G. fossarum* [21]. However, the proposal of this protein as an endocrine disruption (ED) biomarker for male feminization was discarded after laboratory and field experiments that yielded high variability and low Vtg inductions in organisms exposed to contaminants [22]. Similar conclusions were reached when analyzing expression levels of two Vtg genes in *Echinogammarus marinus*, highlighting the fact that invertebrate biomarkers must be developed and validated *de novo,* and not just derived from their eukaryotic counterparts [23]. More recently, a proteogenomic study based on the alliance of transcriptomic and proteomic approaches [24] lead to the construction of a database (GFOSS database) containing 1,873 experimentally validated *Gammarus fossarum* specific-proteins [25]. However, few proteins involved in hormonal regulation could be recognized from this database, and potential key candidates for ED biomarkers were not identified. Indeed, such proteins are expected to be present at very low levels and therefore are hence probably lost in the "background noise" of the proteome.

Taking this context into account, we hypothesized that a gene candidate approach considering a thorough bioinformatics mining of the transcriptomic database obtained in *G. fossarum* could be used to identify candidate biomarkers of endocrine disruptor effects in this species. As discussed in previous studies [17,26], the reproduction cycle of female amphipods such as gammarids (vitellogenesis/oogenesis, embryogenesis) is closely correlated with molting. Disruption of these processes by contamination results in reproductive impairment. In crustaceans, molting is regulated by a multi-hormonal system including specific neuropeptide hormones, juvenoid hormones such as methyl farnesoate (MF), and biological active forms of ecdysteroids such as 20-hydroxyecdysone (20E) or ponasterone A (reviewed in [26] and [3]). Among arthropods, while the ecdysone pathway is well characterized in insects, much less is understood in crustacean species. In insects, 20E binds to and activates the ecdysone receptor/retinoid-X-receptor (EcR/RXR, mainly known as USP in insects) nuclear receptor complex, initiating a cascade of gene-regulatory events that will mediate molting and reproduction (**Figure 1**). Early response genes from the insect ecdysone pathway include the nuclear receptors ecdysone induced protein 75B (E75), ecdysone induced protein 78C (E78) and hormone receptor 3 (HR3), and the transcription factors broad-complex (BR) and ecdysone induced protein 74EF (E74) [27]. In crustaceans, early response genes are thought to be similar to those of insects [3,28,29].

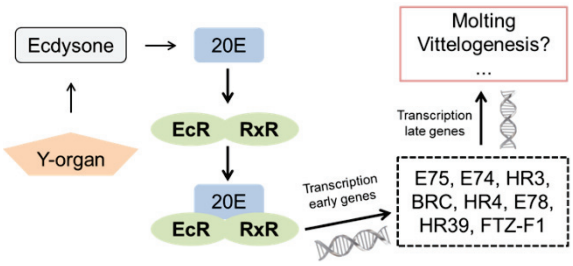


Figure 1: Putative ecdysone regulatory cascade in crustaceans. Ecdysone is secreted by the Y-organ, leading to its active form 20-hydroxyecdysone (20E). The heterodimer EcR/RXR receives 20E and initiates the transcription of the "early" response genes, which in turn will regulate the transcription of "late" genes responsible for initiating molting and vitellogenesis. After (Bonneton and others 2008).

The present study aimed at identifying gene candidate biomarkers of ED effects in the crustacean *G. fossarum*. Assuming a possible homology between crustacean and insect endocrine regulator genes, we mined the transcriptomic GFOSS database and performed sequence similarity searches and phylogenetic analysis in order to identify some key players potentially involved in endocrine regulation in *G. fossarum*. Since the ecdysone pathway is crucial for arthropod molting and reproduction, our work focused on ecdysone-responsive genes whose homologous sequences were identified in GFOSS. Through molecular phylogenetic analysis, we validated the annotations of the nuclear receptors E75 and RXR, and of the transcription factor BR in *G. fossarum* sequences. In order to obtain reliable nucleotide sequences of gene candidates and

primers for subsequent gene expression studies, these genes were cloned and sequenced. Their involvement in the endocrine control of molting and reproductive process and sensitivity to EDCs were evaluated by studying their profile during the female reproductive cycle and following laboratory exposure to three chemicals acting on the hormonal regulation of target arthropods: pyriproxyfen, tebufenozide and piperonyl butoxide.

Materials and Methods

Candidate selection and phylogenetic validation

Selection of Gammarus fossarum specific sequences

Given the importance of ecdysone in arthropod molting and reproduction, we focused on three ecdysone-responsive genes: RXR, EcR and BR. For each candidate, a BLASTP homology search was performed against our RNAseq-derived transcriptome database GFOSS [25], using query sequences from phylogenetically close species obtained from the NCBI protein database. The EcR and BR query sequences used for blast were from the recently genome-sequenced amphipod *Hyallela azteca*, while the RXR sequence belonged to the crab *Eriocheir sinensis*. For this gene, decapods were the closest order to have sequenced RXR sequences in the NCBInr protein database. The BLAST search was conducted against both assembled contigs and unassembled reads databases. The top 3 blast scores with E-values inferior to E^{-10} and alignment length superior to 25% were chosen for the following analysis (Table 1).

Sequence alignment and phylogenetic analysis

For each candidate, a homologous sequence dataset was constructed using ortholog and paralog sequences from the same multigenic family as the candidate gene. The deduced *G. fossarum* sequences were aligned with other known sequences from a diversity of crustacean and insect species, obtained from GenBank (National Center for Biotechnology Information – NCBI) (listed in **Supplementary Table 1**).

Multiple sequence alignments and phylogenetic trees were performed in the SeaView software version 4.6.1 [58]. Alignments were made using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) [59] program using default parameters. Phylogenetic tree model was determined using the IQ-TREE web server model selection [60]. Trees were built also in the SeaView, using the PhyML program based on the maximum-likelihood principle [61], with the LG model with 4 substitution rate categories to estimate the gamma parameter shape, and 100 bootstrap replicates for branch support. Default settings were used for all other parameters.

Gene expression studies

Sampling and maintenance of organisms

Gammarids were collected from the River Le Pollon in France (45°57'25.8"N 5°15'43.6"E) and acclimatized to laboratory conditions as previously described [33]. Organisms were collected by kick sampling using a net, and quickly transported to the laboratory in plastic buckets containing freshwater from the station. In the laboratory, organisms were kept in 30 L tanks continuously supplied with drilled groundwater (500µS/cm) and under constant aeration for at least 10 days. A 16/8h light/dark photoperiod was maintained and the temperature was kept at 12±1°C. Organisms were fed *ad libitum* with alder leaves (*Alnus glutinosa*), previously conditioned for 6±1 days in water. Sexually mature male and female organisms (selection based on the developmental stage of the embryos; well-developed juveniles in the female marsupium) were sampled for RNA extraction and gene sequencing protocols.

Selection of females at different phases of reproductive cycle

For gene expression studies, five females from each main and intermediate stages of the molt cycle (A, B, C1, C1/C2, C2, C2/D1, D1, D1/D2, D2, and final D2) were sampled. The selection of females at specific stages was determined through observation of characteristic integumental changes on the dactylopodite and protopodite during the molt cycle (described in [17]). Briefly, in *Gammarus fossarum* females, oogenesis and embryogenesis occurs simultaneously with the molting cycle. Six molt stages were previously defined [17], A and B for postmolt, C1 and C2 for intermolt, and D1 and D2 for premolt. Organisms were incubated overnight at 4ºC in 300 µL of RNAlater, frozen in liquid nitrogen and immediately stored at -80ºC until total RNA extraction.

Laboratory exposures to endocrine disrupting chemicals

Coupled male and female organisms were sampled and isolated in a separated aquarium. Immediately after spawning, females were isolated and placed in 500 mL glass beakers (10 females per beaker, 3 beakers per condition). There were three conditions corresponding to three different exposure times: 4, 9 and 14 days. Solvent controls

(uncontaminated drilled groundwater spiked with 0.005% acetone) were used to take into account any potential solvent effect in the interpretation of the subsequent gene expression analysis. Organisms were thus exposed to three EDCs: pyriproxyfen (PYR 5 µg/L), tebufenozide (TEB 0.5 µg/L) and piperonyl butoxide (PBO 150 µg/L). Stock solutions were prepared in acetone at concentrations of 100, 10 and 3000 mg/L for PYR, TEB and PBO, respectively, adding small aliquots of these stock solutions to the dilution water. Renewal of the media was performed manually every two days, and temperature was maintained at 16°C throughout the experiment. At the end of the exposure (days 4, 9, and 14), five females from each condition were individually sampled, rapidly weighed, and incubated overnight at 4°C with 300 µL of RNAlater. After incubation, samples were frozen in liquid nitrogen and stored at -80°C until the RNA extraction protocol.

Total RNA extraction and cDNA synthesis

Whole-body organisms were disrupted with a TissueRuptor (Qiagen). For total RNA extraction, the RNeasy® Fibrous Tissue Mini Kit was used according to the manufacturer's instructions (Qiagen). DNase I was used during the protocol to remove possible genomic contaminations. Total RNA was subsequently quantified using a NanoDrop 2000 spectrophotometer (Thermo Scientific), followed by a qualitative analysis using an Agilent Bioanalyzer 2100 RNA Nanochip (Agilent technologies, Santa Clara, CA). A quantity of 1 µg of RNA from each sample was converted into cDNA using the SuperScript® III First-Strand Synthesis System (Thermo Fisher Scientific) following the manufacturer's protocol, and conserved at -20 °C until further analysis.

Gradient PCR amplification, cloning and sequencing

PCR was performed with a 50 µL mix of Millipore water containing 1 µL of cDNA, 2 µL of 10mM dNTP, 2 µL of each primer at 10 µM, 5 µL of RT buffer, 3 µL of 25mM $MgCl_2$, and 0.5 µL of native Taq DNA polymerase (Thermo Fisher Scientific) (200U/µL). Primers (listed in **Table 2**) were manually designed, based on sequence alignments of *G. fossarum* deducted sequences with other arthropod sequences of the same gene.

The following PCR program was used (Biometra TGradient): initial denaturation step at 94°C for 3 min, followed by 39 amplification cycles (denaturation 30s at 94°C; annealing 30s at 56, 58.3, 59.7, and 62°C; elongation 2min at 72°C). A final extension step was made with 10 min incubation at 72°C. The PCR products were run on a 1% agarose gel (**Supplementary Figure 1**).

PCR amplicons were purified using the QIAquick®PCR Purification Kit (Qiagen). Purified fragments were inserted into pGEM®-T Vectors (pGEM®-T Vector System I, Promega), and transformed into DH5α competent cells. Clones containing target fragments were isolated and submitted for direct colony PCR verification using the following program: initial denaturation step at 94°C for 10 min, followed by 40 amplification cycles (denaturation 30s at 94°C; annealing 30 s at 55°C; elongation 1 min at 72°C). A final extension step was made with 5 min incubation at 72°C. The PCR products were run on a 1% agarose gel. To obtain high-purity plasmids, positive clones were purified with the QIAprep®Spin Miniprep Kit (Qiagen) following the manufacturer's protocol. Samples were sent to Beckman Coulter Genomics (GENEWIZ®, UK) for sequencing.

Quantitative real-time polymerase chain reaction (qPCR)

For target genes, the same primers were used as for gene cloning (**Table 2**). Primers for reference genes, also listed in **Table 2**, were selected from previous studies that performed qPCR studies in *Gammarus* species [62,63]. Three reference genes were tested: Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), 18S RNA, and elongation factor 1 α (EF). In order to select a suitable reference gene, the stability of these three genes was assessed in cDNA samples from the physiological and exposure experiments. EF was considered as the most stable, presenting the lowest standard deviation for all Ct values obtained in the different samples, as shown in **Table 2**. The EF gene was therefore chosen to normalize expression data in both experiments.

For the quantitative PCR reactions, the iTaq™ Universal SYBR® Green Supermix (Biorad) kit was used. For the reactions, 25 ng (and 8.5 ng in the contamination experiment) of template cDNA and 300 mM of each primer were used, and the procedure was according the manufacturer's instructions. PCR was performed in a CFX96 Touch™ Real-Time PCR Detection System (Biorad), using the following program: 30 sec at 95°C, 40 cycles of 95°C for 5 s, 60°C for 30 s then a temperature increment was programmed for the melting curve (65–95 °C with 0.5°C increments at 5 s/step). Samples for each condition were a pool of five biological replicates (RNAs extracted from the individual females were pooled together for the reverse transcription and subsequent qPCR). For the physiological experiment, we also analyzed the individual biological replicates in specific stages of the female reproductive cycle. The stages were

selected based on the results from the pools, with the purpose of validating the observed peaks of expression. Therefore, for E75 we analyzed individuals from the C2D1, D1D2, and D2F conditions, while for RXR we analyzed individuals from the A, C2D1, and D2F conditions. All samples were analyzed in triplicate, and controls without DNA were run in every plate. Gene expression levels were analyzed using the relative quantification method (ΔCt) [64].

Statistical analysis

Statistical analyses for testing differences between conditions were performed with the GraphPad Prism Version 7.02 software, using unpaired t-tests corrected for multiple comparisons, according to the the Holm-Sidak method. Fold change was calculated using the $2^{-\Delta\Delta Ct}$ method normalized to the control condition. Significant differences were accepted when meeting both $p < 0.05$ and fold change >1.5.

Table 1: Sequence similarity search results after blastp of RXR, EcR and BR query sequences against the GFOSS database.

Query sequence (Accesion n° - name - [species])	Top 3 blast results	Score	Identitites	Percentage	Expect	GFOSS annotation
XP_018017941.1 PREDICTED: ecdysone receptor-like isoform X1 [Hyalella azteca]	GFOSS_83504_fr5	94	39/86	45	1,0E-18	ultraspiracle
	GFOSS_2900_fr4	70	95/389	24	1,0E-11	Nuclear hormone receptor E75
	seq406923_fr4	70	30/47	63	2,0E-11	-
XP_018008501.1 PREDICTED: broad-complex core protein isoforms 1/2/3/4/5-like isoform X4 [Hyalella azteca]	GFOSS_1452_fr3	281	162/230	70	7,00E-75	broad-complex, isoform Z2
	seq986730_fr4	276	132/139	94	2,00E-73	-
	seq1061041_fr1	274	132/139	94	7,00E-73	-
AHF65151.1 retinoid X receptor [Eriocheir sinensis]	seq307974_fr2	228	113/198	57	3,0E-59	-
	GFOSS_83504_fr5	159	71/90	78	2,0E-38	ultraspiracle
	GFOSS_2900_fr4	114	81/307	26	1,0E-24	Nuclear hormone receptor E75

Results and discussion

Phylogenetic sequence validation

In the present study, based on the GFOSS RNAseq-derived transcriptome database, we aimed to identify *G. fossarum* specific sequences homologue to key genes known to be involved in endocrine regulation of arthropods. Seeing the importance of ecdysone pathway in arthropod molting and reproduction, we focused on three ecdysone-responsive genes: RXR, EcR and BR. After blast sequence similarity searches against GFOSS, the top three results were kept for phylogenetic validation of their annotations. As nuclear receptors are paralog genes with a highly evolutionary conserved domain called DNA-binding domain (DBD) and other conserved amino-acyl sequences in functional areas in the ligand-binding domain (LBD), the BLAST searches for RXR and EcR sequences yielded common results (**Table 1**). Sequences GFOSS_83504 (annotated as ultraspiracle - same as RXR - in GFOSS) and GFOSS_2900 (annotated as nuclear hormone receptor E75 in GFOSS database) were found among the top three blast results of both searches using the query sequences of RXR and EcR from the decapod *Eriocheir sinensis* and the amphipod *Hyalella azteca*, respectively. Concerning BR, blast results yielded one sequence GFOSS_1452 (annotated as broad-complex isoform Z2) and two unassembled reads seq986730_fr4 and seq1061041_fr1.

The sequence annotations available in GFOSS (reported in **Table 1**) were previously generated by an automatic bioinformatic pipeline based on reciprocal blast procedure against public sequence databases [25]. Such approach can be flawed due to affectation of functional annotations from a paralog sequence to the gene identified in the species of interest. Then, in order to select the *G. fossarum* ortholog genes of RXR, EcR, and BR, phylogenetic analyses were conducted to validate the annotation of the candidate sequences. Each *G. fossarum* sequence from **Table 1** was thus aligned with ortholog sequences from different crustacean/insect species, and paralog sequences from different members of the corresponding multigenic family, *i.e.* nuclear receptor family for RXR and EcR, and BTB-ZF family for BR (sequence accession numbers are listed in **Supplementary Table 1**). The phylogenetic trees for candidate sequences are represented in **Figure 2** (EcR), **Figure 3** (BR) and **Figure 4** (RXR).

As presented in **Figure 2**, sequence GFOSS_2900 did not cluster with EcR sequences, but rather with sequences from the nuclear receptor E75, confirming the bioinformatic

functional annotation on GFOSS. The unassembled read GFOSS_seq406923 (47 amino acids long) also clustered with E75 sequences (data not shown) and therefore it also does not correspond to a *G. fossarum* EcR sequence. As expected, both these *G. fossarum* sequences showed the closest relationship with E75 from *Hyalella azteca*, the only genome-sequenced crustacean amphipod. These results show that no EcR transcripts were sequenced in the GFOSS RNAseq-derived transcriptome database. Because seq406923 is a short non-assembled read prone to have sequence errors, it was not used for primer design and sequence amplification. However, the E75 annotated sequence GFOSS_2900 was kept for subsequent studies. E75 is in fact one of the primary target genes of the EcR/RXR heterodimer, and regulates further downstream transcriptional cascades for molting and reproductive processes in both insects and decapod crustaceans [30-32].

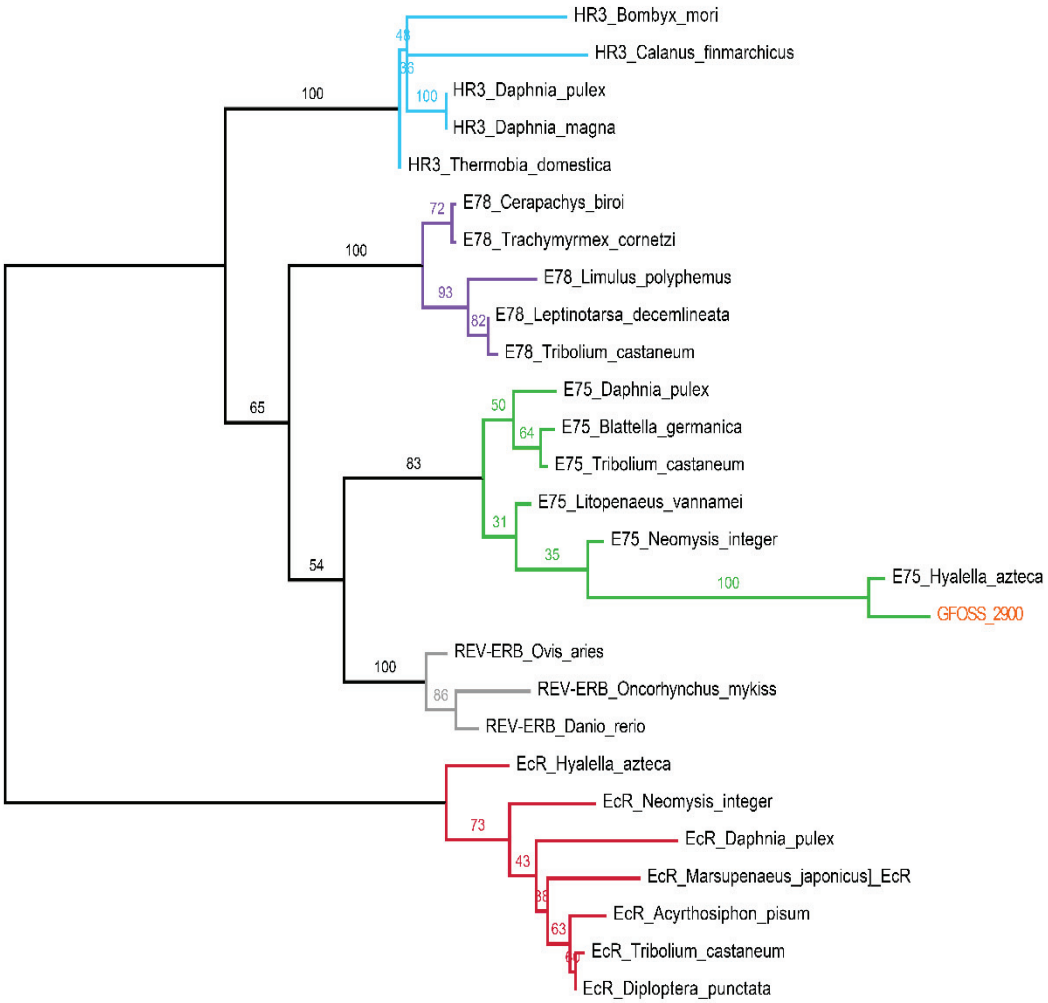


Figure 2: Phylogenetic tree of the GFOSS_2900 sequence. Tree construction was based on a homologous sequence dataset constructed using ortholog and paralog sequences from the same multigenic family as EcR (nuclear receptors E78, E75 and HR3). REV-ERB was used as an out group.

Figure 3 shows the phylogenetic tree obtained from the sequence dataset constructed for BR. CHINMO, TTK and lola genes from crustaceans and insects were considered to build this tree for the BTB-Zf family. As expected, the *G. fossarum* sequence GFOSS_1452 clustered together with the BR sequence from *H. azteca*, along with the other BR branches for crustaceans and insects. Similarly to the EcR results, the unassembled reads seq986730 and seq1061041 were not considered for the subsequent studies. Of note, this tree evidences that available annotation of crustacean sequences from this BTB-Zf family could be questioned due to potential duplication events

unfound in the insects, and which apparently may have led to misannotations. For example, lola genes from *Caligus clemensi* and *Lepeophteirus salmonis* clustered with the crustacean BR branch. This means that this specific family of genes is still relatively unknown and poorly described in crustacean species, and some of the annotations available by successive sequence similarity derivation when annotating new genomes could be erroneous. However, given the good blast score obtained and the posterior phylogenetic analysis for sequence GFOSS_1452, we validated the BR annotation for this sequence.

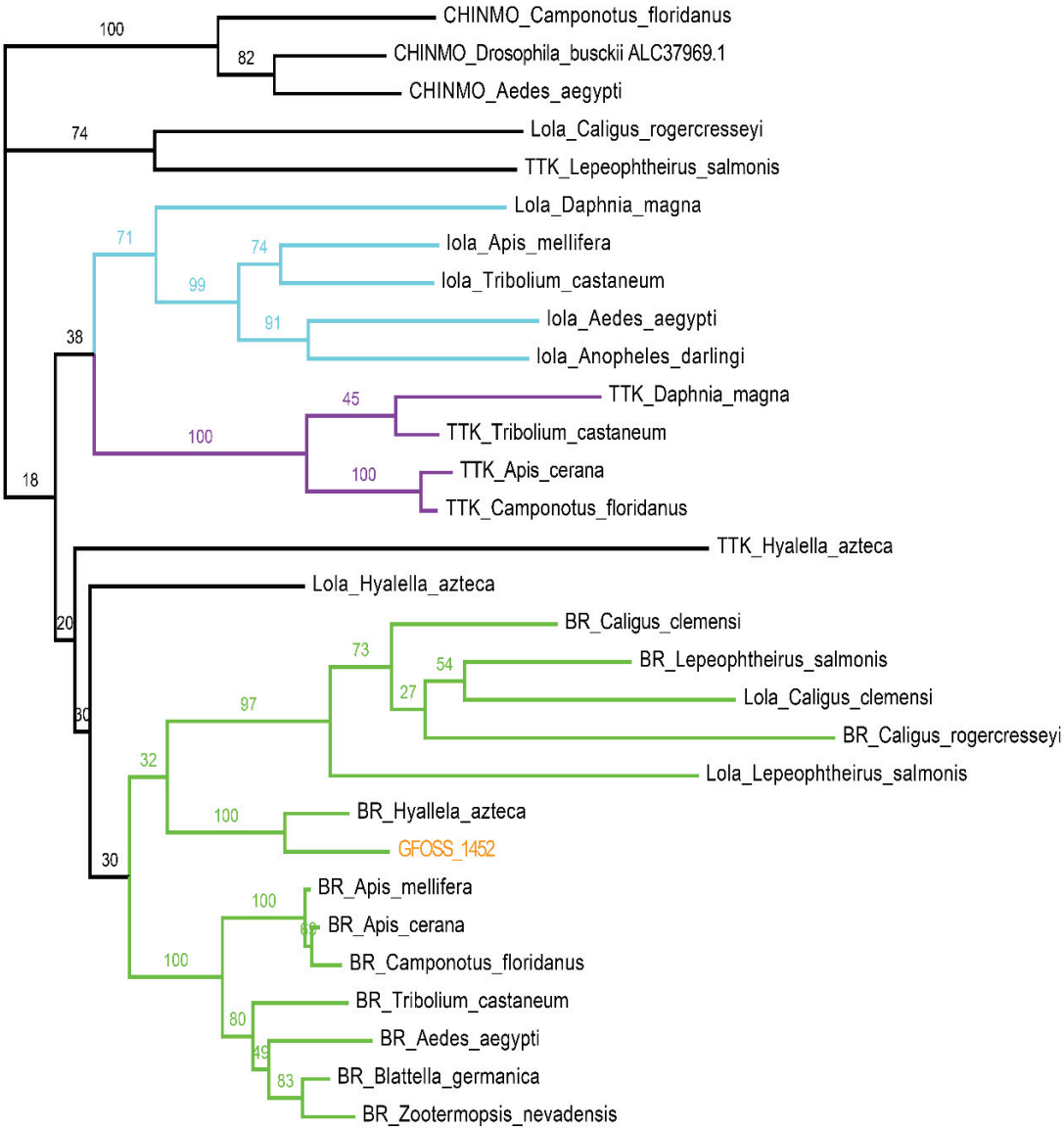


Figure 3: Phylogenetic tree of the GFOSS_1452 sequence. Tree construction was based on a homologous sequence dataset constructed using ortholog and paralog sequences from the same multigenic family as BR (transcription factors tramtrack TTK and longitudinals-like protein lola). CHINMO was used as an out group.

In **Figure 4** are represented two phylogenetic trees, one for each RXR *G. fossarum* sequence identified by blast. Two sequences were retained for this gene in order to obtain a higher alignment length. In fact, sequence GFOSS_83504 covers only the DBD domain of the RXR gene, while the unassembled read GFOSS_seq307974 covers another part of the gene, the ligand-binding domain (LBD). Analyzing the trees, both sequences clustered with the RXR branch, and were thus validated as being RXR sequences.

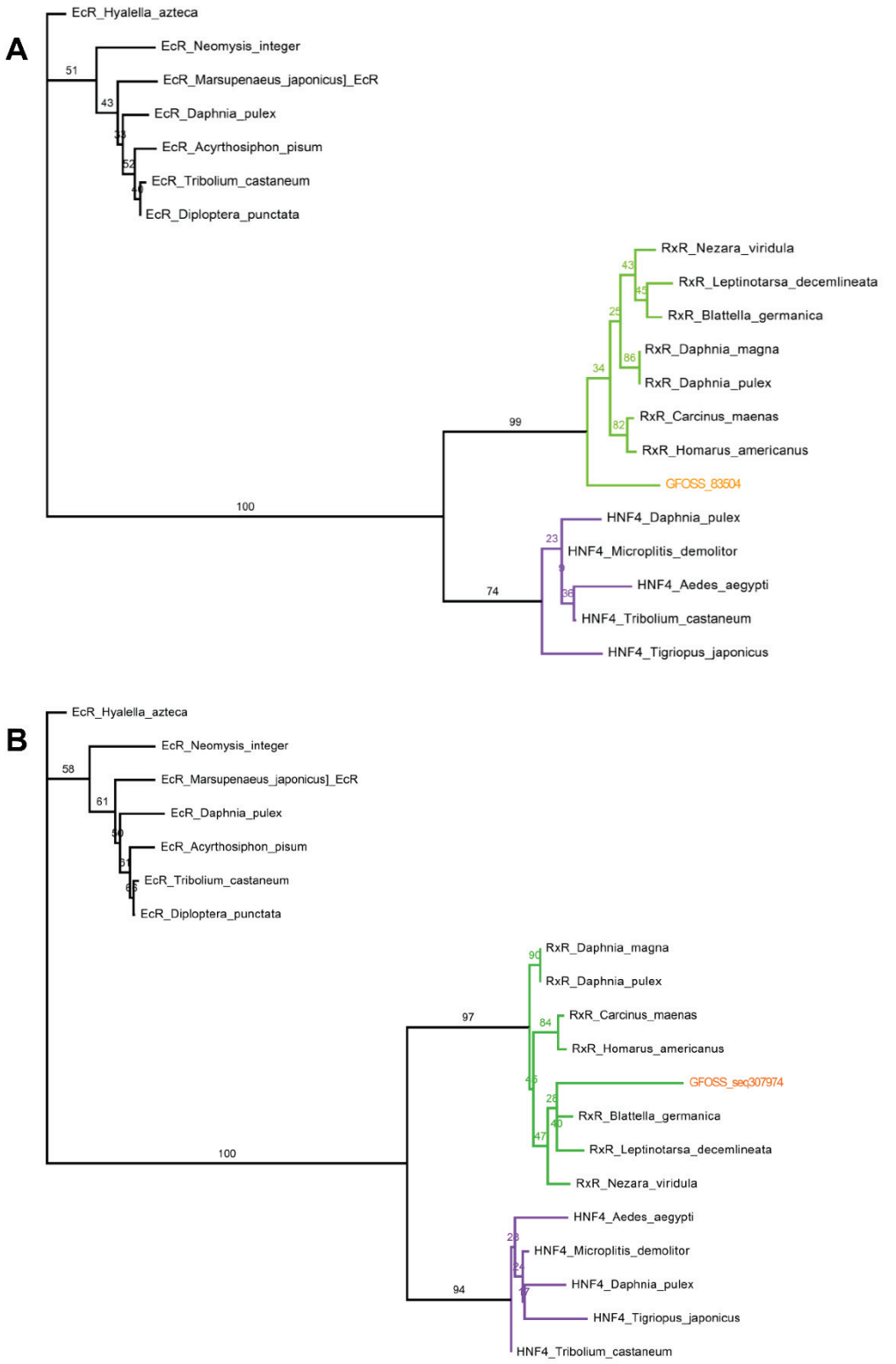


Figure 4: Phylogenetic tree of the (A) GFOSS_83504 sequence and (B) GFOSS_seq307974 read. Tree construction was based on a homologous sequence dataset constructed using ortholog and paralog sequences from the same multigenic family as RXR (nuclear receptor HNF4). EcR was used as an out group.

RXR, E75 and BR were successfully cloned and sequenced using the primers from **Table 2**. For E75, a 419 nucleotide sequence was obtained, covering a portion of both the DBD and LBD of this nuclear receptor. Smaller sequences were obtained for BR and RXR (278 and 220 nucleotides, respectively). The 220 nucleotide RXR sequence corresponds to primers designed in the DBD

domain, using only the GFOSS_83504 sequence. Other primer pairs were tested using forward sequences on the GFOSS_83504 and reverse sequences on seq307974, in order to obtain a sequence covering both DBD and LBD. However, we failed to amplify the RXR sequence with the pairs of primers tested.

Table 2: Nucleotide sequences of the primers used in the PCR and qPCR experiments.

	Primer ID	5' -> 3' nucleotide sequence	Fragment (bp)	Ct standard deviatio	
				Experiment 1	Experim
Target genes	RXR FW1	AAACCTTTGTGCCATCTGTG	220	–	–
	RXR RV3	GAACAGCCTCCCTCTTCATG			
	BRC FW1	GGGAGCAGACCAACAGTTCT	278	–	–
	BRC RV2	TGACCTCGCCATGGTACATA			
	E75 FW1	TCGTCAATGCAGCAAGAATC	419	–	–
	E75 RV2	GGATGTTCTTGGCGAAGGT			
Reference genes	EF FW	TTCAAGTATGCCTGGGTGCT	82	0,42	0,34
	EF RV	CGAACTTCCAGAGAGCAATGTC			
	GADPH FW	CGCTGGCCAGAACATCATTC	–	0,84	0,25
	GADPH RV	CGGCCTTGATGTCGTCGTAA			
	18S FW	TGGGGGAGGTAGTGACGAAATC	–	0,81	0,71
	18S RV	CCTGCGCTCGATACAGACATTC			

Expression during female reproductive cycle

In order to assess the possible involvement of the three ecdysone-inducible genes in reproductive processes, we quantified their expression during the female reproductive/molt cycle. The bar graphs in **Figure 5** represent the expression levels of the three genes through the different stages. E75 and RXR transcripts showed important modulations during the female reproductive cycle, while BR remained relatively stable.

E75

The expression pattern of E75 presents a strong peak at the C2/D1 stage, with a fold change of 12 relative to stage B. The C2 developmental stage is characterized by the quick increase in yolk formation and oocyte size, namely the start of secondary vitellogenesis [17,33]. This continuous increase from the C1/C2 until the C2/D1 stage suggests that the expression of E75 could be related to the onset of secondary vitellogenesis and oogenesis. This expression peak was validated through the analysis of individual replicates, where statistically significant difference was observed between C2D1 and D1D2 stages (p value<0.0001, **Supplementary Figure 2**). A second slight increase in the levels of this transcript in the final D2 stage (Figure 5) could indicate the existence of a second short peak of expression, as observed by the individual-specific increases obtained in biological replicates (**Supplementary Figure 2**). However, no statistically significant results corroborate the existence of this peak. It has been demonstrated that E75 plays a role in molting, oogenesis and vitellogenesis of several insects [34-37]. It has been also suggested that E75 plays an important role in the interplay between the ecdysone and JH pathways [38,39].

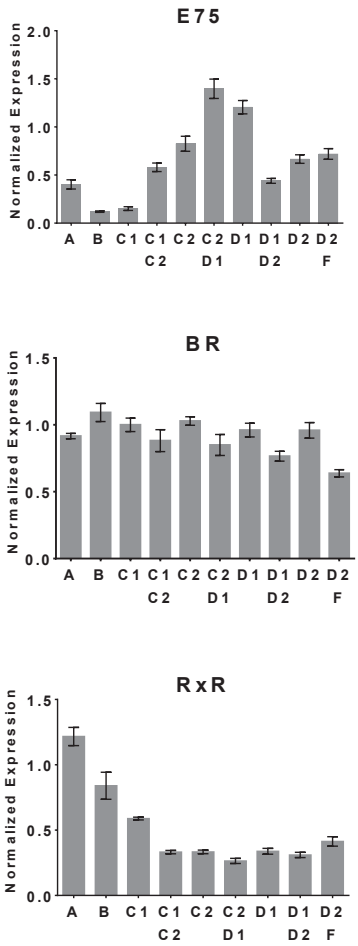


Figure 5: Expression profiles of E75, Br and RXR during the female reproductive cycle. Expression values of target genes were normalized to the expression of the reference gene EF. Each bar is the mean ± SD obtained from three experimental replicates. Each condition (A, B, C1, C1/C2, C2, C2/D1, D1, D1/D2, D2 and D2 F) contains mRNAs from five individuals pooled together.

In *Litopenaeus vannamei*, a crustacean decapod that is grouped with amphipods in Malacostraca crustaceans, E75 is positively regulated by ecdysone, regulates subsequent chitinase activities, and has molt stage-specific expression patterns [32]. The levels of EcR and E75 elevated significantly during vitellogenic stages I and II in the hepatopancreas of the crab *Oziothelphusa senex senex* [31]. Priya et al. [40] showed that E75 is modulated through the molt process of *Fenneropenaeus chinensis*, and silencing the gene leads to a complete arrest of molting. However, other studies [41] showed that E75 is marginally induced by ecdysteroids and presents constant

levels of expression during the molt cycle and embryo development of the branchiopod *Daphnia magna*. The results obtained herein suggest that, similarly to other malacostraca crustaceans, E75 may play a role in female reproductive-related processes of *G. fossarum*, notably in vitellogenesis.

BR

BR presented an up-and-down expression pattern, tending to a slight decrease in the later stages of the molt cycle. However, no deep modulations can be observed. BR is known to play a role in ecdysone and JH responsive processes such as metamorphosis [42,43], vitellogenesis and oogenesis [42,44] in several insects. Little is known about this gene in crustacean species, however it has been suggested that BR acts as a regulator of the action of methyl farnesoate (MF) [28]. Another study highlighted the importance of BR during ovarian development of *P. monodon* [45]. In that study, BR expression was more abundantly expressed in the ovaries than other tissues and injections of serotonin, progesterone and ecdysone all increased BR expression levels. This little information on BR function in crustacean species is complicated by the unknown evolutionary origin of this gene. Because there was no evidence of a BR gene ortholog in the non-insect and non-crustacean arthropod genomes, it has been suggested that BR was gained in the Pancrustacea [46]. Our results also highlight the limitations of using molecular information obtained from insects (or other phylogenetically distant species) for functional annotation of crustacean sequences (cross-species homology). The fact that the GFOSS putative BR sequence was not modulated during the reproductive processes opens several questions. The fact that this gene is a complex locus with many isoforms can be a limiting factor for the analysis. In order to investigate the role of different possible isoforms in *G. fossarum*, it would be necessary to test other BR-annotated sequences obtained by sequence homology and to check their expression patterns. The fact that the whole body of gammarids was used could have also diluted the response of this gene, and tissue specific modulations were therefore not detected.

RXR

An intense peak at the post-molt stage A was observed for the RXR transcript. This peak was validated through the analysis of individual replicates, where a significant difference was observed between stage A and C2D1 (p value<0.0001 , **Supplementary Figure 2**). This stage lasts for only one to two days and is the shortest of the *G. fossarum* molting cycle. After this expression peak, RXR expression gradually decreased until the C1/C2 stage, stabilizing until the end of the cycle. This expression pattern suggests that RXR may be involved in ecdysis, since the expression peak takes place during this process. There are several studies on both RXR tissue-specific expression, and during the reproductive cycle of crustaceans. However, different trends are observed between studies, due to inter-species variability and to the existence of several RXR isoforms. For example, in the lobster *Homarus americanus*, RXR was more expressed in the mandibular organ (MO) than in other tissues [47]. In MO and in ovaries, an expression peak occurred in the intermediate stage of the reproductive cycle (secondary vitellogenesis), while for hepatopancreas the peak occurred only during the early stages [47]. In the decapod shrimp *Fenneropenaeus chinensis*, two isoforms of RXR were more expressed in both the gills and nerve cord, and their expression during the molt cycle using whole-body organisms showed a peak at late premolt and postmolt stages, suggesting a role in ecdysis [48]. Another decapod, *Portunus trituberculatus*, presented higher expressions in both the Y-organ (YO) and the ovaries, as well as a further expression peak in the YO during the post-molt stages [49]. The authors highlight that RXR has an effect on molting but its expression does not correspond to the ecdysone titers observed in their previous study [50]. Some studies already proposed that MF and RXR may have a specific signaling pathway independent of the ecdysone pathway in decapods. This pathway is thought to participate in Vtg gene expression and ovarian development [47,51,52]. Nevertheless, in line with most crustacean literature published, the results presented herein suggest that RXR might be involved in the ecdysone-mediated ecdysis process of *G. fossarum*.

Expression following exposure to endocrine disrupting chemicals

In order to understand the susceptibility of E75, RXR and BR to EDCs, gammarids were exposed to three chemicals known to act on endocrine control of insect development. The chemicals were used at sub-lethal concentrations (PYR 5 µg/L, TEB 0.5 µg/L and PBO 150 µg/L) chosen based on previous experiments made in our laboratory, and therefore known to provoke reproductive impairments in female *G. fossarum*, particularly in vitellogenesis and oogenesis (**Supplementary Figure 3**).

Pyriproxyfen (PYR) is a potent juvenile hormone analog with insecticidal activity and little

mammalian toxicity [53], being very effective in inhibiting metamorphosis and embryogenesis in insects. Tebufenozide (TEB) is an analog for the molting hormone 20E. In lepidopteran species, the action of TEB results in the initiation of a precocious and incomplete molt (the new cuticle is formed before the shed of the old cuticle). This pesticide is not cleared by the organism, consequently inducing a continuous perturbation that maintains the expression of "early" response genes and the repression of genes normally expressed after the ecdysone peak for completing the molt [54]. Piperonyl butoxide (PBO) has no direct pesticidal activity of its own, but acts as a synergist of other insecticides by inhibiting some detoxification enzymes, namely the cytochrome P450 monooxygenases and some esterases [55].

The durations of exposure were chosen based on the results presented above (**Figure 5**). Our aim was to target the peaks of maximum expression for RXR in the beginning of the molting cycle (4 days, stage B) and for E75 in the onset of vitellogenesis (14 days, stage C2/D1). For RXR, since the stage A is very short, the stage B was selected for the first termination date. The nine day exposure (C1 molt stage) was intended to be an intermediate stage between the latter two in order to verify the occurrence of possible RXR/E75 inductions, as well as to confirm the trends between stages observed in the previous experiment (comparing the controls).

The values for the normalized expression in each condition, for the three target genes and at every exposure time, are shown in **Figure 6**. Analyzing solely the control samples we can observe the same trends as before between the B, C1 and C2/D1 stages for the three genes analyzed, reinforcing the repeatability of the described expression pattern during female cycle. The more evident example is E75 expression, very similar between day-4 and day-9 experiments, with the peak appearing at day-14 when females are at their C2/D1 reproductive stage (fold change of 13 between day-4 and day-14). For both RXR and BR control samples, the expression pattern throughout the reproductive stage is also similar to the previous experiment, as shown in **Table 3**. Comparing control and exposed samples, some effects of contamination were observed, suggesting that the female endocrine system was disturbed upon exposure to the pesticides.

To this respect, the **E75** transcript was the more responsive to contamination. At day-4 of exposure there was an increase in the order of 1.5 to 2-fold of E75 expression in all three contaminated conditions, a trend that repeated itself at day-9 (fold changes between 1.6 and 2.2). At day-14, transcript levels returned to normal in PYR and PBO exposures, while the TEB exposure led to a 2.8-fold decrease of E75 expression. The first increase in E75 expression means that the contaminants may induce, in a first phase, an over-expression of the "early" ecdysone-responsive genes in *G. fossarum*. These effects are concomitant with the role of this insecticide as an analog of the molt hormone 20E. However, the inhibition observed for the TEB exposed organisms at day-14 suggests that the ecdysteroid agonist effect of TEB is no longer taking place. Although in other species TEB is known to induce precocious molts through the continuous expression of the "early" ecdysone-responsive genes, in the studied species we have rather observed an inhibition of the early response-gene E75 after a 14-day exposure, which could reflect a general disruption of the normal ongoing of the female reproductive molting cycle.

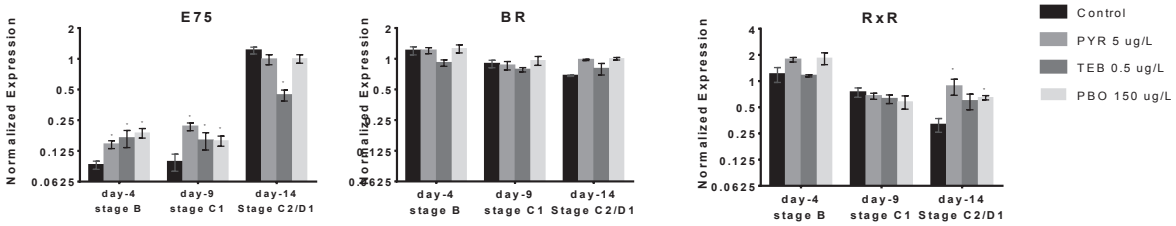


Figure 6: Transcriptional values of E75, RXR and BR in control and exposed organisms for the three exposure times studied. Expression values of target genes were normalized to the expression of the reference gene EF. Each bar is the mean ± SD obtained from three experimental replicates. Each condition (control, PYR, TEB, PBO) contains mRNAs from five individuals pooled together. Statistical significance was accepted with fold changes > 1.5 and p-value>0.05.

E75 is a key molecule in the reproductive processes of arthropods, and since is an early ecdysone-responsive gene, is prone to be affected by interactions of chemical compounds with the ecdysteroid receptor which is at the upstream of the ecdysone regulatory cascade. In a recent study [56], an Adverse Outcome Pathway (AOP) was developed for explaining the molecular events that

lead to lethal molting due to ecdysone receptor agonism in arthropods. The mechanism proposed was based on the literature available for both insects and crustaceans. The expression of the E75 gene was considered the first key event of the molecular chain leading to the adverse outcome, reinforcing its potential to be used as a biomarker of endocrine disturbances.

Table 3: Fold change expressions observed between different reproductive stages (B, C1 and C2/D1) in the two experiments performed.

		Experiment 1 Reproductive cycle	Experiment 2 Contaminant exposures
		fold change	fold change
E75	B/C1	1,27	1,07
	B/C2D1	11,68	13,30
	C1/C2D1	9,22	12,38
BR	B/C1	-1,19	-1,35
	B/C2D1	-1,32	-1,76
	C1/C2D1	-1,11	-1,31
RXR	B/C1	-1,43	-1,61
	B/C2D1	-3,18	-3,81
	C1/C2D1	-2,23	-2,36

Some effects were also seen in the expression of the **RXR** transcript at day-14 of exposure. PYR and PBO led to an over-expression of RXR at the C2/D1 stage (fold changes of 2.8 and 2.1 respectively). As discussed before, RXR and MF may have an ecdysone-independent regulatory pathway, which can explain the over-expression of RXR following exposure to the MF analog PYR. However, the possible mechanism that led to an induction of RXR provoked by PBO is not clear. It is known that this molecule does not have any insecticidal activity, but it inhibits the cytochrome P450 oxydases [57], which participate in the ecdysone synthesis and inactivation pathways. If this occurs, we would rather expect an inhibition of RXR expression due to an inhibition of the 20E expression.

Concerning **BR** expression, some statistically significant modulations were observed, but with fold changes < 1.5. This slight over-expression of this transcript at day-14 of exposure with PYR and PBO (fold changes of 1.43 and 1.47 respectively and p value < 0.001) follows the modulations observed for RXR, but establishing an eventual connection is difficult since the other early gene E75 was not modulated by the same exposure.

Conclusion

Based on a RNAseq-derived transcriptomic database and phylogenetic analyses, we identified potential key players of the hormonal system of *G. fossarum*: the nuclear receptors E75, RXR and the regulator BR. Our analyses, in line with most of the information available in the literature, provided some relevant insights into the possible roles of E75 and RXR in vitellogenesis, oogenesis and ecdysis. Exposure to environmental concentrations of three EDCs provoked significant modulations in the levels of these transcripts. The role of BR in reproductive *G. fossarum* female processes was not elucidated in this study. Allied to the fact that this gene revealed less sensitivity to contamination than the others, BR is not proposed as a biomarker of EDC exposure in female *G. fossarum*.

These results provide the first elements for understanding the molecular mechanisms of endocrine regulation in gammarids. Despite calling for more mechanistic studies in order to unravel the roles of these genes in specific physiological processes, their susceptibility to being modulated by an exposure to EDCs was verified in this study. The rupture of the expression cycle of these genes is therefore proposed as a biomarker of EDCs effects in female *G. fossarum*. To our knowledge, our study is the first to analyze the impact of endocrine disruptors in ecdysone-responsive genes of *G. fossarum*.

Acknowledgements

We thank the “Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture” (Irstea, France), the “Institut de Génomique Fonctionnelle” (IGFL, France), the “Commissariat pour l'Energie Atomique et aux Energies Alternatives” (CEA, France), and the ANR program “ProteoGam” (ANR-14-CE21-0006-02) for financial support. We also want to thank Abderrahman Khila, Fabienne Gas, Virginie Nouvel, Séverine Viala and Florent Decugis for providing useful insights and technical support in the experiments performed.

The authors declare no conflict of interest.

References

Armengaud J, Trapp J, Pible O, Geffard O, Chaumot A, Hartmann EM. 2014. Non-model organisms, a species endangered by proteogenomics. J Proteomics 105:5-18.

Bialecki M, Shilton A, Fichtenberg C, Segraves WA, Thummel CS. 2002. Loss of the ecdysteroid-inducible E75A orphan nuclear receptor uncouples molting from

metamorphosis in Drosophila. Developmental Cell 3(2):209-220.

Birnbaum LS. 2013. State of the Science of Endocrine Disruptors. Environmental health perspectives 121(4):a107-a107.

Bonneton F, Chaumot A, Laudet V. 2008. Annotation of Tribolium nuclear receptors reveals an increase in evolutionary rate of a network controlling the ecdysone cascade. Insect biochemistry and molecular biology 38(4):416-429.

Buaklin A, Sittikankaew K, Khamnamtong B, Menasveta P, Klinbunga S. 2013. Characterization and expression analysis of the Broad-complex (Br-c) gene of the giant tiger shrimp Penaeus monodon. Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology 164(4):280-289.

Buszczak M, Freeman MR, Carlson JR, Bender M, Cooley L, Segraves WA. 1999. Ecdysone response genes govern egg chamber development during mid-oogenesis in Drosophila. Development 126(20):4581-4589.

Cottin D, Foucreau N, Hervant F, Piscart C. 2015. Differential regulation of hsp70 genes in the freshwater key species Gammarus pulex (Crustacea, Amphipoda) exposed to thermal stress: effects of latitude and ontogeny. Journal of Comparative Physiology B 185(3):303-313.

Coulaud R, Geffard O, Xuereb B, Lacaze E, Quéau H, Garric J, Charles S, Chaumot A. 2011. In situ feeding assay with Gammarus fossarum (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring. Water Research 45(19):6417-6429.

Cravedi J-P, Zalko D, Savouret J-F, Menuet A, Jégou B. 2007. Le concept de perturbation endocrinienne et la santé humaine. Med Sci (Paris) 23(2):198-204.

Dubrovskaya VA, Berger EM, Dubrovsky EB. 2004. Juvenile hormone regulation of the E75 nuclear receptor is conserved in Diptera and Lepidoptera. Gene 340(2):171-177.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5):1792-1797.

Ford AT, Fernandes TF, Rider SA, Read PA, Robinson CD, Davies IM. 2004. Endocrine disruption in a marine amphipod? Field observations of intersexuality and de-masculinisation. Marine environmental research 58(2-5):169-73.

Geffard O, Xuereb B, Chaumot A, Geffard A, Biagianti S, Noël C, Abbaci K, Garric J, Charmantier G, Charmantier-Daures M. 2010. Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment. Environmental toxicology and chemistry / SETAC 29(10):2249-59.

Girish BP, Swetha CH, Reddy PS. 2015. Expression of RXR, EcR, E75 and VtG mRNA levels in the hepatopancreas and ovary of the freshwater edible crab, Oziothelphusa senex senex (Fabricius, 1798) during different vitellogenic stages. Science of Nature 102(3-4).

Gong J, Huang C, Shu L, Bao C, Huang H, Ye H, Zeng C, Li S. 2016. The retinoid X receptor from mud crab: New insights into its roles in ovarian development and related signaling pathway. Scientific reports 6.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010 Feb;27(2):221-4. doi: 10.1093/molbev/msp259. Epub 2009 Oct 23.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59(3):307-21.

Guo WC, Liu XP, Fu KY, Shi JF, Lü FG, Li GQ. 2016. Nuclear receptor ecdysone-induced protein 75 is required for larval-pupal metamorphosis in the Colorado potato beetle Leptinotarsa decemlineata (Say). Insect Molecular Biology 25(1):44-57.

Hannas BR, LeBlanc GA. 2010. Expression and ecdysteroid responsiveness of the nuclear receptors HR3 and E75 in the crustacean Daphnia magna. Molecular and Cellular Endocrinology 315(1-2):208-218.

Hardstone MC, Strycharz JP, Kim J, Park IK, Yoon KS, Ahn YJ, Harrington LC, Lee SH, Clark JM. 2015. Development of multifunctional metabolic synergists to suppress the evolution of resistance against pyrethroids in insects that blood feed on humans. Pest Management Science 71(6):842-849.

Höss S, Weltje L. 2007. Endocrine disruption in nematodes: Effects and mechanisms. Ecotoxicology 16(1):15-28.

Hyne RV. 2011. Review of the reproductive biology of amphipods and their endocrine regulation: identification of mechanistic pathways for reproductive toxicants. Environmental toxicology and chemistry / SETAC 30(12):2647-57.

Issartel J, Boulo V, Wallon S, Geffard O, Charmantier G. 2010. Cellular and molecular osmoregulatory responses to cadmium exposure in Gammarus fossarum (Crustacea, Amphipoda). Chemosphere 81(6):701-10.

Jobling S, Tyler CR. 2003. Endocrine disruption in wild freshwater fish. Pure and Applied Chemistry. p. 2219.

Jubeaux G, Simon R, Salvador A, Lopes C, Lacaze E, Quéau H, Chaumot A, Geffard O. 2012a. Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: Inconclusive experience. Aquatic Toxicology 122-123:9-18.

Jubeaux G, Simon R, Salvador A, Quéau H, Chaumot A, Geffard O. 2012b. Vitellogenin-like proteins in the freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males. Aquatic toxicology (Amsterdam, Netherlands) 112-113:72-82.

Jungmann D, Ladewig V, Ludwichowski K-U, Petzsch P, Nagel R. 2004. Intersexuality in Gammarus fossarum KOCH - a common inducible phenomenon? Archiv für Hydrobiologie 159(4):511-529.

King-Jones K, Thummel CS. 2005. Nuclear receptors--a perspective from Drosophila. Nat Rev Genet 6(4):311-23.

Kirby MF, Allen YT, Dyer RA, Feist SW, Katsiadaki I, Matthiessen P, Scott AP, Smith A, Stentiford GD, Thain JE et al. . 2004. Surveys of plasma vitellogenin and intersex in male flounder (Platichthys flesus) as measures of endocrine disruption by estrogenic contamination in United Kingdom estuaries: temporal trends, 1996 to 2001. Environ Toxicol Chem 23(3):748-58.

Kloas W, Urbatzka R, Opitz R, Würtz S, Behrends T, Hermelink B, Hofmann F, Jagnytsch O, Kroupova H, Lorenz C et al. . 2009. Endocrine disruption in aquatic vertebrates. Annals of the New York Academy of Sciences. p. 187-200.

Kunz PY, Kienle C, Gerhardt A. 2010. Gammarus spp. in aquatic ecotoxicology and water quality assessment: toward integrated multilevel tests. Reviews of environmental contamination and toxicology 205:1-76.

Lacaze E, Devaux A, Mons R, Bony S, Garric J, Geffard A, Geffard O. 2011. DNA damage in caged Gammarus fossarum amphipods: a tool for freshwater genotoxicity assessment. Environ Pollut 159(6):1682-91.

LeBlanc GA. 2007. Crustacean endocrine toxicology: A review. Ecotoxicology 16(1):61-81.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25(4):402-8.

Miyamoto J, Hirano M, Takimoto Y, Hatakoshi M. 1993. Insect Growth Regulators for Pest Control, with Emphasis on Juvenile Hormone Analogs. Pest Control with Enhanced Environmental Safety. American Chemical Society. p. 144-168.

Nagaraju GPC, Rajitha B, Borst DW. 2011. Molecular cloning and sequence of retinoid X receptor in the green crab Carcinus maenas: A possible role in female reproduction. Journal of Endocrinology 210(3):379-390.

Oehlmann J, Di Benedetto P, Tillmann M, Duft M, Oetken M, Schulte-Oehlmann U. 2007. Endocrine disruption in prosobranch molluscs: Evidence and ecological relevance. Ecotoxicology 16(1):29-43.

Oetken M, Bachmann J, Schulte-Oehlmann U, Oehlmann J. 2004. Evidence for endocrine disruption in invertebrates. International review of cytology 236:1-44.

Pierceall WE, Li C, Biran A, Miura K, Raikhel AS, Segraves WA. 1999. E75 expression in mosquito ovary and fat body suggests reiterative use of ecdysone-regulated hierarchies in development and reproduction. Molecular and Cellular Endocrinology 150(1-2):73-89.

Priya TAJ, Li F, Zhang J, Wang B, Zhao C, Xiang J. 2009. Molecular characterization and effect of RNA interference of retinoid X receptor (RXR) on E75 and chitinase gene expression in Chinese shrimp Fenneropenaeus chinensis. Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology 153(1):121-129.

Priya TAJ, Li F, Zhang J, Yang C, Xiang J. 2010. Molecular characterization of an ecdysone inducible gene E75 of Chinese shrimp Fenneropenaeus chinensis and elucidation of its role in molting by RNA interference. Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology 156(3):149-157.

Qian Z, He S, Liu T, Liu Y, Hou F, Liu Q, Wang X, Mi X, Wang P, Liu X. 2014. Identification of ecdysteroid signaling late-response genes from different tissues of the Pacific white shrimp, Litopenaeus vannamei. Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology 172:10-30.

Qu Z, Kenny NJ, Lam HM, Chan TF, Chu KH, Bendena WG, Tobe SS, Hui JHL. 2015. How Did Arthropod Sesquiterpenoids and Ecdysteroids Arise? Comparison of Hormonal Pathway Genes in Noninsect Arthropod Genomes. Genome Biology and Evolution 7(7):1951-1959.

Regier JC, Shultz JW, Kambic RE. 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc R Soc B 272.

Retnakaran A, Gelbic I, Sundaram M, Tomkins W, Ladd T, Primavera M, Feng Q, Arif B, Palli R, Krell P. 2001. Mode of action of the ecdysone agonist tebufenozide (RH-5992), and an exclusion mechanism to explain resistance to it. Pest Management Science 57(10):951-957.

Reza AMS, Kanamori Y, Shinoda T, Shimura S, Mita K, Nakahara Y, Kiuchi M, Kamimura M. 2004. Hormonal control of a metamorphosis-specific transcriptional factor Broad-Complex in silkworm. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 139(4):753-761.

Rodriguez EM, Medesani DA, Fingerman M. 2007. Endocrine disruption in crustaceans due to pollutants: a review. Comp Biochem Physiol A Mol Integr Physiol 146(4):661-71.

Scott AP, Sanders M, Stentiford GD, Reese RA, Katsiadaki I. 2007. Evidence for estrogenic endocrine disruption in an offshore flatfish, the dab (Limanda limanda L.). Mar Environ Res 64(2):128-48.

Short S, Yang G, Kille P, Ford AT. 2014. Vitellogenin is not an appropriate biomarker of feminisation in a crustacean. Aquatic toxicology (Amsterdam, Netherlands) 153:89-97.

Simon R, Jubeaux G, Chaumot A, Lemoine Jrm, Geffard O, Salvador A. 2010. Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: Application to vitellogenin in Crustacea (Gammarus fossarum). Journal of Chromatography A 1217:5109-5115.

Sin YW, Kenny NJ, Qu Z, Chan KW, Chan KWS, Cheong SPS, Leung RWT, Chan TF, Bendena WG, Chu KH et al. . 2014. Identification of putative ecdysteroid and juvenile hormone pathway genes in the shrimp Neocaridina denticulata. General and comparative endocrinology.

Smith BS. 1971. SEXUALITY IN THE AMERICAN MUD SNAIL, NASSARIUS OBSOLETUS SAY. Journal of Molluscan Studies 39(5):377-378.

Song Y, Villeneuve DL, Toyota K, Iguchi T, Tollefsen KE. 2017. Ecdysone Receptor Agonism Leading to Lethal Molting Disruption in Arthropods: Review and Adverse Outcome Pathway Development. Environmental Science and Technology 51(8):4142-4157.

Sumpter JP, Jobling S. 1995. Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. Environ Health Perspect 7:173-8.

Swevers L, Eystathioy T, Iatrou K. 2002. The orphan nuclear receptors BmE75A and BmE75C of the silkmoth Bombyx mori: Hornmonal control and ovarian expression. Insect biochemistry and molecular biology 32(12):1643-1652.

Tang J, Zhu DF, Cui XY, Xie X, Qiu XE. 2014. Molecular cloning, characterization and expression analysis of the retinoid X receptor in the swimming crab, portunus trituberculatus (Miers, 1876) (Decapoda, Portunidae). Crustaceana 87(3):312-327.

Tiu SHK, Hult EF, Yagi KJ, Tobe SS. 2012. Farnesoic acid and methyl farnesoate production during lobster reproduction: Possible functional correlation with retinoid X receptor expression. General and comparative endocrinology 175(2):259-269.

Trapp J, Geffard O, Imbert G, Gaillard JC, Davin AH, Chaumot A, Armengaud J. 2014. Proteogenomics of gammarus fossarum to document the reproductive system of amphipods. Molecular and Cellular Proteomics 13(12):3612-3625.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44(W1):W232-W235.

Xuereb Bt, Chaumot A, Mons R, Garric J, Geffard O. 2009. Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda). Intrinsic variability, reference levels, and a reliable tool for field surveys. Aquatic Toxicology 93:225-233.

Yang C, Lin Y, Liu H, Shen G, Luo J, Zhang H, Peng Z, Chen E, Xing R, Han C et al. . 2014. The Broad Complex isoform 2 (BrC-Z2) transcriptional factor plays a critical role in vitellogenin transcription in the silkworm Bombyx mori. Biochimica et Biophysica Acta (BBA) - General Subjects 1840(9):2674-2684.

Zhou B, Riddiford LM. 2001. Hormonal Regulation and Patterning of the Broad-Complex in the Epidermis and Wing Discs of the Tobacco Hornworm, Manduca sexta. Developmental Biology 231(1):125-137.

Zimmer CT, Panini M, Singh KS, Randall EL, Field LM, Roditakis E, Mazzoni E, Bass C. 2017. Use of the synergist piperonyl butoxide can slow the development of alpha-cypermethrin resistance in the whitefly Bemisia tabaci. Insect Molecular Biology 26(2):152-163.

Supplementary Table 1

GenBank accession numbers of the sequences used in sequence similarity searches and alignments/tree construction.

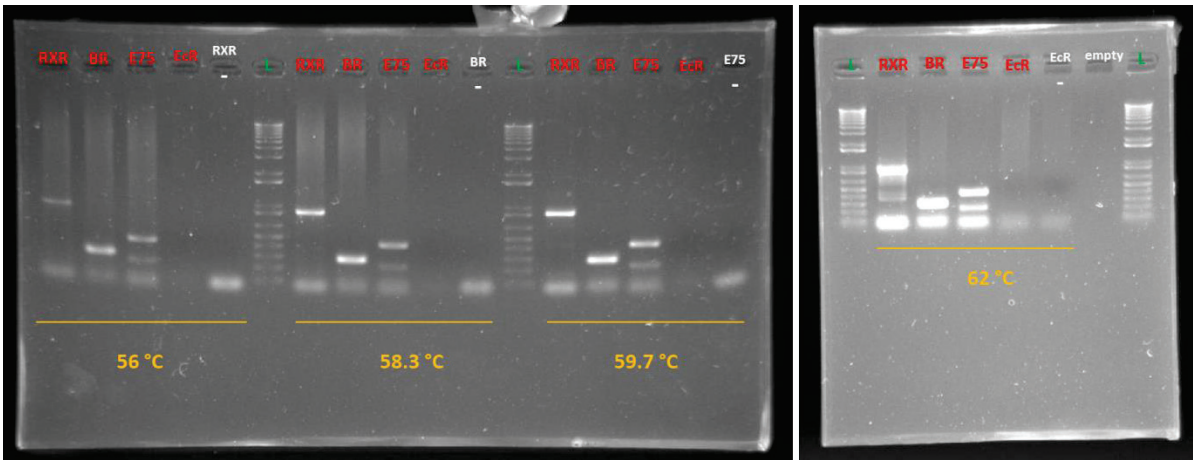
Name	GenBank accession nº	Name	GenBank accession nº
BR_Aedes_aegypti	AAS80326.1	HR3_Bombyx_mori	AAK62804.1
BR_Apis_mellifera	NP_001035356.1	HR3_Thermobia_domestica	BAP76394.1
BR_Tribolium_castaneum	NP_001104734.1	HR3_Daphnia_magna]	ACY56690.1
BR_Blattella_germanica	CBJ05857.1	HR3_Daphnia_pulex	ACY56691.1
BR_Zootermopsis_nevadensis	KDR08439.1	HR3_Calanus_finmarchicus	AHN16686.1
BR_Apis_cerana	XP_006558529.1	HR3 _Hyalella_azteca	XP_018025373.1
BR_Camponotus_floridanus	EFN67811.1	E78_Leptinotarsa_decemlineata	AQN67834.1
BR_Hyallela_azteca	XP_018008501.1	E78_Limulus_polyphemus	XP_013782694.1
BR_Neocaridina_denticulata	AIY69133.1	E78_Cerapachys_biroi	XP_011332703.1
BR_Lepeophtheirus_salmonis	ADD38891.1	E78_Tribolium_castaneum	EFA01351.2
BR_Caligus_clemensi	ACO15753.1	E78_Trachymyrmex_cornetzi	KYN26678.1
BR_Caligus_rogercresseyi	ACO11696.1	TTK_Tribolium_castaneum	NP_001157610.1
Lola_Hyalella_azteca	XP_018023991.1	TTK_Apis_cerana	XP_016907014.1
Lola_Daphnia_magna	JAN74307.1	TTK_Camponotus_floridanus	XP_011258243.1
Lola_Lepeophtheirus_salmonis	ACO12742.1	TTK_Lepeophtheirus_salmonis	CDW19648.1
Lola_Caligus_clemensi	ACO15552.1	TTK_Hyalella_azteca	XP_018012514.1
Lola_Caligus_rogercresseyi	ACO11667.1	TTK_Daphnia_magna	JAN81132.1
Lola_Apis_mellifera	XP_006567635.1	REV-ERB_Ovis_aries	ABV53446.1
Lola_Tribolium_castaneum	NP_001157310.1	REV-ERB_Danio_rerio	AAK76396.1
Lola_Aedes_aegypti	XP_001659827.1	REV-ERB_Oncorhynchus_mykiss	AAK76397.1
Lola_Anopheles_darlingi	ETN65068.1	RxR_Eriocheir_sinensis	AHF65151.1
E75_Hyalella_azteca	XP_018025430.1	RxR_Nezara_viridula	ADQ43369.1
E75_Neomysis_integer	ACL55050.1	RxR_Leptinotarsa_decemlineata	BAD99298.1
E75_Litopenaeus_vannamei	AGS94407.1	RxR_Daphnia_pulex	ABF74729.1
E75_Daphnia_pulex	ADB79814.1	RxR_Daphnia_magna	ABF74729.1
E75_Blattella_germanica	CAJ87513.1	HNF4_Daphnia_pulex	EFX72163.1
E75_Tribolium_castaneum	KYB25584.1	HNF4_Tigriopus_japonicus	AID52844.1
EcR_Hyalella_azteca	XP_018017941.1	HNF4_Tribolium_castaneum	KYB26571.1
EcR_Neomysis_integer	ACJ68423.1	HNF4_Aedes_aegypti	AAC72723.1
EcR_Marsupenaeus_japonicus	BAF75375.1	HNF4_Microplitis_demolitor	XP_008543491.1
EcR_Daphnia_pulex	EFX79409.1	CHINMO_synthetic_construct	NP_001188680.1
EcR_Tribolium_castaneum	NP_001135390.1	CHINMO_Drosophila_busckii	ALC37969.1
EcR_Diploptera_punctata	AEZ64364.1	CHINMO_synthetic_construct2	AOQ14382.1
EcR_Acyrthosiphon_pisum	NP_001152832.1		

Supplementary Figure 1

Fragments obtained after the temperature gradient PCR.

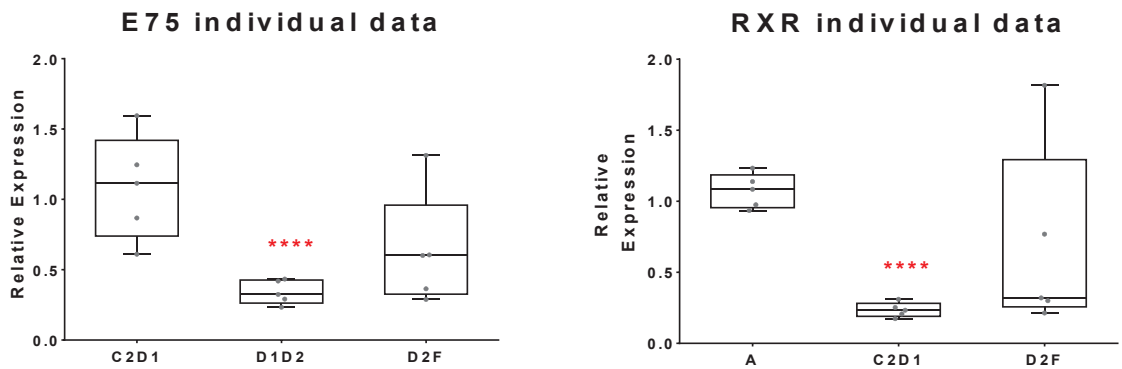
RXR -, BR -, E75 -, and EcR – correspond to negative (no DNA) controls.

L stands for ladder DNA.



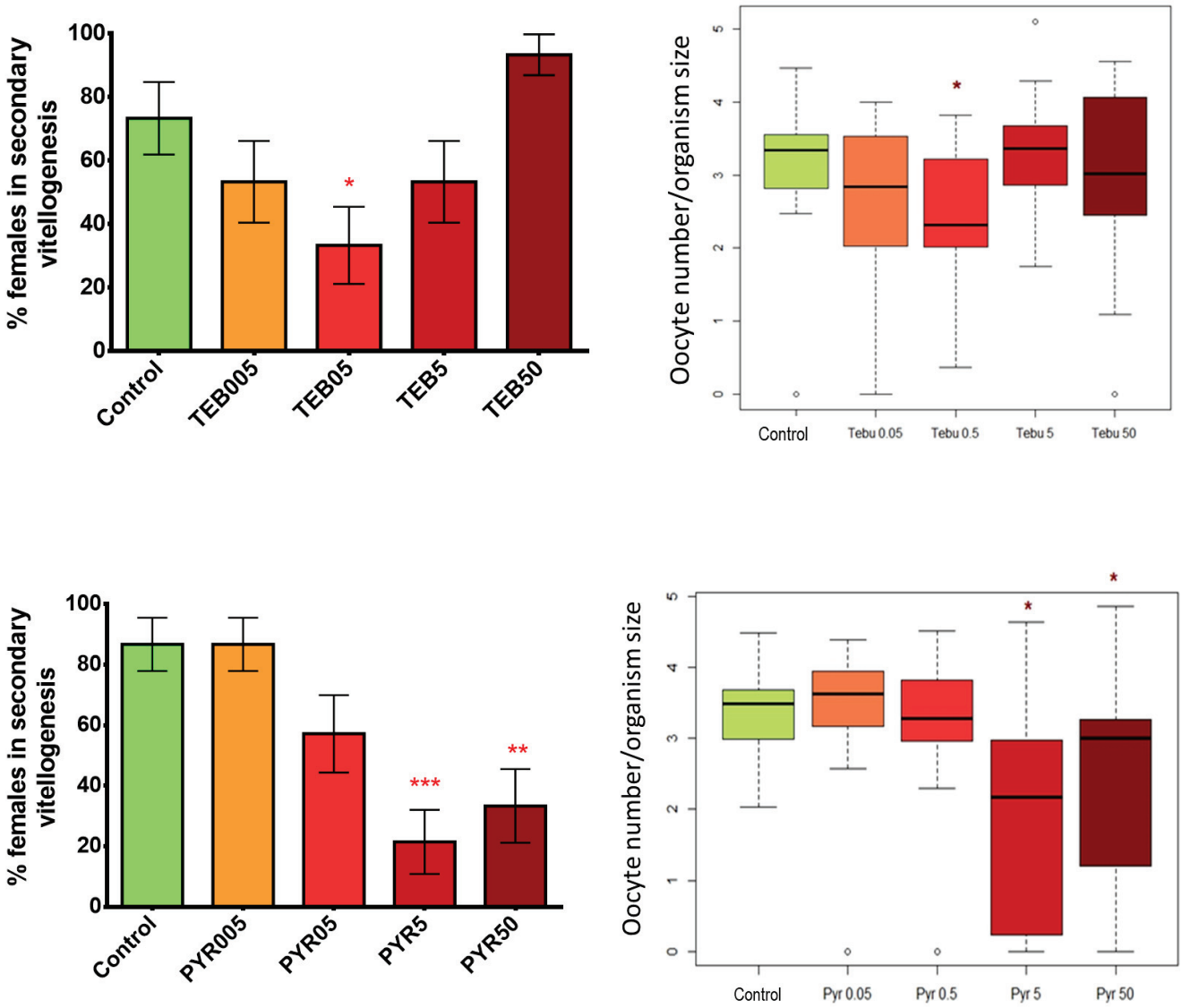
Supplementary Figure 2

Data of individual replicates from the physiological experiment. Statistically significant differences were calculated through an unpaired t test with Welch's correction. ****=p value<0.0001



Supplementary Figure 3

Impact of tebufenozide and pyriproxyfen on oocyte development (bar graphs) and oocyte number (normalized to the size of the female) (boxplots) in *G. fossarum* females. Statistical differences were performed using a unilateral proportion comparisons test for the comparison of percentages (bar graphs), and non-parametric Mann-Whitney tests for the comparison of number of oocytes (boxplots) between control and contaminated samples. Significantly statistical differences were accepted at $p<0.05$.



1. Development of a new MS-based multibiomarker absolute quantification in *Gammarus fossarum*

Analytical development and validation

A first approach, based on the identification of proteins of interest through automatic functional prediction, allowed the establishment of a list of 177 peptides reporting for 55 candidate biomarker proteins. For this proof of concept study, we selected candidate biomarkers with functional annotations that suggested a role in important physiological processes such as hormonal metabolism and reproduction. These proteins were selected from an extensive catalog of proteins that was built from previous physiological and toxicological whole-proteome experiments (Trapp et al., 2015; Trapp et al., 2014b).

The first stage of analytical optimizations performed in samples containing only the isotopically labelled peptides resulted in the elimination of 52 peptides due to poor ionizations or excessive charge state distribution. Briefly, a list of twenty transitions per peptide (3540 transitions total) was generated *in silico* using the Skyline software. A complete b- and y-product ion series for doubly and triply charged precursor ions with m/z ranges between 300-1200 was produced. Posteriorly the most suitable transitions for a targeted peptide were selected. This involved choosing the fragment ions for each precursor-ion charge state that provide the highest signal intensity and lowest level of interfering signals. Optimizations were performed in both isotopically labelled peptide samples and endogenous peptides obtained from *G. fossarum* protein extracts. In parallel, the MS parameters were optimized for the most sensitive and peptide-specific transitions. The top three performing transitions per peptide were selected for the tests in the biological matrix.

The robustness of the methodology for the parallel acquisition of the remaining 125 peptides and 750 transitions was then evaluated in fifty protein extracts from male and female organisms in contrasted physiological conditions. This procedure allowed verifying if the transitions "co-eluted" in the same peak, and if their contribution to the total area of the peak was constant among biological replicates. Figure V-1 exemplifies a peptide elution with a well-formed interference-free peak (peptide 1), and another peptide (peptide 2) that despite having also a well-formed peak, presents some small

interferences around its retention time. Moreover, as shown in the histogram from peptide 2, the area ratios of the three transitions are not equal among the several samples. Peptides presenting these profiles were eliminated.

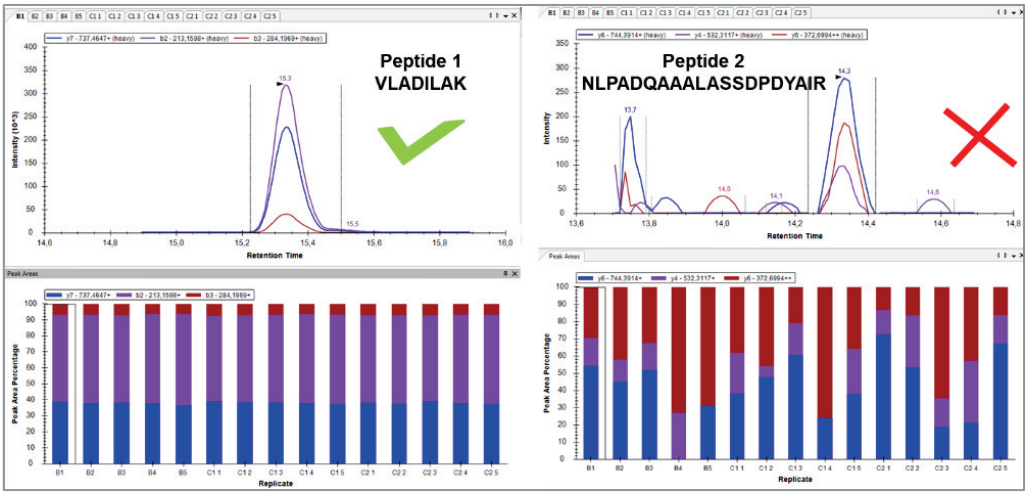


Figure V-1 – Example of chromatographic peaks obtained for two peptides in fifteen different samples (histograms). The different transitions are represented with different colors (violet, blue and brown). Histograms represent the percentage of contribution of each transition to the total area of the peak.

These data also allowed us to verify the correlation between peptides presumably reporters of a same protein (because they were anchored on a common contig from the GFOSS database). Depending on the characteristics of the peptides, they will behave differently during the sample preparation procedures, and sometimes losses occur. In some cases, mass spectrometry signals from different tryptic peptides from a same protein can differ as much as 100-fold in intensity (Picotti and Aebersold, 2012). Therefore, for each protein, the areas of proteotypic peptides were correlated between themselves for all sample extracts. Two peptides from a same protein will have the same area ratio regardless the concentration of the peptide, thus the correlation coefficient (r^2) between the areas of peptide pairs in all the samples was analyzed.

Figure V-2 exemplifies the correlations obtained between three peptides from a same protein. As we can observe, peptide TSEVFLPLTNELYQQTK correlates perfectly with peptide IFNVLQPIAESK, representing reliable reporter peptides for their corresponding protein. This means that they have similar specificity, ionization conditions, digestion efficiency, and that matrix effects are similar (or non-existent) for

both peptides. However, peptide ITMQEDGSGEVQLK has lower correlation coefficients with the other peptides from the protein. This suggests that this peptide could have some losses during the sample preparation procedure, or that it may be present in two or more distinct proteins. If the protein has several isoforms, it is also plausible that the peptides chosen as reporters are not present in all of them, giving rise to this kind of poor correlations. Only peptides presenting good correlations ($r^2 > 0.90$) were kept in the assay.

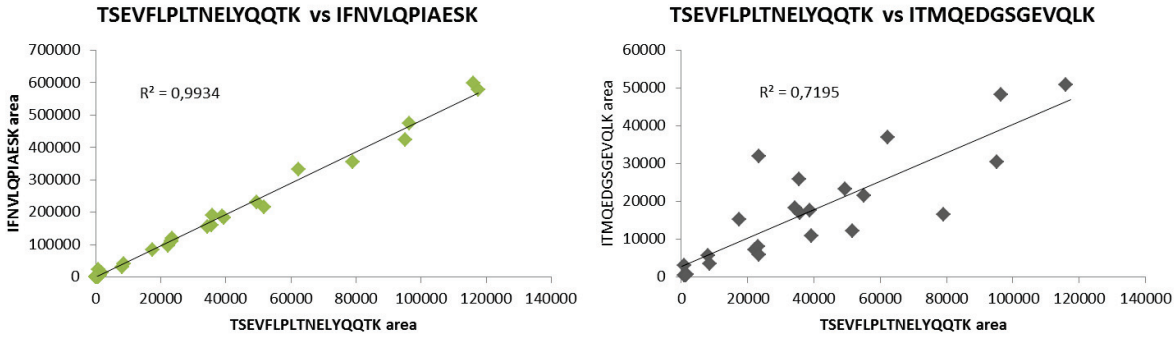


Figure V-2 – Example of the area correlations for three peptides from protein 17046. Data points were obtained from samples analyzed with the QTRAP 4000.

This analysis led to the identification of the most sensitive and specific peptides, and to a final assay that allowed the reliable quantification of a maximum of 71 peptides reporting for 40 proteins. Table V-1 lists the principal parameters of the LC-MS/MS run, and the maximum number of proteins simultaneously quantified with the two mass spectrometers used.

Table V-1 – Fundamental characteristics of the validated SRM assay.

	Peptide fractionation	**SPE desalting method**
	Chromatography	**2-mm-internal diameter C18 reverse phase column**
	LC run time	**35 minutes**
	Precursor charges	**+2 and +3**
4000 QTRAP	*CE voltage*	**20 - 64 V**
	Total transitions	**276**
	Total peptides	**46**
	Total proteins	**29**
5500 QTRAP	*CE voltage*	**10 - 60 V**
	Total transitions	**426**
	Total peptides	**71**
	Total proteins	**40**

Other analytical parameters such as the LOD, LOQ, and intra- and inter-assays reproducibility, were determined for each validated peptide and were detailed in publication nº1.

Verification of the role of candidate biomarkers in biological functions

The measurement of peptide modulations during the female reproductive/molt cycle allowed assessing the physiological importance of several protein biomarkers. The clearest example is the continuous increase in concentration of the Vtg-like proteins along the cycle, as shown in Figure V-3.

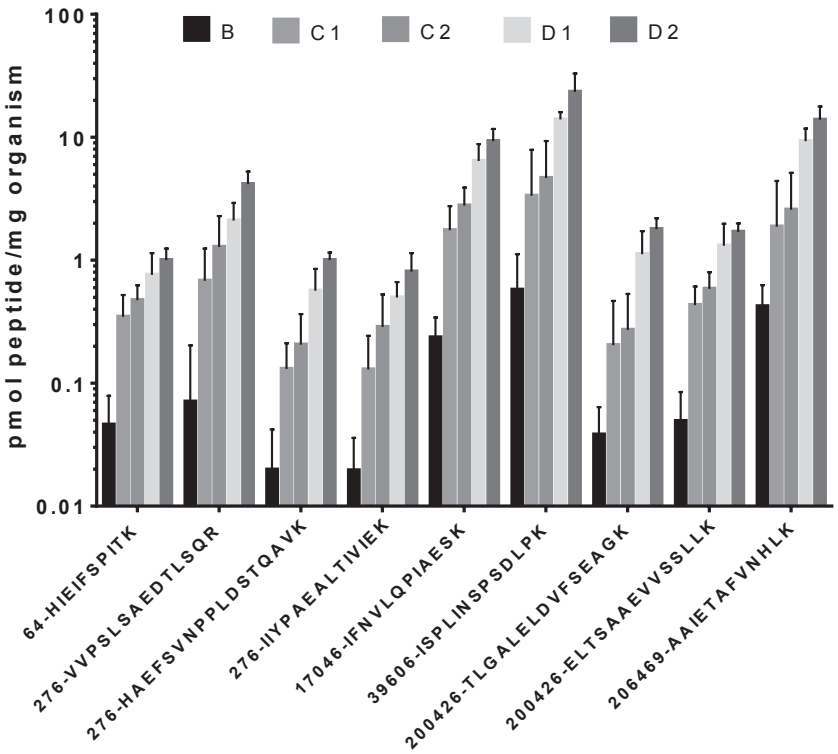


Figure V-3 – Percentage increases of the several Vtg-like proteins at each stage of the female reproductive/molt cycle.

In amphipods, the molting cycle occurs in parallel with oogenesis (progressive accumulation of vitellogenin proteins in oocytes) and embryonic development, because females mate and spawn immediately after ecdysis (Geffard et al., 2010). Sharp increases between the B/C1 and C2/D1 stages are related to the first and secondary

stages of the vitellogenesis process. Vitellogenins are precursors of major egg yolk proteins and are essential for providing the energy required for embryonic development. *G. fossarum* presents a large variety of Vtg-like proteins, fact that was already underlined in a recent publication (Trapp et al., 2016). Some of these proteins, validated as having yolk protein functions, originated from different families of the large lipid transfer protein superfamily (notably proteins phylogenetically related to clotting proteins or apolipocrustaceins in decapod crustaceans). Therefore, they were automatically annotated as being "clotting proteins". The absolute quantification results obtained herein reinforce and validate that these proteins are deeply implicated in vitellogenesis process and oocyte development/growth.

Osmoregulation-related Na^+K^+ ATPases, the digestive enzyme cellulase, and the immune-related prophenoloxidase and transglutaminase enzymes were also found modulated throughout the female reproductive cycle (Figure V-4).

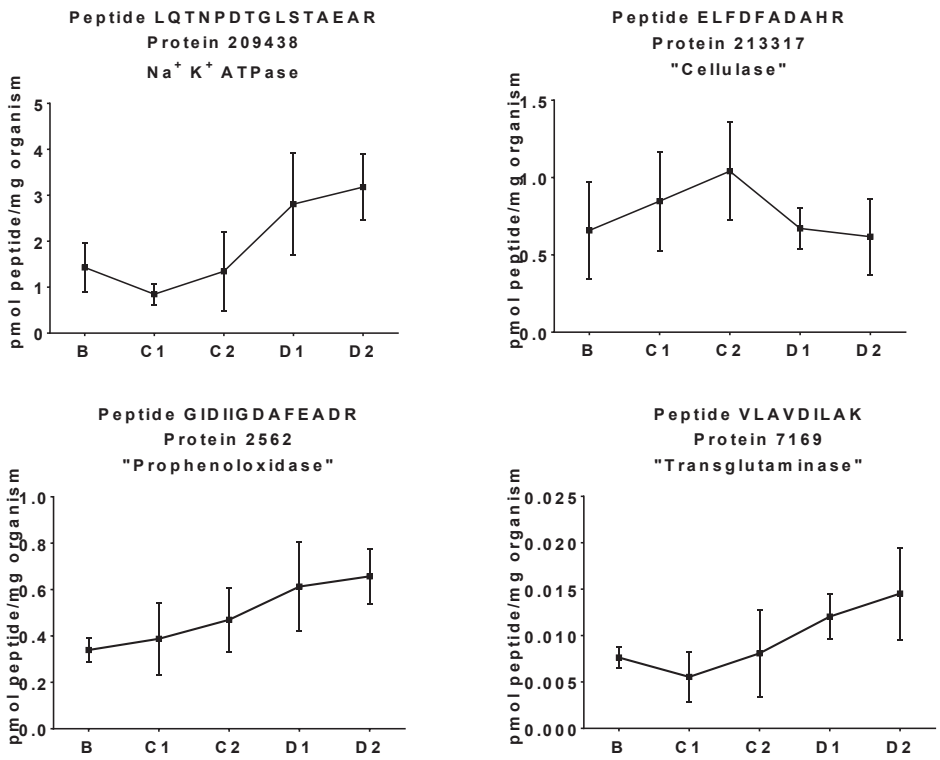


Figure V-4 – Molt-cycle-dependent modulation of proteins involved in several key functions in female gammarids.

The increase in concentration of Na^+K^+ ATPase proteins in the pre-molt phase is linked with the entry of water into the body and a consequent increase in volume at the time of ecdysis (Charmantier, 1998). Na^+K^+ pumps are therefore strongly modulated due to the great effort of the organism in order to maintain its osmotic equilibrium during this process. Cellulase presented an increase in the inter-molt stage, followed by a decrease in pre-molt, which is in accord with energy accumulation after ecdysis and the formation of the amplexus in the inter-molt. When coupled, females have a much restrictive access to food. Prophenoloxidase and transglutaminase, although normally associated with immune functions, were shown to have important modulations in the molt cycle of other crustaceans (Liu et al., 2006; Liu et al., 2011; Yeh et al., 2009).

This experiment allowed evaluating the use of selected peptides/proteins as reporters of major physiological functions presumed from their bioinformatic annotations previously defined in the GFOSS database. These automatic functional predictions were previously done through an automatic pipeline that is based on cross-species sequence homology, which is challenging for non-model species distantly related from well-characterized genomes. In this context, the molecular responses were interpreted by using the expertise of our research team in *Gammarus* physiology, thus demonstrating the interest of following the modulations of these proteins in an ecotoxicological context.

Evaluation of the relevance of protein candidates as biomarkers of toxicity: a laboratory case study with metals

A twenty-one day exposure to environmentally realistic concentrations of Cd and Pb (2 and 10 µg/L respectively) led to the modulation of several proteins linked with reproduction and general detoxification processes. In females, the vitellogenesis/oogenesis process was blocked in some organisms by both contaminants, as illustrated in Figure V-5. This was concluded based on the inhibition of the several yolk proteins in some exposed females (Cd F6, and Pb F1, F3, and F14).

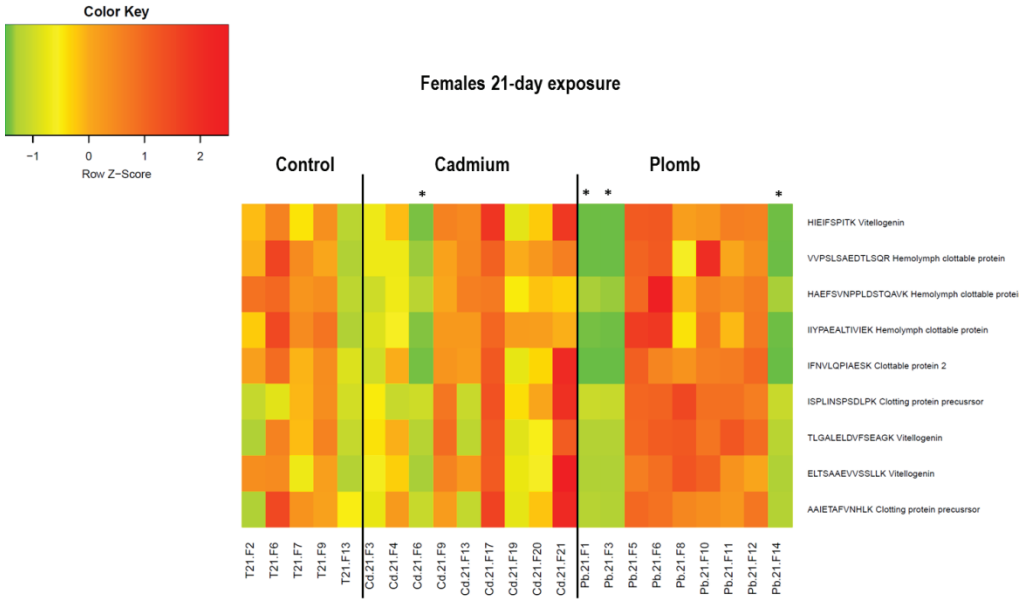


Figure V-5 – Heatmap illustrating the yolk protein levels in 21-day exposed females. Asterisks represent females with significant lower levels of yolk proteins.

Male organisms presented several significant protein modulations, illustrated in Figure V-6. Modulated proteins were associated with defense mechanisms such as detoxification (GST) and antioxidative processes (CAT), molt-related hormonal metabolism (JHE-like carboxylesterase and PPO), and immune responses (transglutaminase). Cadmium-induced molt inhibitions and vitellogenesis impairments have been previously observed in other studies and invertebrates (Geffard et al., 2010; Luo et al., 2015; Yang et al., 2015), and represent a common outcome following heavy-metal exposure in these species. GST and CAT are commonly used as biomarkers of detoxification and oxidative stress for a wide range of contaminants. The inhibitions observed herein are probably related to an incapability of the organism to detoxify, or to possible adaptation mechanisms implemented by the organism to cope with presence of metals in its environment. Mechanisms involved in heavy metal response are related to intracellular sequestration, repairing mechanisms, anti-oxidative enzymes, and modification of the uptake/elimination rates involved in metal bioaccumulation. The implementation of these processes can lead to energetic reallocations favoring the defense mechanisms and survival of the individual, but hindering other biological processes such as reproduction, energy acquisition, and growth.

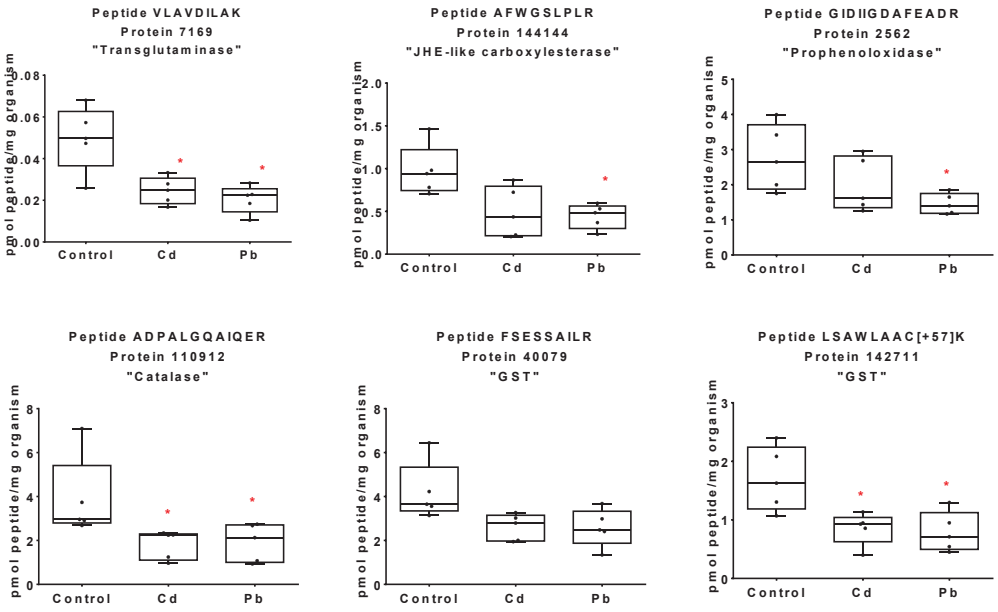


Figure V-6 – Down-regulated proteins in 21-day exposed male organisms. Red asterisks indicate statistical significant differences between the contaminated and control condition (non-parametric Mann-Whitney t-tests, $p<0.05$).

Towards a new generation multibiomarker assay for biomonitoring

Following the proposition of the new multibiomarker absolute quantification methodology, we evaluated the robustness of the approach in the context of a caging deployment operated through a river monitoring program from the RMC French water agency. Male and female gammarids were caged in several field sites presenting a very low (references) or high risk (contaminated) of degrading water quality due to the presence of chemical contaminants.

The comparative analysis of biomarker responses between control and contaminated field sites allowed for obtaining several protein modulations linked to contaminant effects. As shown in Figure V-7, contaminated sites exerted responses of Vtg-like proteins in both sexes. In females, Vtg-like concentration decreases are probably related to an inhibition in vitellogenesis, similarly to the previous laboratory experiment with Cd and Pb. In males, Vtg-like concentration increases in the order of 5-18 fold change were observed in some contaminated sites. These inductions were individual- and site-specific, in the same molds as previous studies that tried to study Vtg induction in male gammarids (Jubeaux et al., 2012b; Xuereb et al., 2011).

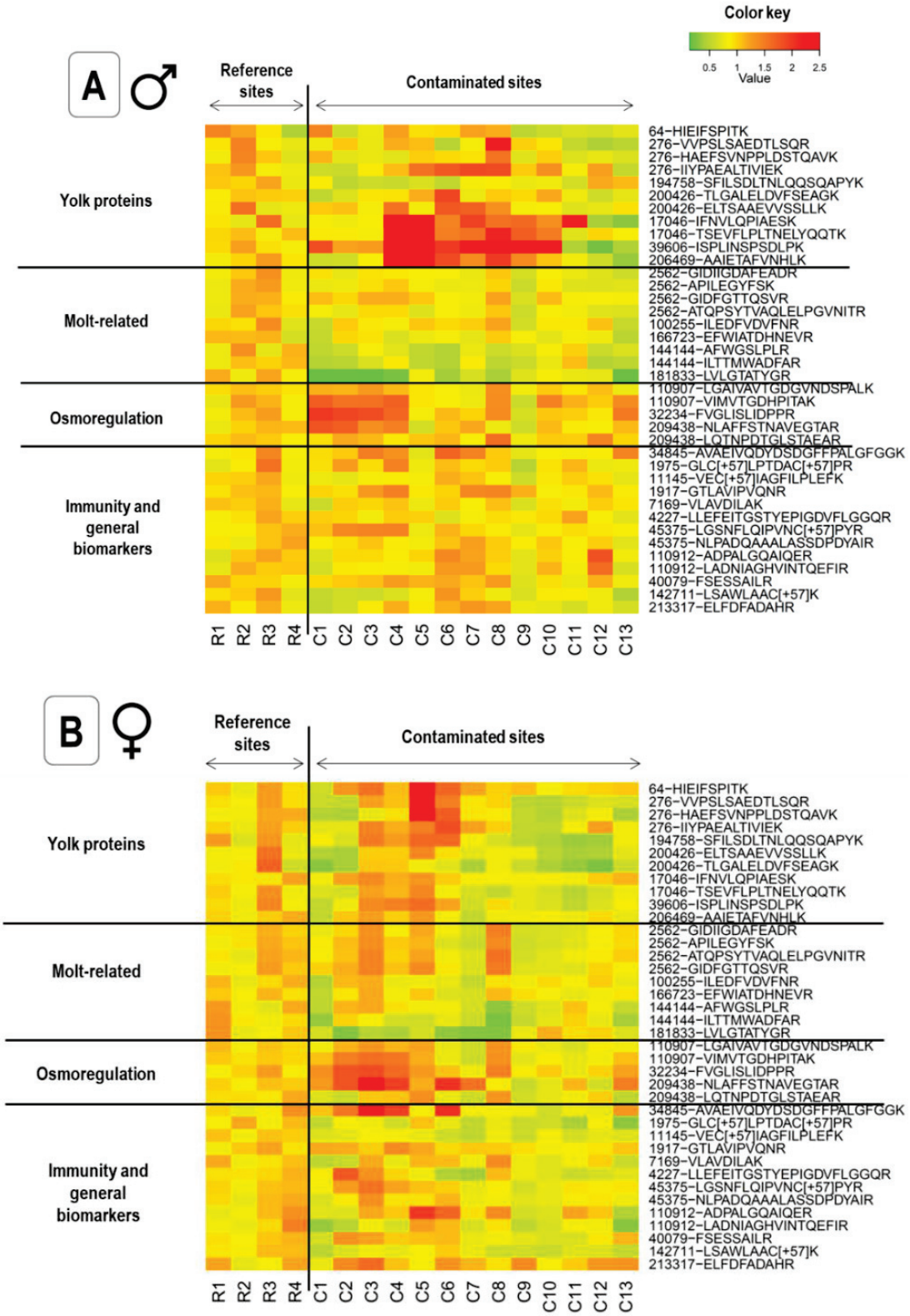


Figure V-7 – Heatmap illustrating the global biomarker responses in male (A) and female (B) organisms. Field sites are represented in columns, and individual biomarker peptides in lines. The color gradient ranges from green (lowest concentrations) to red (highest concentrations).

The results advocate for further studies to evaluate the pertinence of Vtg-like proteins as biomarkers of toxicity in gammarids, regarding their unusual diversity observed in *G. fossarum* (Trapp et al., 2016). Despite their primary role in oocyte development, some of the gammarid vitellogenins can have roles in other processes such as immunity, as already suggested in the literature for other invertebrates (Shi et al., 2006; Zhang et al., 2005).

Other than Vtg-like modulations, the exposure in contaminated sites also provoked inhibitions in two molt-related enzymes: chitinase (protein ID 181833) and JHE-like carboxylesterase (protein ID 144144). Strong decreases in the concentration of these proteins were observed in several sites. Chitinase is regulated by ecdysone, plays a key role in crustacean molting, and represents a sensitive biomarker to different kinds of contaminant exposure. JHE carboxylesterases participate in the inactivation of insect juvenile hormone (and presumably MF in crustaceans). Among the several biomarkers of contamination developed in *Gammarus fossarum*, the female molt cycle has been found to be very sensitive to chemical contamination (Geffard et al., 2010).

Hence, we could hypothesize that these inhibitions were the consequence of the action of molt-interfering contaminants interfering at the hormonal level. Such clear inhibitions also demonstrate the high sensitivity of these two biomarkers to contamination. JHE-like carboxylesterase was also found down-regulated after heavy metal exposure in the laboratory. Osmoregulatory Na^+K^+ ATPases, detoxification GST (protein 142711), and two immune-related proteins (proteins 1975 and 11145) were also modulated in contaminated sites. Decreases in GST and immune-related protein levels reflected the general stress of the organism that probably stopped counteracting the contamination effects after twenty-one days of exposure. Na^+K^+ ATPase pumps are often used as indicators of exposure to contamination (Issartel et al., 2010; Li et al., 2010; Pałecz et al., 2005).

The inductions observed in this experiment had similar fold changes to the variations obtained in the previous functional study, in which the response of these peptides was evaluated during the female molt cycle. In that study the increase of the levels of these proteins was related to the physiological events occurring during the molting process, thus reinforcing the physiological relevance of the modulations observed here. Nevertheless, further studies must be developed for identifying the consequences of

these modulations in the fitness of the organism. Again, using the extensive knowledge of gammarid physiology, some links can be established between molecular responses and individual outcomes. In previous studies the authors managed to establish clear links between digestive enzyme activities and reproductive impariments (Charron et al., 2015), or AChE levels and feeding rate / locomotion (Xuereb et al., 2009a) in *G. fossarum*. It would also be interesting to quantify and test the sensitivities to contamination of these proteins in different tissues. It is possible that modulations of osmotic pumps due to the effects of chemical disturbances will be more intense in the gills, or for JHE-like carboxylesterase in the cephalon.

2. Endocrine disruption biomarkers

2.1. Comparative shotgun proteomics for ED biomarker discovery

The high-throughput comparative proteomics experiment using the new Orbitrap mass analyzer from the Q Exactive HF led to the attainment of five times more spectra, 8 times more assigned spectra, and five times more peptides than its predecessor LTQ Orbitrap XL. This allowed for the identification of 4031 proteins. Moreover, the 14 most abundant proteins only accounted for 13% of the proteome in the QEx analysis, in contrast with the 29% in the LTQ. For the 20% less abundant proteome, the LTQ analysis identified 225 proteins, while the QEx identified 692 proteins. As expected, the QEx largely outperforms the LTQ platform for detecting proteins with lower abundances. It is known that low-abundant proteins (including taxon-specific proteins) carry great diagnostic potential, being major targets for biomarker research.

The combined quantification data from the QEx comparative proteome analysis between control and contaminated individuals led to the observation of a total of 53 statistically-validated modulated proteins. The two tested contaminant doses (0.5 and 50 μg/L) provided similar proteomic signatures, with a general decrease in metabolic and cytoskeleton proteins. Modulated proteins and their associated functional annotations are listed in Table V-2. Despite providing an interesting number of candidate biomarkers of pyriproxyfen exposure, few of them can be associated with an endocrine disruption effect. Nevertheless, with the high-throughput detection and relative quantification of thousands of proteins, it was possible to quickly assess the affected pathways by pyriproxyfen exposure. Through the analysis of modulated proteins, it was possible to draw some proteome signatures of pyriproxyfen-exposure in male gammarids that suggested disturbance in cell homeostasis and cytoskeleton formation. The fact that several proteins related to cytoskeleton formation and remodeling were strongly inhibited could be associated with molt-related perturbations. This is a common effect exerted by contaminants in gammarids, as previously observed in the SRM experiments, notably after the field exposures in sites with putative pesticide contamination.

Table V-2 – Statistically modulated proteins after PYR exposure.

				NCBInr first homolog protein		
	Fold Change	pValue	Evalue	Accession NCBInr	Functional annotation	Organism
Pyriproxyfen 50 µg/L	-7,33	0,00873	8,3E-202	XP_018015368.1	PREDICTED: methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial-like	*Hyalella azteca*
	-13,00	0,00735	1E-111	XP_018023847.1	PREDICTED: uncharacterized protein LOC108679669 isoform X2	*Hyalella azteca*
	-6,67	0,02266	3,5E-125	XP_018006603.1	PREDICTED: uncharacterized protein LOC108664518	*Hyalella azteca*
	-2,83	0,00340	1,4E-238	XP_015122631.1	PREDICTED: adenosylhomocysteinase 2-like isoform X1	*Diachasma alloeum*
	-3,25	0,00023	5,1E-273	XP_018024895.1	PREDICTED: papilin-like	*Hyalella azteca*
	-5,75	0,01831	3,4E-110	XP_018022683.1	PREDICTED: heat shock protein HSP 90-alpha-like	*Hyalella azteca*
	-18,00	0,00003	1,1E-89	XP_018008175.1	PREDICTED: 3-ketoacyl-CoA thiolase, mitochondrial-like	*Hyalella azteca*
	-7,00	0,00936	6,9E-115	XP_018020867.1	PREDICTED: transketolase-like isoform X1	*Hyalella azteca*
	-9,33	0,00883	4,5E-37	XP_018020867.1	PREDICTED: transketolase-like isoform X1	*Hyalella azteca*
	-8,00	0,00196	1,7E-196	XP_018027744.1	PREDICTED: sodium bicarbonate cotransporter 3-like	*Hyalella azteca*
	6,57	0,02445	4,2E-177	XP_018018795.1	PREDICTED: endoplasmin-like isoform X1	*Hyalella azteca*
	-8,57	0,00233	4,6E-38	XP_014260234.1	PREDICTED: glutathione S-transferase 1, isoform D-like	*Cimex lectularius*
	16,00	0,01230	2E-20	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*
	-6,75	0,03449	9,8E-174	XP_018026505.1	PREDICTED: succinate dehydrogenase	*ubiquinone*
	3,67	0,00790	6,7E-69	XP_013219096.1	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC101971546	*Ictidomys tridecemlineatus*
	21,00	0,03839	1,5E-146	XP_018019979.1	PREDICTED: alanine--glyoxylate aminotransferase 2, mitochondrial-like	*Hyalella azteca*
	-8,50	0,00158	1E-55	XP_018011744.1	PREDICTED: MICOS complex subunit Mic60-like isoform X2	*Hyalella azteca*
	-7,00	0,04909	1,4E-53	XP_018010288.1	PREDICTED: uncharacterized protein LOC108667737	*Hyalella azteca*
	5,33	0,02250	2,7E-28	XP_018019546.1	PREDICTED: uncharacterized protein LOC108676008	*Hyalella azteca*
	6,25	0,01942	1,6E-47	XP_018022389.1	PREDICTED: myosin heavy chain, muscle-like isoform X2	*Hyalella azteca*
	-7,00	0,02249	5,8E-161	XP_018016824.1	PREDICTED: dihydrolipoyl dehydrogenase, mitochondrial-like	*Hyalella azteca*
Pyriproxyfen 0.5 µg/L	-23,00	0,0006	4,5E-193	XP_018018893.1	PREDICTED: ubiquitin-like modifier-activating enzyme 1 isoform X2	*Hyalella azteca*
	-4,33	0,0174	6,4E-105	XP_018014531.1	PREDICTED: carboxypeptidase B-like	*Hyalella azteca*
	-5,33	0,0002	0	XP_018019307.1	PREDICTED: coatomer subunit beta-like	*Hyalella azteca*
	-6,50	0,0127	1E-111	XP_018023847.1	PREDICTED: uncharacterized protein LOC108679669 isoform X2	*Hyalella azteca*
	-6,67	0,0227	3,5E-125	XP_018006603.1	PREDICTED: uncharacterized protein LOC108664518	*Hyalella azteca*
	-3,74	0,0009	1,4E-238	XP_015122631.1	PREDICTED: adenosylhomocysteinase 2-like isoform X1	*Diachasma alloeum*
	-8,67	0,0058	2,6E-195	XP_018026866.1	PREDICTED: pyruvate carboxylase, mitochondrial-like	*Hyalella azteca*
	-17,50	0,0259	3E-116	XP_018022237.1	PREDICTED: PDZ and LIM domain protein Zasp-like isoform X6	*Hyalella azteca*
	-9,42	0,036	2,2E-20	XP_018022395.1	PREDICTED: myosin heavy chain, muscle-like isoform X7	*Hyalella azteca*
	-10,00	0,0023	2,2E-155	XP_018023846.1	PREDICTED: uncharacterized protein LOC108679669 isoform X1	*Hyalella azteca*
	-5,14	0,0068	1,7E-58	XP_018011652.1	PREDICTED: putative metallocarboxypeptidase ECM14	*Hyalella azteca*
	-2,70	0,0054	4,1E-71	XP_018009947.1	PREDICTED: 60S ribosomal protein L13a-like	*Hyalella azteca*
	-4,20	0,0207	2,8E-108	XP_018022055.1	PREDICTED: 97 kDa heat shock protein-like isoform X2	*Hyalella azteca*
	-4,80	0,0029	1,7E-196	XP_018027744.1	PREDICTED: sodium bicarbonate cotransporter 3-like	*Hyalella azteca*
	-2,14	0,0022	2E-98	XP_018010811.1	PREDICTED: glycogen phosphorylase-like	*Hyalella azteca*
	-3,73	0,0084	1E-12	AFN20597.1	lectin 1	*Macrobrachium rosenbergii*
	4,46	0,0105	3,4E-58	AAC78681.1	actin 1	*Penaeus monodon*
	29,00	0,0342	2E-20	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*
	-3,57	0,0041	7,1E-32	XP_018017725.1	PREDICTED: basement membrane-specific heparan sulfate proteoglycan core protein-like	*Hyalella azteca*
	-6,00	0,0133	1,9E-12	XP_003478750.1	PREDICTED: collagen alpha-2(V) chain	*Cavia porcellus*
	-6,67	0,0187	1,1E-55	XP_018012586.1	PREDICTED: sulfotransferase 1A1-like isoform X1	*Hyalella azteca*
	-2,38	0,0021	1,8E-84	XP_018017147.1	PREDICTED: leukocyte elastase inhibitor-like isoform X3	*Hyalella azteca*
	7,00	0,0067	7,5E-121	XP_018026505.1	PREDICTED: succinate dehydrogenase	*ubiquinone*
	5,33	0,0342	3E-111	XP_015917698.1	PREDICTED: glycogen phosphorylase-like	*Parasteatoda tepidariorum*
	35,00	0,0403	3,8E-124	XP_011347332.1	PREDICTED: glycogen phosphorylase	*Cerapachys biroi*
	3,10	0,0056	3,9E-74	AIM43585.1	L-lactate dehydrogenase	*Halocaridina rubra*
	-6,00	0,0231	9,1E-21	KHC67646.1	hypothetical protein MGE_01014	*Candida albicans P75010*
	-4,00	0,0128	2,6E-79	XP_018012529.1	PREDICTED: carboxypeptidase B-like	*Hyalella azteca*
	-5,60	0,0282	2,5E-52	XP_018019546.1	PREDICTED: uncharacterized protein LOC108676008	*Hyalella azteca*
	-4,75	0,0214	1,3E-202	XP_018023175.1	PREDICTED: LOW QUALITY PROTEIN: gelsolin, cytoplasmic-like	*Hyalella azteca*
	-20,00	0,0102	3,7E-180	XP_018023891.1	PREDICTED: myosin heavy chain, non-muscle-like isoform X4	*Hyalella azteca*
	-7,00	0,0267	8,2E-140	XP_018006437.1	PREDICTED: moesin/ezrin/radixin homolog 1-like isoform X4	*Hyalella azteca*

The proteins listed in Table V-2 were obtained after application of stringent statistical criteria: *p* value<0.05, FDR<0.05, fold-change>2, and seen at least in 4/5 samples. However, the study comprised a limited number of biological replicates (five per condition) that are not sufficient for a rigorous statistical power analysis regarding biomarker validation. By applying only fold-change criteria, we unraveled proteins whose functional annotations suggest a role in reproductive and hormonal processes. These proteins are listed on Table V-3.

Table V-3 – Physiological relevant proteins presenting strong fold changes in contaminated samples but needing further tests for validation.

				NCBInr first homolog protein		
	Fold Change	**pValue**	**evalue**	**Accession NCBInr**	**Functional annotation**	**Organism**
PYR 0.5 µg/L	29,00	0,03422	2E-20	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*
	16,00	**0,12322**	**4,3E-39**	**XP_018024698.1**	**PREDICTED: hemolymph clottable protein-like isoform X1**	***Hyalella azteca***
	11,00	**0,19586**	**1,9E-67**	**XP_018023978.1**	**PREDICTED: hemocyanin B chain-like**	***Hyalella azteca***
	10,00	0,04257	4,1E-55	XP_018023786.1	PREDICTED: glutathione S-transferase 1-like	*Hyalella azteca*
	-3,67	0,00361	8,3E-69	XP_018016475.1	PREDICTED: catalase-like	*Hyalella azteca*
	-7,00	0,07996	1,9E-28	AAX98287.1	hepatopancreas trypsin, partial	*Astacus leptodactylus*
	-7,50	0,07811	1,1E-44	XP_018016810.1	PREDICTED: endoglucanase A-like	*Hyalella azteca*
	-8,00	**0,05138**	**1,2E-51**	**XP_018028141.1**	**PREDICTED: prostaglandin reductase 1-like**	***Hyalella azteca***
	-11,00	**0,05754**	**1,5E-84**	**XP_018015212.1**	**PREDICTED: juvenile hormone epoxide hydrolase 1-like**	***Hyalella azteca***
	-13,50	**0,08262**	**1,6E-49**	**XP_018016813.1**	**PREDICTED: endoglucanase E-4-like**	***Hyalella azteca***
	-23,00	0,15785	5,7E-24	XP_014260234.1	PREDICTED: glutathione S-transferase 1, isoform D-like	*Cimex lectularius*
PYR 50 µg/L	**-18,00**	**0,08898**	**4,2E-39**	**XP_018015052.1**	**PREDICTED: cuticle protein AM1159-like**	***Hyalella azteca***
	-12,00	**0,08815**	**3,9E-136**	**XP_018024668.1**	**PREDICTED: endochitinase-like**	***Hyalella azteca***
	-12,00	0,16996	5,4E-38	XP_018027468.1	PREDICTED: glutathione S-transferase 2-like	*Hyalella azteca*
	-8,57	0,00233	4,6E-38	XP_014260234.1	PREDICTED: glutathione S-transferase 1, isoform D-like	*Cimex lectularius*
	8,00	**0,04153**	**8,8E-42**	**XP_018018288.1**	**PREDICTED: prostaglandin E synthase 3-like**	***Hyalella azteca***
	16	0,01230	2E-20	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*

From the proteins in this list, we can associate several functional annotations to possible endocrine disruption effects (proteins in bold). The hemolymph clottable protein-like isoform X1 is most likely a yolk protein, thus female specific. The fact that this protein is sixteen times up regulated can imply a possible induction of yolk proteins (Vtg-like) in male organisms, an ED effect, similar to the ones observed in the *in situ* SRM experiment. The protein annotated as juvenile hormone epoxide hydrolase 1-like is involved in the degradation of the juvenile hormone. This enzyme was found inhibited by low dose exposure to pyriproxyfen. Inhibitions of a similar protein involved in JH degradation were observed in the previous SRM experiments after metal exposure in the laboratory and multicontamination in field sites. The same observation can be made for protein annotated as “endochitinase”, inhibited here at a high dose of pyriproxyfen, and in the SRM experiments after field exposures. Hemocyanins, responsible for providing

oxygen during the molting process, are also known to have phenoloxidase activity and therefore to play an important in the molting process of crustaceans (Adachi et al., 2005; Glazer et al., 2013). Prostaglandin-related proteins are known for playing a role in endocrine regulation of crustacean reproduction, especially in oocyte maturation (Kornthong et al., 2014; Sagi et al., 1995; Spaziani et al., 1995; Sumpownon et al., 2015). Modulations of this protein in male organisms can also suggest possible ED effects.

In summary, we demonstrated the advantages of the use of new generation mass spectrometers in a context of ecotoxicological biomarker research. Due to its faster scan rate and higher resolution, the QEx improved greatly protein discovery and identification, as well as the extension of the candidate biomarker list previously reported with the LTQ (Trapp et al., 2015). The high-throughput analysis allowed establishing proteome signatures of pyriproxyfen exposure. By applying less stringent statistical criteria due to the limited number of biological technical replicates, this experiment highlighted several potential ED biomarker proteins that can be further studied and validated by targeted proteomics.

2.2. Candidate gene approach

As illustrated in Figure I-8, each gene origins several protein products, giving rise to very complex proteomes with thousands of different proteins that can still undergo post-translational modifications. Despite the advances in MS instrumentation and bioinformatics, large-scale proteome analysis only allow for the analysis of a subset of the proteome (especially for complex organisms). Protein samples are often dominated by a few high-abundance proteins, which overshadow low-abundant proteins. Low-abundant proteins are believed to carry great diagnostic potential (da Costa et al., 2017), but fall often below the limits of detection of the technique used for analysis. Moreover, even small proteomes present large dynamic ranges of protein concentrations that can span over 12 orders of magnitude (Righetti and Boschetti, 2013), rendering the analysis of the entire proteome even more challenging. In some human biofluids used for biomarker research, such as plasma or saliva, only 10 to 20 proteins constitute more than 90% of the total proteome (Figure V-8).

Low-abundant proteins may be present permanently or transiently, and generally comprise proteins associated with transcription and translation mechanisms, membrane receptors, allergens, polypeptides, and post-translationally modified proteins. All of these proteins, despite being present at very low concentrations, can exert physiological effects involved in key biological processes (Righetti and Boschetti, 2013).

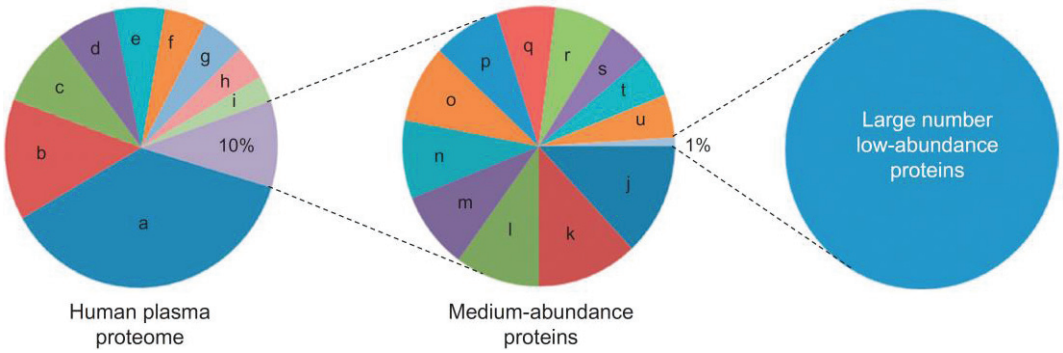


Figure V-8 - Schematic representation of the proportion of low-abundance proteins in human plasma. While a dozen proteins represent more than 90% of the protein content, less than 0.1% is accounted, comprising many proteins of low and very low-abundance. (Righetti and Boschetti, 2013)

Transcription factors, for example, play a pivotal role in gene expression through binding to specific DNA regions and regulating the transcription rate of genetic information (da Costa et al., 2017). Low-abundant molecules implicated in reproductive functions also diversified greatly throughout evolution. This originated distinct endocrine systems with species-specific proteins. In non-model species, these proteins are extremely difficult to assess and characterize in biological samples through homology-driven proteomics experiments.

There are several approaches available for detecting low-abundant proteins in proteomic analyses: depletion, enrichment, equalization, and fractionation techniques (Gianazza et al., 2016). These approaches, although very useful, are still time-consuming and limit the number of samples to be treated and compared. The careless application of these strategies can also lead to the possibility of inadvertently removing low-abundant proteins from the samples during the protocols, thus compromising the quantification of the target molecules. From an ecotoxicological point of view, important information can therefore be lost when comparing control and contaminated biological samples for

biomarker search. The identification and quantification of low-abundant proteins in biological samples remains one of the key technical challenges in proteome analysis and biomarker discovery.

Identification of candidate genes

EDCs may act upon hormonal signaling through binding to the hormonal receptors and acting as agonist or antagonist, or by provoking anti-hormonal effects through interference with the synthesis of the hormone (LeBlanc, 2007; Rodríguez et al., 2007). Hence, for this part of the thesis, we proposed developing an alternative sequence homology strategy for identifying key molecules involved in the hormonal system of *G. fossarum*. Based on the assumption that crustacean and insect endocrine regulator genes could be similar, we have mined the literature in order to identify key players involved in endocrine regulation. We therefore developed the workflow represented in Figure V-9 for discovering and validating the functions of the candidate sequences obtained from the research.

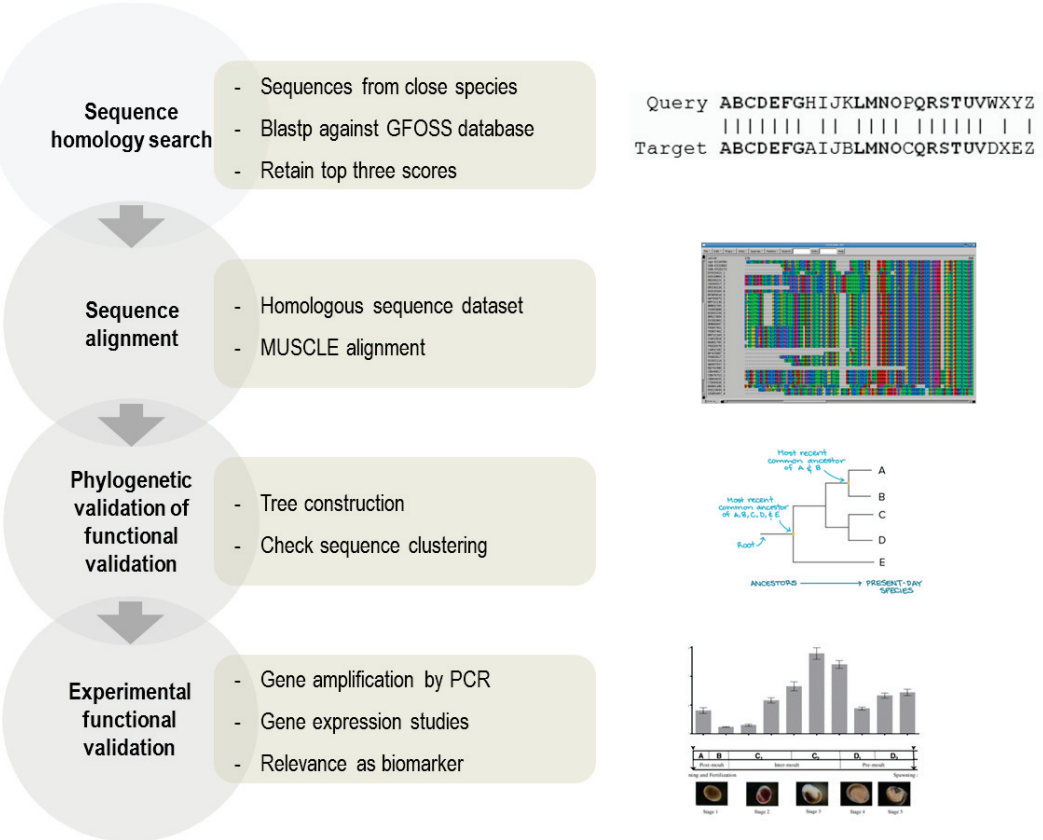


Figure V-9 – Workflow for the discovery and functional validation of *in silico* deduced biomarker candidates.

A candidate list was built after literature mining, and the presence or absence of sequence information in the reference transcriptome and protein catalog was verified through blast searches. This workflow eliminated the problematic background noise derived from high abundant proteins in the analysis, and the high specificity of the PCR allowed us to perform reliable quantifications of these low-abundant molecules at the transcript level.

Molecular phylogeny for validating bioinformatic functional annotations

Seen the importance of ecdysone pathway in arthropod molting and reproduction, we focused on three ecdysone-responsive genes for the phylogenetic and experimental validations (RXR, E75, and BR). Phylogenetic analysis allowed verifying the bioinformatic functional annotations of the candidate gene sequences. The putative RXR and BR sequences clustered with their orthologs from other species and were thus validated. However, the putative EcR sequences obtained from the blast search were found to be from another nuclear receptor, E75, a primary target of the RXR/EcR heterodimer. Still, the E75 annotated sequence GFOSS_2900 was kept for subsequent studies, since this nuclear receptor is one of the primary target genes of the EcR/RxR heterodimer, and regulates further downstream transcriptional cascades for molting and reproductive processes (Song et al., 2017).

In general, this phylogenetic analysis highlighted the need to question sequence homology search results when working with non-model species. As highlighted earlier, the first annotations available on GFOSS were generated through an automatic bioinformatic pipeline based on reciprocal blast procedure against public sequence databases (Trapp et al., 2014b 198). Because of the lack of molecular information available for these species, functional annotations predicted from sequence homology of phylogenetically distant species can sometimes be erroneous. Therefore, it is common to observe functional annotations from other paralog genes assigned to the gene of the species of interest. As an example, Table V-4 shows that among the results from a blast search with query sequences from RXR and EcR against the GFOSS database, we find the same homologous sequences for both genes, highlighted in green (GFOSS_83504 and GFOSS_2900).

Table V-4 – Sequence similarity search results from blastp of RXR and EcR query sequences against GFOSS.

Query sequence (Accesion nº - name - [species])	Top 3 blast results	Score	Identitites	Percentage	Expect	GFOSS annotation
AHF65151.1 retinoid X receptor [Eriocheir sinensis]	seq307974_fr2	228	113/198	57	3,0E-59	-
	GFOSS_83504_fr5	159	71/90	78	2,0E-38	ultraspiracle
	GFOSS_2900_fr4	114	81/307	26	1,0E-24	Nuclear hormone receptor E75
XP_018017941.1 PREDICTED: ecdysone receptor-like isoform X1 [Hyalella azteca]	GFOSS_83504_fr5	94	39/86	45	1,0E-18	ultraspiracle
	GFOSS_2900_fr4	70	95/389	24	1,0E-11	Nuclear hormone receptor E75
	seq406923_fr4	70	30/47	63	2,0E-11	-

This is due to the high sequence conservation among these paralog genes, as demonstrated in Figure V-10. The alignment presents a highly conserved part of the DBD for paralog and ortholog genes of RXR.

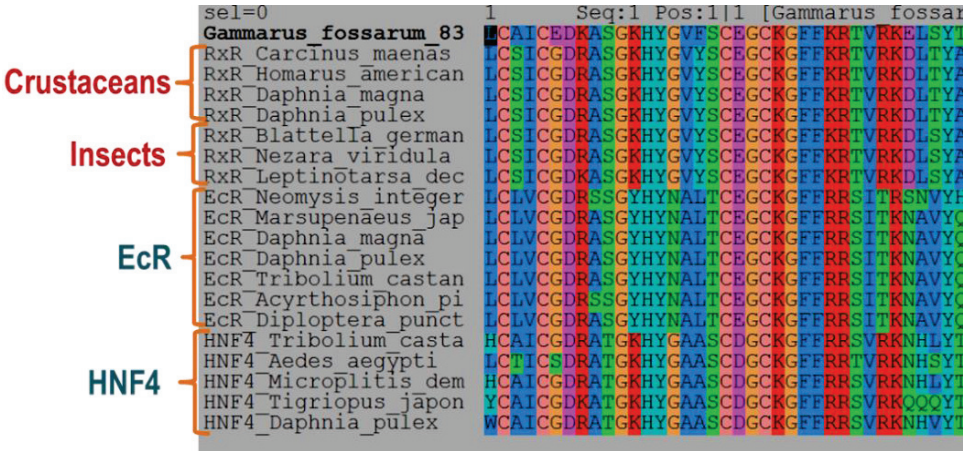


Figure V-10 – Protein sequence alignment in SeaView, using ortholog (same gene for different arthropod species) and paralog sequences (different members of the multigenic family of the target gene) of the candidate sequence Gammarus_fossarum_83504_fr5 (potential RxR).

The data presented in Figure V-11 also supports this fact, due to a strange clustering of the different crustacean sequences from the BTB-Zf family. The fact that lola genes from *Caligus clemensi* and *Lepeophteirus salmonis* clustered with the crustacean BR branch, and the lola from *Hyalella azteca* did not clustered with the other crustacean lola sequences, demonstrates the lack of knowledge concerning this specific gene family in crustacean species. It is possible that these crustacean genes were annotated (erroneously) only via sequence homology without any further verification. Taking again the example from the Vtg-like proteins in *G. fossarum*, their automatic functional annotations in the GFOSS database were mostly from decapod clotting

proteins. Based on these annotations and without further analysis (Trapp et al., 2016), these proteins would have been discarded as being reproduction-related.

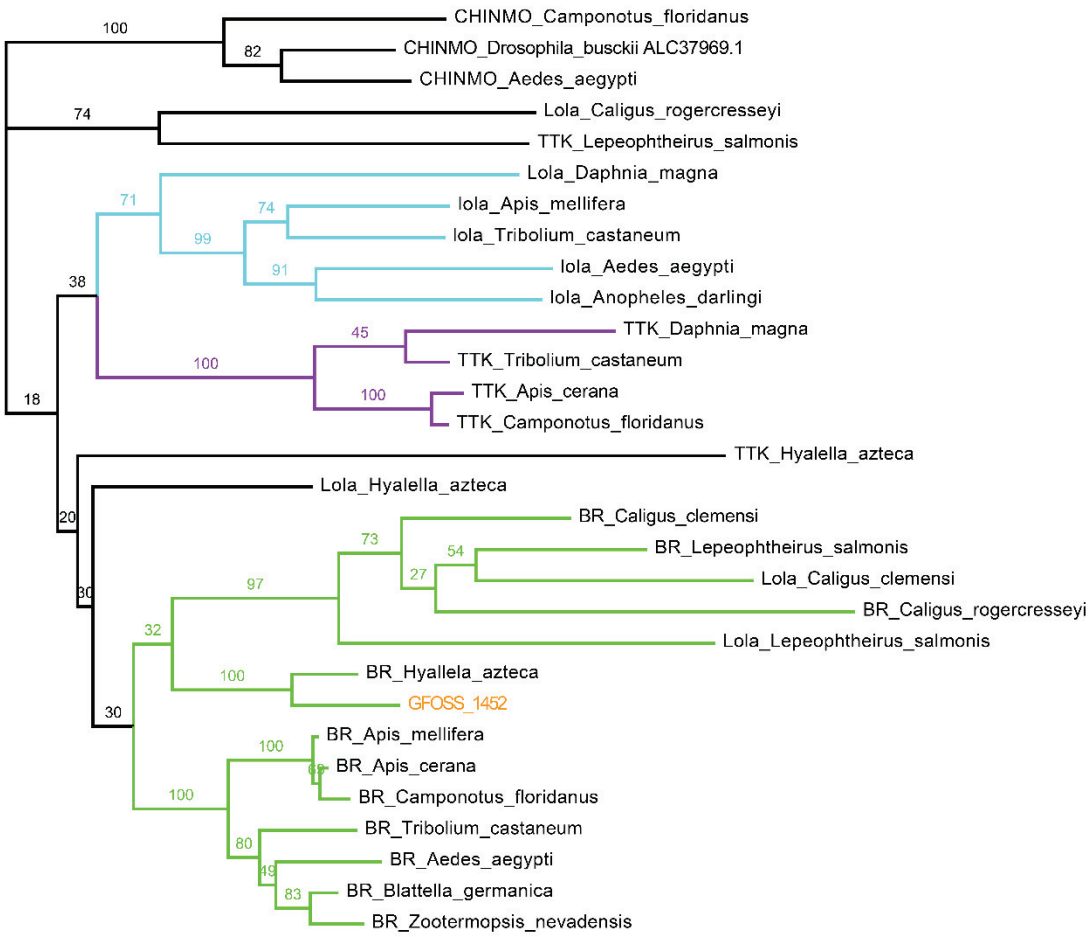


Figure V-11 – Phylogenetic tree of the homologous sequence dataset constructed for the functional validation of the GFOSS BR sequence 1452.

Expression of three ecdysone-responsive genes during female physiological processes

After the *in silico* validation of the RNAseq-derived gene sequences, specific sequences for each gene (RXR, E75, BR) were obtained by PCR amplification. This allowed having nucleotide sequences without assembly errors, and functional primers for the subsequent gene expression studies. Gene expression studies were performed only in female organisms, which have a well-characterized reproductive/molt cycle and thus were the obvious choice for studying the role of the three candidate genes in these processes. Results from gene expression throughout an entire reproductive cycle provided important insights into the role of each candidate in molt/reproduction of

female gammarids. As shown in Figure V-12, both E75 and RXR presented important modulations along the cycle, while BR remained relatively stable.

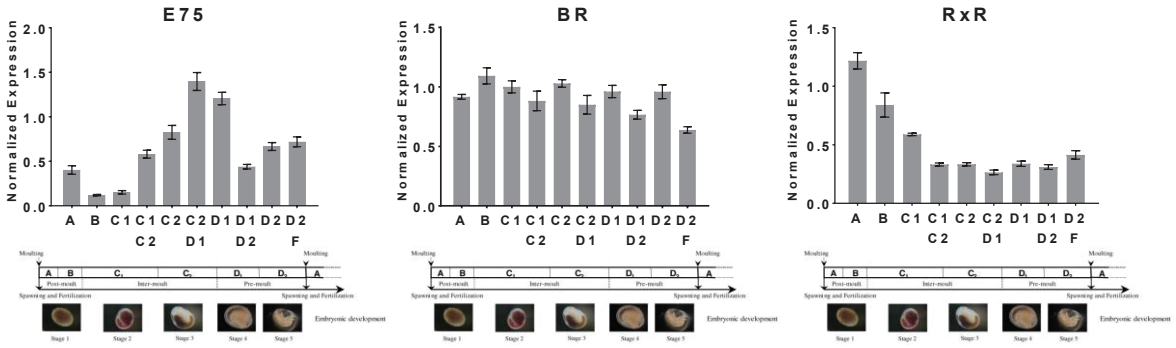


Figure V-12 – Expression profiles of E75, BR, and RXR during the female reproductive cycle. Expression values of target genes were normalized to the expression of the reference gene EF. Each bar corresponds to the mean ± SD obtained from the three experimental replicates.

The post-molt peak of RXR, just after exuviation, suggests a role of this gene in ecdysis. Similar conclusions were withdrawn from RXR expression in decapod species (Priya et al., 2009; Tang et al., 2014). E75 has two expression peaks that highlight its primary role in the onset of secondary vitellogenesis and ecdysis. The expression peak at the final premolt stage D2 that stretches until the postmolt stage A could point out a potential interaction/relationship between RXR and E75. Priya and coworkers (Priya et al., 2009) showed that RXR might be involved in the downstream regulation of E75 and chitinase gene transcription, thus influencing ecdysis. It is also known that E75 can act upon nuclear receptors HR3, HR4, and E75 isoforms for molt regulation (Hiruma and Riddiford, 2009; Horner et al., 1995; Mané-Padrós et al., 2008).

Surprisingly, BR did not present major variations in expression along the cycle, despite being known to play a major role in insect reproductive processes (mainly in metamorphosis) (Karim et al., 1993; Piulachs et al., 2010; Yang et al., 2014; Zhang and Zheng, 2017). As highlighted before, it is possible that the fact of using whole-body gammarids for the analysis diluted possible modulations of this gene. Thus it would be interesting to check in what tissue this gene is more highly expressed and to study its modulations in that tissue (cephalon or female gonads for example). It is also possible

that this gene could have gained different functions along evolution, in this particular group of arthropods, thus not participating in the female reproductive processes. Nevertheless, this stable expression of BR along the cycle demonstrates the reliability of the measurement and reinforces the results obtained for the other two genes.

Evaluation of RXR, E75, and BR as biomarkers of endocrine disruption in female gammarids

The comparative gene expression study between control and exposed-organisms highlighted the strong susceptibility of RXR and E75 to the EDCs used (Figure V-13). Concerning BR, no significant modulations were observed between control and contaminated conditions. Once more, this leads to question the role of this gene in endocrine-related female processes and to the need of performing tissue-specific studies. E75 was the more responsive to contamination, presenting a profile consistent with an endocrine disruption event. The inductions observed in the first days of exposure indicated that the three contaminants provoked an over-expression of the "early" ecdysone genes. The strong inhibition observed at day-14 suggests that tebufenozide disrupts the ecdysone cascade in crustacean gammarids. Several studies have shown the susceptibility of the ecdysone pathway to different types of endocrine disruptors in different arthropod species. Modulations of RXR were observed in TBT-exposed decapods *Crangon crangon* (Verhaegen et al., 2011), and in embryos from *Daphnia magna* females exposed to PYR (Wang et al., 2007). EcR was found induced in *Chironomus riparius* after exposure to a wide range of contaminants: BBP (Herrero et al., 2015), triclosan (Martínez-Paz et al., 2017), TBT (Morales et al., 2013), PCP (Morales et al., 2014), bisphenol A (Planelló et al., 2008). Inductions were also observed after bisphenol A and TEB injection in *S. nonagrioides* (Kontogiannatos et al., 2014), and after 30h exposure to fipronil in *A. tenuiremis* (Gaertner et al., 2012).

The surprising RXR induction at day-14 in pyriproxyfen-exposed females supports the theory of possible interaction of this gene with MF, since this contaminant is an MF analogue. In decapod crustaceans, a possible ecdysone-independent regulatory pathway for female reproduction control was suggested, through interactions of RXR with MF (Gong et al., 2016; Nagaraju et al., 2011; Tiu et al., 2012).

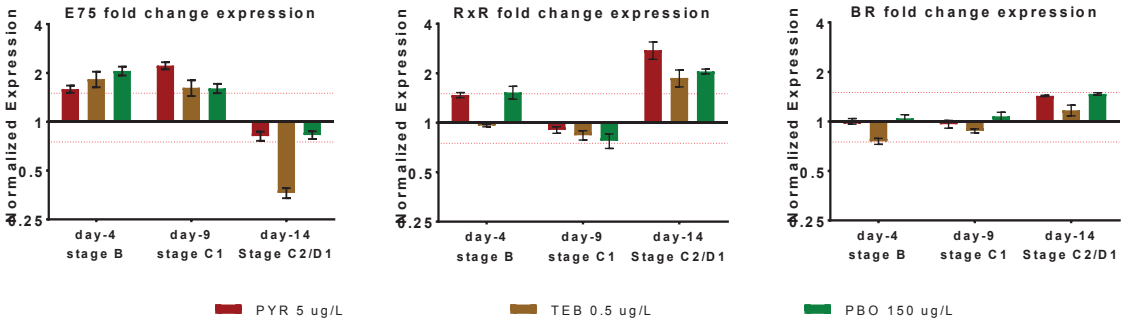


Figure V-13 – Fold change gene expressions of E75, RXR, and BR regarding the control condition. Red lines represent the 1.5 fold change threshold.

The ensemble of results demonstrated the important roles of RXR and E75 in endocrine related processes, but also call for further studies in order to better understand the regulation mechanisms of the ecdysone response cascade, and to elucidate if RXR and E75 have indeed some kind of reciprocal regulation that allows the organism to perform the ecdysis process. Analyzing the expression of these three genes in different tissues, such as cephalon and/or gonads, and different types of pollutants, could elucidate some of these points, notably through the tissue-specific inductions obtained. It is possible that stronger modulations will be observed in specific tissues, validating and reinforcing the use of the expression of these genes as indicators of endocrine disruption events.

Moreover, we are closer to the acquisition of other *G. fossarum* specific sequences of key molecules involved in hormonal processes, such as EcR. New studies are underway that aim at deciphering the inter-species variability of the *Gammaridea* suborder, through high-throughput transcriptome analyses. These new in-depth transcriptomes, allied with the high sensitivity of the QExactive HF mass spectrometer, will facilitate gene/protein discovery and allow more mechanistic studies of hormonal processes in gammarids. Of note, by using a draft transcriptomic database containing information sequence from several amphipods to interpret MS/MS spectra (as a test), we already managed to identify some EcR-annotated peptide sequences.

The use of non-model sentinel organisms for biomonitoring defies the application of classical biomarkers previously developed and validated for model species, especially vertebrates. The deficiency of molecular biology data and analytical tools for non-model species hindered the development of specific biomarkers for understanding the mode of action of contaminants and diagnosing their impact. In order to address these limitations, our research laboratory proposed, in recent projects, to develop new generation of biomarkers in *Gammarus fossarum* using high-throughput technologies such as transcriptomics and proteomics. Following the proteogenomics studies performed in *G. fossarum*, which allowed the creation of a large transcriptomic and proteomic species-specific database, this thesis project aimed at identifying and validating new biomarkers of toxicity for use in future biomonitoring programs(Garcia-Reyero et al., 2004).

The innovative MS-based multibiomarker approach developed in the first part of the thesis was the first methodology proposed for simultaneous quantification of several species-specific biomarkers in invertebrate ecotoxicological assessments. The assay allows for the detection of 20-40 biomarker proteins simultaneously, depending on the mass spectrometer used. Biomarkers were proved to have important roles in the normal physiological processes of the organisms, and to be sensitive to contamination. The results from the application of the assay in field studies were very satisfactory. Through its combination with the caging strategy developed for *G. fossarum*, variability sources related to biotic confounding factors were reduced, and the realism of exposure conditions were improved. Despite presenting some normal inter-individual variability, clear contamination-related biomarker responses were observed, and the high-throughput capabilities allowed recording these responses in a short period. After exposure, one week of sample preparation and analytical runs was enough for obtaining the absolute levels of 25 biomarkers in 170 organisms that were exposed in 17 different field sites, totaling as much as 6460 concentration values. From a biomonitoring point of view, this comprises a major breakthrough since it allows to rapidly determining the health status of organisms after exposure in a specific environment. Moreover, the

modulations obtained in molt-related proteins suggested precocious molecular responses - molt disturbances - that were not assessed through the physiological tests performed in the same organisms. Among the other responses obtained, we also observed Vtg-like inductions in male organisms that could indicate some feminization events taking place due to the presence of hormonal-disturbing molecules. These inductions can also be related to different functions acquired by these proteins in male organisms. The fact that nowadays we possess specific sequences for the eight different yolk proteins of *Gammarus fossarum*, allied to the modulations obtained herein, allows re-opening the old question of whether Vtg proteins can be used as ED biomarkers in amphipod crustaceans.

These works proved that the multibiomarker SRM approach is a promising strategy to be applied as a functional monitoring tool for regulatory purposes. However, there are still many considerations to take into account before this application can take place. Notably, further tests must be made for definite biomarker interpretation in terms of contaminant-induced modulations in biomonitoring. This comprises measuring the biomarkers in a large set of samples (in the order of the hundreds), and establishing reference values for each biomarker. These reference values must account for the effects of abiotic factors on biomarker responses, and for their natural variability, in order to propose threshold values. Further studies must also be performed in order to characterize each biomarker, determine their possible isoforms, and understand the mechanistic basis of their roles in the physiology of the organism. The diversity of vitellogenins, for example, calls for a deeper evaluation of their turnovers due to their cleavage process and suspected multiple functions. A similar case can be made for the ionic pumps Na^+K^+ ATPases, known to have several isoforms.

From a technical point of view, the continuous improvements in new mass spectrometers are allowing researchers to reach lower detection limits and higher dynamic ranges. This will allow easily detecting low-abundant proteins and augmenting the multiplexing capabilities of analysis. We have already stated that the use of more recent mass spectrometers leads to a significant increase in the proteins analyzed. Moreover, a new project is underway in our research teams that plan to use a new MS-targeted approach that will enhance the screening capability of the SRM approach. The approach, called SCOUT-MRM (Rougemont et al., 2017), uses spiked scout peptides that

will trigger complex transition lists regardless of the retention time of the targeted peptides. This eliminates the scheduling procedure around each peptides retention time, increasing greatly the multiplexing capability of the analysis. Preliminary *in silico* predictions point to the possibility of following simultaneously around 1000 peptides reporting for 250 proteins specific of *G. fossarum*. Considering the biomarker validation workflow presented in Figure I-15, SCOUT-MRM would be ideal as an intermediate step between biomarker discovery by shotgun proteomics and targeted absolute quantification assay (Figure i). This way, the elimination of false positives could be done with a much higher initial peptide/protein list, originating higher number of reliable surrogate peptide biomarkers for developing the absolute quantification assay.

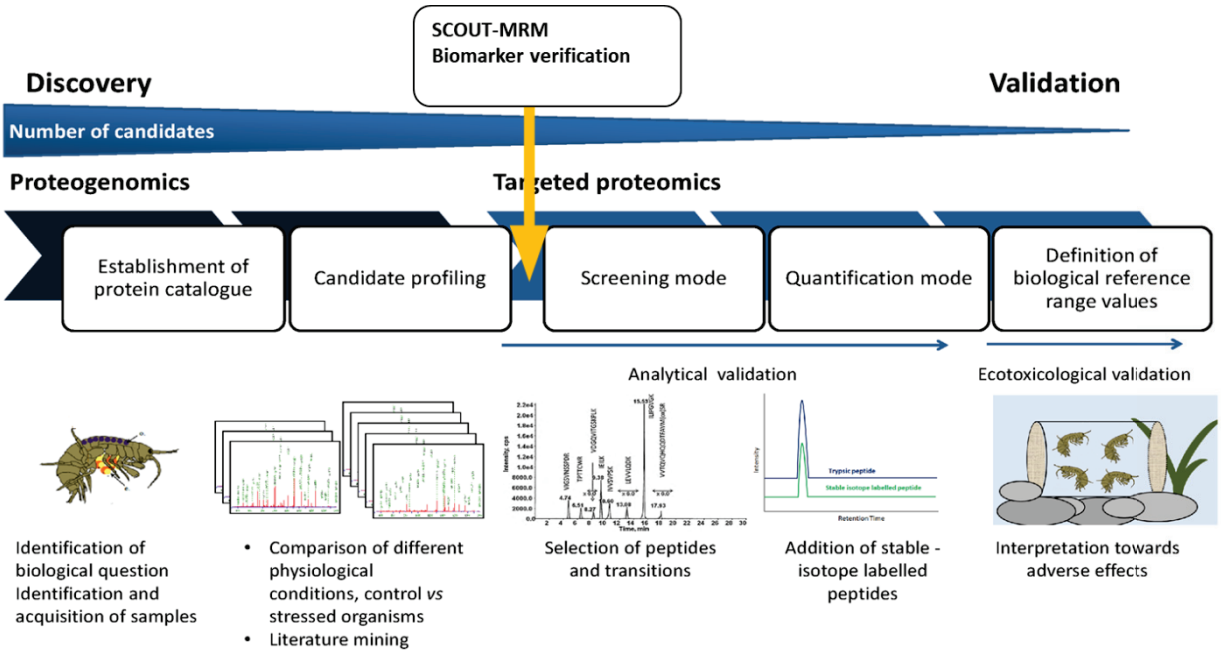


Figure i – Integration of the new SCOUT-MRM approach in the environmental biomarker pipeline proposed previously by (Trapp et al., 2014a).

The high-throughput capabilities of this methodology can also provide information on protein conservation among species. Similar to the works performed by Jubeaux et al. (Jubeaux et al., 2012a) with vitellogenins, one can study the inter-species transferability of the biomarker peptide sequences. Through the determination of conserved peptides among different gammarid species, for example, "universal" biomarkers can be developed and applied to the *Gammaridae* family. If successful, this

biomarker transferability will allow establishing our reference population of *G. fossarum* as a surrogate species towards gammarids diversity, one-step closer to be used as a model species in ecotoxicology.

The perspectives offered by these new-generation multibiomarker methodologies can also be applied to the discovery of specific signatures of contaminant modes of action such as endocrine disruption. With relevant surrogate biomarkers of major physiological functions, this kind of high-throughput multibiomarker approaches can give information about the type of contaminants present in the exposure media through these signatures. In this context, the second axis of this thesis focused specifically on endocrine disruption, one of the major concerns in environmental risk assessment. The works performed herein allowed gaining important insights into endocrine regulation and disruption in *Gammarus fossarum.* The shotgun proteomics study performed in male gonads demonstrated the great evolution that is taking place in mass spectrometry technology, which is bringing researchers closer to the big objective of analyzing the entire proteome of an organism. With the 4031 proteins detected in a non-model species, this study allowed us to enlarge the protein catalog for our reference species *G. fossarum*, which previously contained 1873 proteins. The fact that we can simultaneously follow thousands of proteins also opens the door to possible future studies aiming at determining contaminant modes of action, affected pathways, as well as studying inter-population and inter-species proteome diversity that can explain, for example, adaptation mechanisms to contamination. The modulated male proteins observed after pyriproxyfen exposure also allowed unraveling potential biomarkers of reproductive dysfunctions effects. These biomarkers were proposed for further validation studies by targeted proteomics.

During this thesis, we also developed an alternative workflow developed for the discovery and detection of specific poorly expressed target molecules in the context of ED biomarker development. The implementation of this workflow allowed us to discover and functionally validate, for the first time in gammarids, key genes involved in the ecdysone pathway. RXR and E75 genes were proved to have an important role in female physiological processes such as vitellogenesis, oogenesis and ecdysis, and showed a high sensitivity to pesticide exposure. These works highlighted the importance

of using sentinel species whose physiology is extensively known in order to understand the function of biomarker candidates. Moreover, due to the lack of knowledge of endocrine systems in the majority of crustacean species, we also observed some possible erroneous functional annotations in public databases that can mislead gene annotations by blast homology. Phylogenetic and experimental functional assessments must be done to overcome this limitation in non-model species. In order to fully understand the molecular basis of gammarid hormonal processes, deeper mechanistic and functional studies must be yet performed. However, from a biomarker point of view, the molecules targeted herein proved to be contaminant-sensitive. Moreover, specific protein sequences can be deduced from the nucleotidic sequence, and can be introduced in the targeted proteomics multibiomarker approach for further validation and use as biomarkers of exposure to EDC. Since RXR and E75 are also known to have several isoforms, characterization of these molecules must be also performed in order to choose reliable reporter peptides.

Summarizing, these works allowed addressing two important limitations of biomarker use in ecotoxicology and especially biomonitoring programs. Species-specific general and ED biomarkers were discovered and validated, and incorporated into a new multibiomarker absolute quantification approach. However, the biomarkers developed and proposed herein still do not reflect ecologically meaningful effects. Nevertheless, as highlighted in Chapter I, biomarker-based environmental risk assessment can be used to develop AOPs to establish links between sub-individual biomarker responses and possible effects at the higher biological levels of organisation. AOPs can potentially predict adverse outcomes on the individual, population, community, and ecological levels according to biomarker responses, thus yielding a better understanding of the implication of a given molecular event in environmental risk assessments (Lee et al., 2015). Moreover, the incorporation of existing knowledge in the literature to support each event constitutes an alternative that minimizes resource-intensive testing approaches. As suggested by (Martyniuk and Simmons, 2016), multi-omics researches must be applied to develop detailed mechanism-based AOPs that are applicable for use in both effects monitoring and risk assessment framework. Large-scale omics approaches can allow identifying the initial molecular markers related to the AOPs, as

well as filling the gaps between the molecular markers and individual effects (Lee et al., 2015).

Finally, a *Gammarus fossarum* genome project is also being considered by our research team. Having a sequenced genome would greatly elucidate some of the questions that were posed during these works, notably in terms of protein identification by mass spectrometry, determination of protein isoforms and discovery of endocrine pathways present or absent in gammarids.

Abu-Asab, M., Chaouchi, M., Amri, H., 2008. Evolutionary medicine: A meaningful connection between omics, disease, and treatment. Proteomics - Clinical Applications 2, 122-134.

Adachi, K., Endo, H., Watanabe, T., Nishioka, T., Hirata, T., 2005. Hemocyanin in the exoskeleton of crustaceans: enzymatic properties and immunolocalization. Pigment Cell Res 18, 136-143.

Aggarwal, K., Choe, L.H., Lee, K.H., 2006. Shotgun proteomics using the iTRAQ isobaric tags. Brief Funct Genomic Proteomic 5, 112-120.

Ahn, Y.H., Lee, J.Y., Lee, J.Y., Kim, Y.S., Ko, J.H., Yoo, J.S., 2009. Quantitative analysis of an aberrant glycoform of TIMP1 from colon cancer serum by L-PHA-enrichment and SISCAPA with MRM mass spectrometry. J Proteome Res 8, 4216-4224.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.

America, A.H., Cordewener, J.H., 2008. Comparative LC-MS: a landscape of peaks and valleys. Proteomics 8, 731-749.

Amiard-Triquet, C., Berthet, B., 2015. Chapter 7 - Individual Biomarkers, Aquatic Ecotoxicology. Academic Press, pp. 153-182.

Amiard, J.C., Amiard-Triquet, C., Barka, S., Pellerin, J., Rainbow, P.S., 2006. Metallothioneins in aquatic invertebrates: Their role in metal detoxification and their use as biomarkers. Aquatic Toxicology 76, 160-202.

Anderson, N.L., Anderson, N.G., 1998. Proteome and proteomics: New technologies, new concepts, and new words. ELECTROPHORESIS 19, 1853-1861.

Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrrano, J.A., Tietge, J.E., Villeneuve, D.L., 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environmental Toxicology And Chemistry 29, 730-741.

Aranda, A., Pascual, A., 2001. Nuclear Hormone Receptors and Gene Expression. Physiological Reviews 81, 1269-1304.

Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., Hartmann, E.M., 2014. Non-model organisms, a species endangered by proteogenomics. J Proteomics 105, 5-18.

Bahamonde, P.A., Feswick, A., Isaacs, M.A., Munkittrick, K.R., Martyniuk, C.J., 2016. Defining the role of omics in assessing ecosystem health: Perspectives from the Canadian environmental monitoring program. Environ Toxicol Chem 35, 20-35.

Bahamonde, P.A., McMaster, M.E., Servos, M.R., Martyniuk, C.J., Munkittrick, K.R., 2015. Molecular pathways associated with the intersex condition in rainbow darter (Etheostoma caeruleum) following exposures to municipal wastewater in the Grand River basin, ON, Canada. Part B. Aquatic Toxicology 159, 302-316.

Barnard, J.L., Barnard, C.M., 1983. Freshwater Amphipoda of the World, I. Evolutionary patterns and II. Handbook and bibliography. Hayfield Associates, Mt. Vernon, Virginia.

Behrens, A., Segner, H., 2005. Cytochrome P4501A induction in brown trout exposed to small streams of an urbanised area: results of a five-year-study. Environmental Pollution 136, 231-242.

Beliaeff, B., Burgeot, T., 2002. Integrated biomarker response: A useful tool for ecological risk assessment. Environmental Toxicology And Chemistry 21, 1316-1322.

Benedicto, J., Martínez-Gómez, C., Campillo, J., 2005. Induction of metallothioneins in Mullus barbatus as specific biomarker of metal contamination: A field study in the western Mediterranean. Ciencias Marinas 31, 265-274.

Benninghoff, A., 2007. Toxicoproteomics--the next step in the evolution of environmental biomarkers. Toxicol Sci 95, 1-4.

Besse, J.-P., Coquery, M., Lopes, C., Chaumot, A., Budzinski, H., Labadie, P., Geffard, O., 2013. Caged Gammarus fossarum (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: Towards the determination of threshold values. Water Research 47, 650-660.

Biales, A.D., Bencic, D.C., Flick, R.L., Blocksom, K.A., Lazorchak, J.M., Lattier, D.L., 2011. Proteomic analysis of a model fish species exposed to individual pesticides and a binary mixture. Aquatic Toxicology 101, 196-206.

Birnbaum, L.S., 2013. State of the Science of Endocrine Disruptors. Environmental health perspectives 121, a107-a107.

Bocchetti, R., Lamberti, C.V., Pisanelli, B., Razzetti, E.M., Maggi, C., Catalano, B., Sesta, G., Martuccio, G., Gabellini, M., Regoli, F., 2008. Seasonal variations of exposure biomarkers, oxidative stress responses and cell damage in the clams, Tapes philippinarum, and mussels, Mytilus galloprovincialis, from Adriatic sea. Marine environmental research 66, 24-26.

Bondarenko, P.V., Chelius, D., Shaler, T.A., 2002. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. Analytical chemistry 74, 4741-4749.

Bonneton, F., Chaumot, A., Laudet, V., 2008. Annotation of Tribolium nuclear receptors reveals an increase in evolutionary rate of a network controlling the ecdysone cascade. Insect biochemistry and molecular biology 38, 416-429.

Borst, D.W., Laufer, H., Landau, M., Chang, E.S., Hertz, W.A., Baker, F.C., Schooley, D.A., 1987. Methyl farnesoate and its role in crustacean reproduction and development. Insect Biochemistry 17, 1123-1127.

Bouétard, A., Noirot, C., Besnard, A.-L., Bouchez, O., Choisne, D., Robe, E., Klopp, C., Lagadic, L., Coutellec, M.-A., 2012. Pyrosequencing-based transcriptomic resources in the pond snail Lymnaea stagnalis, with a focus on genes involved in molecular response to diquat-induced stress. Ecotoxicology 21, 2222-2234.

Breitholtz, M., 2013. Crustaceans, Endocrine Disrupters. John Wiley & Sons, Inc., pp. 116-142.

Cairns, J., 1992. The threshold problem in ecotoxicology. Ecotoxicology 1, 3-16.

Cajaraville, M.P., Bebianno, M.J., Blasco, J., Porte, C., Sarasquete, C., Viarengo, A., 2000. The use of biomarkers to assess the impact of pollution in coastal environments of the Iberian Peninsula: a practical approach. Science of the Total Environment 247, 295-311.

Campbell, R.K., Satoh, N., Degnan, B.M., 2004. Piecing together evolution of the vertebrate endocrine system. Trends Genet 20, 359-366.

Cappello, T., Maisano, M., D'Agata, A., Natalotto, A., Mauceri, A., Fasulo, S., 2013. Effects of environmental pollution in caged mussels (Mytilus galloprovincialis). Marine environmental research 91, 52-60.

Carnevali, O., Mosconi, G., Habibi, H.R., Elia, A.C., Cardinali, M., Polzonetti-Magni, A.M., 2003. Validation of an enzyme linked immunosorbent assay (Elisa) for Cyprinus carpio L. vitellogenin, as a biomarker of reproductive disorders. Chemistry and Ecology 19, 5-13.

Chalghmi, H., Zrafi, I., Gourves, P.-Y., Bourdineaud, J.-P., Saidane-Mosbahi, D., 2016. Combined effects of metal contamination and abiotic parameters on biomarker responses in clam Ruditapes decussatus gills: an integrated approach in biomonitoring of Tunis lagoon. Environmental Science: Processes & Impacts 18, 895-907.

Chan, S.-M., Gu, P.-L., Chu, K.-H., Tobe, S.S., 2003. Crustacean neuropeptide genes of the CHH/MIH/GIH family: implications from molecular studies. General and comparative endocrinology 134, 214-219.

Chang, E.S., 1993. Comparative endocrinology of molting and reproduction: insects and crustaceans. Annu Rev Entomol 38, 161-180.

Chapman, P.M., Wang, F., Caeiro, S.S., 2013. Assessing and managing sediment contamination in transitional waters. Environment International 55, 71-91.

Charmantier, G.U.Y., 1998. Ontogeny of osmoregulation in crustaceans: a review. Invertebrate Reproduction & Development 33, 177-190.

Charnot, A., Gouveia, D., Armengaud, J., Almunia, C., Chaumot, A., Lemoine, J., Geffard, O., Salvador, A., 2017. Multiplexed assay for protein quantitation in the invertebrate Gammarus fossarum by liquid chromatography coupled to tandem mass spectrometry. Anal Bioanal Chem 17, 017-0348.

Charron, L., Geffard, O., Chaumot, A., Coulaud, R., Jaffal, A., Gaillet, V., Dedourge-Geffard, O., Geffard, A., 2014. Influence of molting and starvation on digestive enzyme activities and energy storage in Gammarus fossarum. PLoS ONE 9.

Charron, L., Geffard, O., Chaumot, A., Coulaud, R., Jaffal, A., Gaillet, V., Dedourge-Geffard, O., Geffard, A., 2015. Consequences of Lower Food Intake on the Digestive Enzymes Activities, the Energy Reserves and the Reproductive Outcome in Gammarus fossarum. PLOS ONE 10, e0125154.

Charron, L., Geffard, O., Chaumot, A., Coulaud, R., Queau, H., Geffard, A., Dedourge-Geffard, O., 2013. Effect of water quality and confounding factors on digestive enzyme activities in Gammarus fossarum. Environmental Science and Pollution Research 20, 9044-9056.

Chaumot, A., Geffard, O., Armengaud, J., Maltby, L., 2015. Chapter 11 - Gammarids as Reference Species for Freshwater Monitoring A2 - Amiard-Triquet, Claude, in: Amiard, J.-C., Mouneyrac, C. (Eds.), Aquatic Ecotoxicology. Academic Press, pp. 253-280.

Chen, G., Pramanik, B.N., 2009. Application of LC/MS to proteomics studies: current status and future prospects. Drug Discovery Today 14, 465-471.

Chen, Y.T., Chen, H.W., Wu, C.F., Chu, L.J., Chiang, W.F., Wu, C.C., Yu, J.S., Tsai, C.H., Liang, K.H., Chang, Y.S., Wu, M., Yang, W.T.O., 2017. Development of a multiplexed liquid chromatography multiple-reaction-monitoring mass spectrometry (LC-MRM/MS) method for evaluation of salivary proteins as oral cancer biomarkers. Molecular and Cellular Proteomics 16, 799-811.

Chipman, A.D., Ferrier, D.E.K., Brena, C., Qu, J., Hughes, D.S.T., Schröder, R., Torres-Oliva, M., Znassi, N., Jiang, H., Almeida, F.C., Alonso, C.R., Apostolou, Z., Aqrawi, P., Arthur, W., Barna, J.C.J., Blankenburg, K.P., Brites, D., Capella-Gutiérrez, S., Coyle, M., Dearden, P.K., Du Pasquier, L., Duncan, E.J., Ebert, D., Eibner, C., Erikson, G., Evans, P.D., Extavour, C.G., Francisco, L., Gabaldón, T., Gillis, W.J., Goodwin-Horn, E.A., Green, J.E., Griffiths-Jones, S., Grimmelikhuijzen, C.J.P., Gubbala, S., Guigó, R., Han, Y., Hauser, F., Havlak, P., Hayden, L., Helbing, S., Holder, M., Hui, J.H.L., Hunn, J.P., Hunnekuhl, V.S., Jackson, L., Javaid, M., Jhangiani, S.N., Jiggins, F.M., Jones, T.E., Kaiser, T.S., Kalra, D., Kenny, N.J., Korchina, V., Kovar, C.L., Kraus, F.B., Lapraz, F., Lee, S.L., Lv, J., Mandapat, C., Manning, G., Mariotti, M., Mata, R., Mathew, T., Neumann, T., Newsham, I., Ngo, D.N., Ninova, M., Okwuonu, G., Ongeri, F., Palmer, W.J., Patil, S., Patraquim, P., Pham, C., Pu, L.-L., Putman, N.H., Rabouille, C., Ramos, O.M., Rhodes, A.C., Robertson, H.E., Robertson, H.M., Ronshaugen, M., Rozas, J., Saada, N., Sánchez-Gracia, A., Scherer, S.E., Schurko, A.M., Siggens, K.W., Simmons, D., Stief, A., Stolle, E., Telford, M.J., Tessmar-Raible, K., Thornton, R., van der Zee, M., von Haeseler, A., Williams, J.M., Willis, J.H., Wu, Y., Zou, X., Lawson, D., Muzny, D.M., Worley, K.C., Gibbs, R.A., Akam, M., Richards, S., 2014. The First Myriapod Genome Sequence Reveals Conservative Arthropod Gene Content and Genome Organisation in the Centipede Strigamia maritima. PLoS Biology 12, e1002005-e1002005.

Cho, C.K.J., Drabovich, A.P., Batruch, I., Diamandis, E.P., 2011. Verification of a biomarker discovery approach for detection of Down syndrome in amniotic fluid via multiplex selected reaction monitoring (SRM) assay. Journal of Proteomics 74, 2052-2059.

Ciliberti, A., Chaumot, A., Recoura-Massaquant, R., Chandesris, A., François, A., Coquery, M., Ferréol, M., Geffard, O., 2017. Caged Gammarus as biomonitors identifying thresholds of toxic metal bioavailability that affect gammarid densities at the French national scale. Water Research 118, 131-140.

Cobbett, C., Goldsbrough, P., 2002. Phytochelatins and metallothioneins: Roles in heavy metal detoxification and homeostasis, Annual Review of Plant Biology, pp. 159-182.

Collins, M., Tills, O., Spicer, J.I., Truebano, M., 2017. De novo transcriptome assembly of the amphipod Gammarus chevreuxi exposed to chronic hypoxia. Marine genomics 33, 17-19.

Comoglio, L., Amin, O., Roque, A., Betancourt-Lozano, M., Anguas, D., Haro, B.M., 2005. Evaluation of sublethal biomarkers in Litopenaeus vannamei on foodborne exposure to methyl parathion. Ecotoxicol Environ Saf 62, 66-74.

Connon, R.E., Geist, J., Werner, I., 2012. Effect-Based Tools for Monitoring and Predicting the Ecotoxicological Effects of Chemicals in the Aquatic Environment. Sensors (Basel, Switzerland) 12, 12741-12771.

Coulaud, R., Geffard, O., Coquillat, A., Quéau, H., Charles, S., Chaumot, A., 2014. Ecological Modeling for the Extrapolation of Ecotoxicological Effects Measured during in Situ Assays in Gammarus. Environmental science & technology 48, 6428-6436.

Coulaud, R., Geffard, O., Xuereb, B., Lacaze, E., Quéau, H., Garric, J., Charles, S., Chaumot, A., 2011. In situ feeding assay with Gammarus fossarum (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring. Water Research 45, 6417-6429.

Courtenay, S., Grunwald, C., Kreamer, G.-L., Alexander, R., Wirgin, I., 1993. Induction and clearance of cytochrome P4501A mRNA in Atlantic tomcod caged in bleached kraft mill effluent in the Miramichi River. Aquatic Toxicology 27, 225-244.

Covi, J.A., Chang, E.S., Mykles, D.L., 2012. Neuropeptide signaling mechanisms in crustacean and insect molting glands. Invertebrate Reproduction and Development 56, 33-49.

Cravedi, J.-P., Zalko, D., Savouret, J.-F., Menuet, A., Jégou, B., 2007. Le concept de perturbation endocrinienne et la santé humaine. Med Sci (Paris) 23, 198-204.

Crisp, T.M., Clegg, E.D., Cooper, R.L., Wood, W.P., Andersen, D.G., Baetcke, K.P., Hoffmann, J.L., Morrow, M.S., Rodier, D.J., Schaeffer, J.E., Touart, L.W., Zeeman, M.G., Patel, Y.M., 1998. Environmental endocrine disruption: An effects assessment and analysis. Public Health Services, US Dept of Health and Human Services, pp. 11-56.

da Costa, J.P., Santos, P.S.M., Vitorino, R., Rocha-Santos, T., Duarte, A.C., 2017. How low can you go? A current perspective on low-abundance proteomics. TrAC Trends in Analytical Chemistry 93, 171-182.

Dagnino, A., Allen, J.I., Moore, M.N., Broeg, K., Canesi, L., Viarengo, A., 2007. Development of an expert system for the integration of biomarker responses in mussels into an animal health index. Biomarkers 12, 155-172.

Dagnino, A., Viarengo, A., 2014. Development of a decision support system to manage contamination in marine ecosystems. Science of the Total Environment 466, 119-126.

Davies, M.J., 2005. The oxidative environment and protein damage. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1703, 93-109.

Dayras, P., Charmantier, G., Chaumot, A., Vigneron, A., Coquery, M., Quéau, H., Artells, E., Lignot, J.H., Geffard, O., Issartel, J., 2017. Osmoregulatory responses to cadmium in reference and historically metal contaminated Gammarus fossarum (Crustacea, Amphipoda) populations. Chemosphere 180, 412-422.

De Kleijn, D.P., Van Herp, F., 1995. Molecular biology of neurohormone precursors in the eyestalk of Crustacea. Comp Biochem Physiol B Biochem Mol Biol 112, 573-579.

Dedhar, S., Rennie, P.S., Shago, M., Hagesteijn, C.Y., Yang, H., Filmus, J., Hawley, R.G., Bruchovsky, N., Cheng, H., Matusik, R.J., et al., 1994. Inhibition of nuclear hormone receptor activity by calreticulin. Nature 367, 480-483.

Dedourge-Geffard, O., Charron, L., Hofbauer, C., Gaillet, V., Palais, F., Lacaze, E., Geffard, A., Geffard, O., 2013. Temporal patterns of digestive enzyme activities and feeding rate in gammarids (Gammarus fossarum) exposed to inland polluted waters. Ecotoxicology and Environmental Safety 97, 139-146.

Depiereux, S., Liagre, M., Danis, L., De Meulder, B., Depiereux, E., Segner, H., Kestemont, P., 2014. Intersex Occurrence in Rainbow Trout (Oncorhynchus mykiss) Male Fry Chronically Exposed to Ethynylestradiol. PLoS ONE 9, e98531.

Depledge, M.H., 1994. The rational basis for the use of biomarkers as ecotoxicological tools, in: Fossi, M.C., Leonzio, C. (Eds.), Nondestructive biomarkers in vertebrates. Lewis Publishers, pp. 261-285.

DeSouza, L.V., Siu, K.W.M., 2013. Mass spectrometry-based quantification. Clinical Biochemistry 46, 421-431.

Devin, S., Burgeot, T., Giambérini, L., Minguez, L., Pain-Devin, S., 2014. The integrated biomarker response revisited: optimization to avoid misuse. Environmental Science and Pollution Research 21, 2448-2454.

Di Giulio, R.T., Benson, W.H., Sanders, B. M., Van Veld, P.A., 1995. Biochemical mechanisms: Metabolism,adaptation, and toxicity, Fundamentals of aquatic toxicology : effects, environmental fate, and risk assessment

2nd ed. Taylor & Francis, Washington, D.C., pp. 523-561.

Drabovich, A.P., Martínez-Morillo, E., Diamandis, E.P., 2015. Toward an integrated pipeline for protein biomarker development. Biochimica et Biophysica Acta - Proteins and Proteomics 1854, 677-686.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792-1797.

Enis Yonar, M., Mise Yonar, S., Silici, S., 2011. Protective effect of propolis against oxidative stress and immunosuppression induced by oxytetracycline in rainbow trout (Oncorhynchus mykiss, W.). Fish Shellfish Immunol 31, 318-325.

Erickson, A.R., Cantarel, B.L., Lamendella, R., Darzi, Y., Mongodin, E.F., Pan, C., Shah, M., Halfvarson, J., Tysk, C., Henrissat, B., Raes, J., Verberkmoes, N.C., Fraser, C.M., Hettich, R.L., Jansson, J.K., 2012. Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease. PLoS ONE 7.

Escher, B.I., Hermens, J.L., 2002. Modes of action in ecotoxicology: their role in body burdens, species sensitivity, QSARs, and mixture effects. Environ Sci Technol 36, 4201-4217.

Falisse, E., Voisin, A.-S., Silvestre, F., 2017. Impacts of triclosan exposure on zebrafish early-life stage: Toxicity and acclimation mechanisms. Aquatic Toxicology 189, 97-107.

Fan, L., Wang, A., Miao, Y., Liao, S., Ye, C., Lin, Q., 2016. Comparative proteomic identification of the hepatopancreas response to cold stress in white shrimp, Litopenaeus vannamei. Aquaculture 454, 27-34.

Felten, V., 2003. Effets de l'acidification des ruisseaux vosgiens sur la biologie, l'écologie et l'écophysiologie de Gammarus fossarum Koch, 1835 (Crustacea Amphipoda) : Approche intégrée à différents niveaux d'organisation, pp. [340]p.-[340]p.

Forbes, V.E., Palmqvist, A., Bach, L., 2006. The use and misuse of biomarkers in ecotoxicology. Environ Toxicol Chem 25, 272-280.

Ford, A.T., Fernandes, T.F., Rider, S.A., Read, P.A., Robinson, C.D., Davies, I.M., 2004. Endocrine disruption in a marine amphipod? Field observations of intersexuality and de-masculinisation. Marine environmental research 58, 169-173.

Fossi Tankoua, O., Buffet, P.E., Amiard, J.C., Berthet, B., Mouneyrac, C., Amiard-Triquet, C., 2013. Integrated assessment of estuarine sediment quality based on a multi-biomarker approach in the bivalve Scrobicularia plana. Ecotoxicology and Environmental Safety 88, 117-125.

Fulton, M.H., Key, P.B., 2001. Acetylcholinesterase inhibition in estuarine fish and invertebrates as an indicator of organophosphorus insecticide exposure and effects. Environmental Toxicology And Chemistry 20, 37-45.

Gaertner, K., Chandler, G.T., Quattro, J., Ferguson, P.L., Sabo-Attwood, T., 2012. Identification and expression of the ecdysone receptor in the harpacticoid copepod, Amphiascus tenuiremis, in response to fipronil. Ecotoxicology and environmental safety 76, 39-45.

Gallien, S., Duriez, E., Domon, B., 2011. Selected reaction monitoring applied to proteomics. J Mass Spectrom 46, 298-312.

Galloway, T.S., Brown, R.J., Browne, M.A., Dissanayake, A., Lowe, D., Jones, M.B., Depledge, M.H., 2004. Ecosystem management bioindicators: the ECOMAN project – a multi-biomarker approach to ecosystem management. Marine Environmental Research 58, 233-237.

Gao, J., Wang, X., Zou, Z., Jia, X., Wang, Y., Zhang, Z., 2014. Transcriptome analysis of the differences in gene expression between testis and ovary in green mud crab (Scylla paramamosain). BMC genomics 15, 585-585.

Garcia-Reyero, N., Griffitt, R.J., Liu, L., Kroll, K.J., Farmerie, W.G., Barber, D.S., Denslow, N.D., 2008. Construction of a robust microarray from a non-model species largemouth bass, Micropterus salmoides (Lacèpede), using pyrosequencing technology. Journal of Fish Biology 72, 2354-2376.

Garcia-Reyero, N., Raldua, D., Quiros, L., Llaveria, G., Cerda, J., Barcelo, D., Grimalt, J.O., Pina, B., 2004. Use of vitellogenin mRNA as a biomarker for endocrine disruption in feral and cultured fish. Anal Bioanal Chem 378, 670-675.

Gaudet, J., VanderElst, I., Spence, A.M., 1996. Post-transcriptional regulation of sex determination in Caenorhabditis elegans: widespread expression of the sex-determining gene fem-1 in both sexes. Mol Biol Cell 7, 1107-1121.

Geffard, O., Xuereb, B., Chaumot, A., Geffard, A., Biagianti, S., Noël, C., Abbaci, K., Garric, J., Charmantier, G., Charmantier-Daures, M., 2010. Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment. Environmental toxicology and chemistry / SETAC 29, 2249-2259.

Geret, F., Cosson, R.P., 2000. The use of metallothioneins as biomarker of metal contamination: Variability between sites and organs in the oyster Crassostrea gigas. Oceanologica Acta 23, 261-271.

Ghisi, N.d.d.C., 2012. Relationship Between Biomarkers and Pesticide Exposure in Fishes: A Review, in: Soundararajan, R.P. (Ed.), Pesticides - Advances in Chemical and Botanical Pesticides. InTech, Rijeka, p. Ch. 15.

Gianazza, E., Miller, I., Palazzolo, L., Parravicini, C., Eberini, I., 2016. With or without you — Proteomics with or without major plasma/serum proteins. Journal of Proteomics 140, 62-80.

Gillet, L.C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., Aebersold, R., 2012. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics 11, 18.

Gismondi, E., Cossu-Leguille, C., Beisel, J.N., 2013. Do male and female gammarids defend themselves differently during chemical stress? Aquatic Toxicology 140, 432-438.

Gismondi, E., Mazzucchelli, G., De Pauw, E., Joaquim-Justo, C., Thomé, J.P., 2015. Gender differences in responses in Gammarus pulex exposed to BDE-47: A gel-free proteomic approach. Ecotoxicology and Environmental Safety 122, 205-213.

Gismondi, E., Thomé, J.P., 2016. Transcriptome of the freshwater amphipod Gammarus pulex hepatopancreas. Genomics Data 8, 91-92.

Gismondi, E., Thomé, J.P., Urien, N., Uher, E., Baiwir, D., Mazzucchelli, G., De Pauw, E., Fechner, L.C., Lebrun, J.D., 2017. Ecotoxicoproteomic assessment of the functional alterations caused by chronic metallic exposures in gammarids. Environmental Pollution 225, 428-438.

Glazer, L., Tom, M., Weil, S., Roth, Z., Khalaila, I., Mittelman, B., Sagi, A., 2013. Hemocyanin with phenoloxidase activity in the chitin matrix of the crayfish gastrolith. Journal of Experimental Biology 216, 1898-1904.

Goksøyr, A., Förlin, L., 1992. The cytochrome P-450 system in fish, aquatic toxicology and environmental monitoring. Aquatic Toxicology 22, 287-311.

Gómez-Canela, C., Miller, T.H., Bury, N.R., Tauler, R., Barron, L.P., 2016. Targeted metabolomics of Gammarus pulex following controlled exposures to selected pharmaceuticals in water. Science of the Total Environment 562, 777-788.

Gouveia, D., Chaumot, A., Charnot, A., Quéau, H., Armengaud, J., Almunia, C., Salvador, A., Geffard, O., 2017. Assessing the relevance of a multiplexed methodology for proteomic biomarker measurement in the invertebrate species Gammarus fossarum: a physiological and ecotoxicological study. Aquatic Toxicology 190, 199-209.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010 Feb;27(2):221-4. doi: 10.1093/molbev/msp259. Epub 2009 Oct 23.

Greenbaum, D., Colangelo, C., Williams, K., Gerstein, M., 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol 4, 29.

Grossmann, J., Fernández, H., Chaubey, P.M., Valdés, A.E., Gagliardini, V., Cañal, M.J., Russo, G., Grossniklaus, U., 2017. Proteogenomic analysis greatly expands the identification of proteins related to reproduction in the apogamous fern Dryopteris affinis ssp. affinis. Frontiers in Plant Science 8.

Guan, J., Lv, Y., Zhang, Y., Ning, Q., 2016. A shortening effect of KK-42 on the moult cycle of juvenile (De Haan, 1849) (Decapoda, Palaemonidae). Crustaceana 89, 85-95.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307-321.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17, 994-999.

Hagger, J.A., Jones, M.B., Lowe, D., Leonard, D.R.P., Owen, R., Galloway, T.S., 2008. Application of biomarkers for improving risk assessments of chemicals under the Water Framework Directive: A case study. Marine Pollution Bulletin 56, 1111-1118.

Han, X., Aslanian, A., Yates, J.R., 2008. Mass Spectrometry for Proteomics. Current opinion in chemical biology 12, 483-490.

Harlan, R., Zhang, H., 2014. Targeted proteomics: a bridge between discovery and validation. Expert Review of Proteomics 11, 657-661.

Harris, R.R., Santos, M.C.F., 2000. Heavy metal contamination and physiological variability in the Brazilian mangrove crabs Ucides cordatus and Callinectes danae (Crustacea: Decapoda). Marine Biology 137, 691-703.

Helmholz, H., Lassen, S., Ruhnau, C., Pröfrock, D., Erbslöh, H.B., Prange, A., 2015. Investigation on the proteome response of transplanted blue mussel (Mytilus sp.) during a long term exposure experiment at differently impacted field stations in the German Bight (North Sea). Marine environmental research 110, 69-80.

Herrero, Ó., Planelló, R., Morcillo, G., 2015. The plasticizer benzyl butyl phthalate (BBP) alters the ecdysone hormone pathway, the cellular response to stress, the energy metabolism, and several detoxication mechanisms in Chironomus riparius larvae. Chemosphere 128, 266-277.

Hidalgo-Galiana, A., Monge, M., Biron, D.G., Canals, F., Ribera, I., Cieslak, A., 2014. Reproducibility and Consistency of Proteomic Experiments on Natural Populations of a Non-Model Aquatic Insect. PLoS ONE 9, e104734.

Hook, S.E., Gallagher, E.P., Batley, G.E., 2014a. The Role of Biomarkers in the Assessment of Aquatic Ecosystem Health. Integrated environmental assessment and management 10, 327-341.

Hook, S.E., Osborn, H.L., Spadaro, D.A., Simpson, S.L., 2014b. Assessing mechanisms of toxicant response in the amphipod Melita plumulosa through transcriptomic profiling. Aquatic toxicology (Amsterdam, Netherlands) 146, 247-257.

Hook, S.E., Skillman, A.D., Small, J.A., Schultz, I.R., 2006. Gene expression patterns in rainbow trout, Oncorhynchus mykiss, exposed to a suite of model toxicants. Aquatic Toxicology 77, 372-385.

Hook, S.E., Twine, N.A., Simpson, S.L., Spadaro, D.A., Moncuquet, P., Wilkins, M.R., 2014c. 454 pyrosequencing-based analysis of gene expression profiles in the amphipod Melita plumulosa: transcriptome assembly and toxicant induced changes. Aquatic toxicology (Amsterdam, Netherlands) 153, 73-88.

Horiguchi, T., Shiraishi, H., Shimizu, M., Yamazaki, S., Morita, M., 1995. Imposex in Japanese gastropods (neogastropoda and mesogastropoda): Effects of tributyltin and triphenyltin from antifouling paints. Marine Pollution Bulletin 31, 402-405.

Höss, S., Weltje, L., 2007. Endocrine disruption in nematodes: Effects and mechanisms. Ecotoxicology 16, 15-28.

Hu, W., Culloty, S., Darmody, G., Lynch, S., Davenport, J., Ramirez-Garcia, S., Dawson, K.A., Lynch, I., Blasco, J., Sheehan, D., 2014. Toxicity of copper oxide nanoparticles in the blue mussel, Mytilus edulis: A redox proteomic investigation. Chemosphere 108, 289-299.

Huang, Q., Huang, H.Q., 2012. Alterations of protein profile in zebrafish liver cells exposed to methyl parathion: A membrane proteomics approach. Chemosphere 87, 68-76.

Huang, Q., Yang, L., Luo, J., Guo, L., Wang, Z., Yang, X., Jin, W., Fang, Y., Ye, J., Shan, B., Zhang, Y., 2015. SWATH enables precise label-free quantification on proteome scale. Proteomics 15, 1215-1223.

Hui, J.H.L., Hayward, A., Bendena, W.G., Takahashi, T., Tobe, S.S., 2010. Evolution and functional divergence of enzymes involved in sesquiterpenoid hormone biosynthesis in crustaceans and insects. Peptides 31, 451-455.

Hyne, R.V., 2011. Review of the reproductive biology of amphipods and their endocrine regulation: identification of mechanistic pathways for reproductive toxicants. Environmental toxicology and chemistry / SETAC 30, 2647-2657.

Hyne, R.V., Gale, S.A., King, C.K., 2005. Laboratory culture and life-cycle experiments with the benthic amphipod melita plumulosa (zeidler). Environmental Toxicology And Chemistry 24, 2065-2073.

Jemec, A., Drobne, D., Tisler, T., Sepcić, K., 2010. Biochemical biomarkers in environmental studies--lessons learnt from enzymes catalase, glutathione S-transferase and cholinesterase in two crustacean species. Environmental science and pollution research international 17, 571-581.

Jensen, O.N., 2006. Interpreting the protein language using proteomics. Nat Rev Mol Cell Biol 7, 391-403.

Jha, A.N., 2008. Ecotoxicological applications and significance of the comet assay. Mutagenesis 23, 207-221.

Jiang, C., Baehrecke, E.H., Thummel, C.S., 1997. Steroid regulated programmed cell death during Drosophila metamorphosis. Development 124, 4673-4683.

Jin, S., Fu, H., Zhou, Q., Sun, S., Jiang, S., Xiong, Y., Gong, Y., Qiao, H., Zhang, W., 2013. Transcriptome Analysis of Androgenic Gland for Discovery of Novel Genes from the Oriental River Prawn, <italic>Macrobrachium nipponense</italic>, Using Illumina Hiseq 2000. PLoS ONE 8, e76840.

Jobling, S., Sumpter, J.P., Sheahan, D., Osborne, J.A., Matthiessen, P., 1996. Inhibition of testicular growth in rainbow trout (Oncorhynchus mykiss) exposed to estrogenic alkylphenolic chemicals. Environmental Toxicology And Chemistry 15, 194-202.

Jobling, S., Tyler, C.R., 2003. Endocrine disruption in wild freshwater fish, Pure and Applied Chemistry, p. 2219.

Jubeaux, G., Audouard-Combe, F., Simon, R., Tutundjian, R., Salvador, A., Geffard, O., Chaumot, A., 2012a. Vitellogenin-like proteins among invertebrate species diversity: potential of proteomic mass spectrometry for biomarker development. Environmental science & technology 46, 6315-6323.

Jubeaux, G., Simon, R., Salvador, A., Lopes, C., Lacaze, E., Quéau, H., Chaumot, A., Geffard, O., 2012b. Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: Inconclusive experience. Aquatic Toxicology 122-123, 9-18.

Jubeaux, G., Simon, R., Salvador, A., Quéau, H., Chaumot, A., Geffard, O., 2012c. Vitellogenin-like proteins in the freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males. Aquatic toxicology (Amsterdam, Netherlands) 112-113, 72-82.

Karim, F.D., Guild, G.M., Thummel, C.S., 1993. The Drosophila Broad-Complex plays a key role in controlling ecdysone-regulated gene expression at the onset of metamorphosis. Development 118, 977-988.

Kerambrun, E., Henry, F., Marechal, A., Sanchez, W., Minier, C., Filipuci, I., Amara, R., 2012. A multibiomarker approach in juvenile turbot, Scophthalmus maximus, exposed to contaminated sediments. Ecotoxicology and Environmental Safety 80, 45-53.

Khalsa-Moyers, G., McDonald, W.H., 2006. Developments in mass spectrometry for the analysis of complex protein mixtures. Brief Funct Genomic Proteomic 5, 98-111.

Kim, H.J., Koedrith, P., Seo, Y.R., 2015a. Ecotoxicogenomic Approaches for Understanding Molecular Mechanisms of Environmental Chemical Toxicity Using Aquatic Invertebrate, Daphnia Model Organism. International Journal of Molecular Sciences 16, 12261-12287.

Kim, Y.J., Sertamo, K., Pierrard, M.A., Mesmin, C., Kim, S.Y., Schlesser, M., Berchem, G., Domon, B., 2015b. Verification of the biomarker candidates for non-small-cell lung cancer using a targeted proteomics approach. Journal of Proteome Research 14, 1412-1419.

Kirby, M.F., Allen, Y.T., Dyer, R.A., Feist, S.W., Katsiadaki, I., Matthiessen, P., Scott, A.P., Smith, A., Stentiford, G.D., Thain, J.E., Thomas, K.V., Tolhurst, L., Waldock, M.J., 2004. Surveys of plasma vitellogenin and intersex in male flounder (Platichthys flesus) as measures of endocrine disruption by estrogenic contamination in United Kingdom estuaries: temporal trends, 1996 to 2001. Environ Toxicol Chem 23, 748-758.

Kloas, W., Urbatzka, R., Opitz, R., Würtz, S., Behrends, T., Hermelink, B., Hofmann, F., Jagnytsch, O., Kroupova, H., Lorenz, C., Neumann, N., Pietsch, C., Trubiroha, A., Van Ballegooy, C., Wiedemann, C., Lutz, I., 2009. Endocrine disruption in aquatic vertebrates, Annals of the New York Academy of Sciences, pp. 187-200.

Konopova, B., Smykal, V., Jindra, M., 2011. Common and Distinct Roles of Juvenile Hormone Signaling Genes in Metamorphosis of Holometabolous and Hemimetabolous Insects. PLoS ONE 6, e28728.

Kontogiannatos, D., Swevers, L., Zakasis, G., Kourti, A., 2014. The molecular and physiological impact of bisphenol A in Sesamia nonagrioides (Lepidoptera: Noctuidae). Ecotoxicology 24, 356-367.

Kornthong, N., Cummins, S.F., Chotwiwatthanakun, C., Khornchatri, K., Engsusophon, A., Hanna, P.J., Sobhon, P., 2014. Identification of genes associated with reproduction in the mud crab (Scylla olivacea) and their differential expression following serotonin stimulation. PLoS ONE 9.

Kumar, H., Lim, H.-W., More, S.V., Kim, B.-W., Koppula, S., Kim, I.S., Choi, D.-K., 2012. The Role of Free Radicals in the Aging Brain and Parkinson's Disease: Convergence and Parallelism. International Journal of Molecular Sciences 13, 10478.

Kunz, P.Y., Kienle, C., Gerhardt, A., 2010. Gammarus spp. in aquatic ecotoxicology and water quality assessment: toward integrated multilevel tests. Reviews of environmental contamination and toxicology 205, 1-76.

Kussmann, M., Raymond, F., Affolter, M., 2006. OMICS-driven biomarker discovery in nutrition and health. Journal of Biotechnology 124, 758-787.

Lacaze, E., 2011. Un biomarqueur de génotoxicité chez Gammarus fossarum : développement, signification fonctionnelle et application au milieu naturel.

Lacaze, E., Devaux, A., Jubeaux, G., Mons, R., Gardette, M., Bony, S., Garric, J., Geffard, O., 2011a. DNA damage in Gammarus fossarum sperm as a biomarker of genotoxic pressure: intrinsic variability and reference level. Science of the Total Environment 409, 3230-3236.

Lacaze, E., Devaux, A., Mons, R., Bony, S., Garric, J., Geffard, A., Geffard, O., 2011b. DNA damage in caged Gammarus fossarum amphipods: a tool for freshwater genotoxicity assessment. Environ Pollut 159, 1682-1691.

Ladhar-Chaabouni, R., Machreki-Ajmi, M., Hamza-Chaffai, A., 2012. Use of metallothioneins as biomarkers for environmental quality assessment in the Gulf of Gabes (Tunisia). Environ Monit Assess 184, 2177-2192.

Lafontaine, A., Baiwir, D., Joaquim-Justo, C., De Pauw, E., Lemoine, S., Boulangé-Lecomte, C., Forget-Leray, J., Thomé, J.P., Gismondi, E., 2017. Proteomic response of Macrobrachium rosenbergii hepatopancreas exposed to chlordecone: Identification of endocrine disruption biomarkers? Ecotoxicology and Environmental Safety 141, 306-314.

LeBlanc, G.A., 2007. Crustacean endocrine toxicology: A review. Ecotoxicology 16, 61-81.

Lee, J.W., Won, E.-J., Raisuddin, S., Lee, J.-S., 2015. Significance of adverse outcome pathways in biomarker-based environmental risk assessment in aquatic organisms. Journal of Environmental Sciences 35, 115-127.

Lee, S.O., Jeon, J.M., Oh, C.W., Kim, Y.M., Kang, C.K., Lee, D.S., Mykles, D.L., Kim, H.W., 2011. Two juvenile hormone esterase-like carboxylesterase cDNAs from a Pandalus shrimp (Pandalopsis japonica): cloning, tissue expression, and effects of eyestalk ablation. Comp Biochem Physiol B Biochem Mol Biol 159, 148-156.

Leroy, D., Haubruge, E., De Pauw, E., Thomé, J.P., Francis, F., 2010. Development of ecotoxicoproteomics on the freshwater amphipod Gammarus pulex: identification of PCB biomarkers in glycolysis and glutamate pathways. Ecotoxicology and environmental safety 73, 343-352.

Li, Y., Zhang, Z., Robinson, G.E., Palli, S.R., 2007. Identification and characterization of a juvenile hormone response element and its binding proteins. J Biol Chem 282, 37605-37617.

Liu, C.-H., Tseng, D.-Y., Lai, C.-Y., Cheng, W., Kuo, C.-M., 2006. Molecular cloning and characterisation of prophenoloxidase cDNA from haemocytes of the giant freshwater prawn, Macrobrachium rosenbergii, and its transcription in relation with the moult stage. Fish & Shellfish Immunology 21, 60-69.

Liu, C.H., Chang, C.C., Chiu, Y.C., Cheng, W., Yeh, M.S., 2011. Identification and cloning of a transglutaminase from giant freshwater prawn, Macrobrachium rosenbergii, and its

transcription during pathogen infection and moulting. Fish and Shellfish Immunology 31, 871-880.

Liu, F., Wang, D.Z., Wang, W.X., 2012. Cadmium-induced changes in trace element bioaccumulation and proteomics perspective in four marine bivalves. Environmental Toxicology And Chemistry 31, 1292-1300.

Liu, H., Sadygov, R.G., Yates, J.R., 3rd, 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Analytical chemistry 76, 4193-4201.

Liubicich, D.M., Serano, J.M., Pavlopoulos, A., Kontarakis, Z., Protas, M.E., Kwan, E., Chatterjee, S., Tran, K.D., Averof, M., Patel, N.H., 2009. Knockdown of Parhyale Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. Proc Natl Acad Sci U S A 106, 13892-13896.

Lomax, D.P., Roubal, W.T., Moore, J.D., Johnson, L.L., 1998. An enzyme-linked immunosorbent assay (ELISA) for measuring vitellogenin in English sole (Pleuronectes vetulus): Development, validation and cross-reactivity with other pleuronectids. Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology 121, 425-436.

Loo, J.A., Udseth, H.R., Smith, R.D., 1989. Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. Analytical Biochemistry 179, 404-412.

Lopez-Casado, G., Covey, P.A., Bedinger, P.A., Mueller, L.A., Thannhauser, T.W., Zhang, S., Fei, Z., Giovannoni, J.J., Rose, J.K.C., 2012. Enabling proteomic studies with RNA-Seq: The proteome of tomato pollen as a test case. PROTEOMICS 12, 761-774.

Lucu, Č., Towle, D.W., 2003. Na++K+-ATPase in gills of aquatic crustacea. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 135, 195-214.

Luo, J., Pei, S., Jing, W., Zou, E., Wang, L., 2015. Cadmium inhibits molting of the freshwater crab Sinopotamon henanense by reducing the hemolymph ecdysteroid content and the activities of chitinase and N-acetyl-beta-glucosaminidase in the epidermis. Comp Biochem Physiol C Toxicol Pharmacol 169, 1-6.

Lushchak, V.I., 2011. Environmentally induced oxidative stress in aquatic animals. Aquat Toxicol 101, 13-30.

Mæland Nilsen, M., Meier, S., Larsen, B.K., Ketil Andersen, O., Hjelle, A., 2011. An estrogen-responsive plasma protein expression signature in Atlantic cod (Gadus morhua) revealed by SELDI-TOF MS. Ecotoxicology and Environmental Safety 74, 2175-2181.

Maes, G.E., Raeymaekers, J.A.M., Hellemans, B., Geeraerts, C., Parmentier, K., De Temmerman, L., Volckaert, F.A.M., Belpaire, C., 2013. Gene transcription reflects poor health status of resident European eel chronically exposed to environmental pollutants. Aquatic Toxicology 126, 242-255.

Major, K., Soucek, D.J., Giordano, R., Wetzel, M.J., Soto-Adames, F., 2013. The common ecotoxicology laboratory strain of Hyalella azteca is genetically distinct from most wild strains sampled in eastern North America. Environ Toxicol Chem 32, 2637-2647.

Malécot, M., Mezhoud, K., Marie, A., Praseuth, D., Puiseux-Dao, S., Edery, M., 2009. Proteomic study of the effects of microcystin-LR on organelle and membrane proteins in medaka fish liver. Aquatic Toxicology 94, 153-161.

Marigómez, I., Garmendia, L., Soto, M., Orbea, A., Izagirre, U., Cajaraville, M.P., 2013. Marine ecosystem health status assessment through integrative biomarker indices: A comparative study after the Prestige oil spill "mussel Watch". Ecotoxicology 22, 486-505.

Markov, G.V., Tavares, R., Dauphin-Villemant, C., Demeneix, B.A., Baker, M.E., Laudet, V., 2009. Independent elaboration of steroid hormone signaling pathways in metazoans. Proc Natl Acad Sci U S A 106, 11913-11918.

Martínez-Paz, P., Morales, M., Urien, J., Morcillo, G., Martínez-Guitarte, J.L., 2017. Endocrine-related genes are altered by antibacterial agent triclosan in Chironomus riparius aquatic larvae. Ecotoxicology and Environmental Safety 140, 185-190.

Martinez, M., Ramo, J.D., Torreblanca, A., Pastor, A., Diaz-Mayans, J., 1996. Cadmium toxicity, accumulation and metallothionein induction in echinogammarus echinosetosus. Journal of Environmental Science and Health . Part A: Environmental Science and Engineering and Toxicology 31, 1605-1617.

Martyniuk, C.J., Kroll, K.J., Doperalski, N.J., Barber, D.S., Denslow, N.D., 2010. Environmentally relevant exposure to 17α-ethinylestradiol affects the telencephalic proteome of male fathead minnows. Aquatic Toxicology 98, 344-353.

Martyniuk, C.J., Simmons, D.B., 2016. Spotlight on environmental omics and toxicology: a long way in a short time. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 19, 97-101.

Matozzo, V., Gagne, F., Marin, M.G., Ricciardi, F., Blaise, C., 2008. Vitellogenin as a biomarker of exposure to estrogenic compounds in aquatic invertebrates: a review. Environ Int 34, 531-545.

Mermelekas, G., Vlahou, A., Zoidakis, J., 2015. SRM/MRM targeted proteomics as a tool for biomarker validation and absolute quantification in human urine. Expert Review of Molecular Diagnostics 15, 1441-1454.

Meucci, V., Arukwe, A., 2005. Detection of vitellogenin and zona radiata protein expressions in surface mucus of immature juvenile Atlantic salmon (Salmo salar) exposed to waterborne nonylphenol. Aquatic Toxicology 73, 1-10.

Miller, D.H., Jensen, K.M., Villeneuve, D.L., Kahl, M.D., Makynen, E.A., Durhan, E.J., Ankley, G.T., 2007. Linkage of biochemical responses to population-level effects: A case study with vitellogenin in the fathead minnow (Pimephales promelas). Environmental Toxicology And Chemistry 26, 521-527.

Minier, C., Levy, F., Rabel, D., Bocquené, G., Godefroy, D., Burgeot, T., Leboulenger, F., 2000. Flounder health status in the Seine Bay. A multibiomarker study. Marine environmental research 50, 373-377.

Monteoliva, L., Albar, J.P., 2004. Differential proteomics: an overview of gel and non-gel based approaches. Brief Funct Genomic Proteomic 3, 220-239.

Morales, M., Martínez-Paz, P., Martín, R., Planelló, R., Urien, J., Martínez-Guitarte, J.L., Morcillo, G., 2014. Transcriptional changes induced by in vivo exposure to pentachlorophenol (PCP) in

Chironomus riparius (Diptera) aquatic larvae. Aquatic toxicology (Amsterdam, Netherlands) 157C, 1-9.

Morales, M., Martínez-Paz, P., Ozáez, I., Martínez-Guitarte, J.L., Morcillo, G., 2013. DNA damage and transcriptional changes induced by tributyltin (TBT) after short in vivo exposures of Chironomus riparius (Diptera) larvae. Comparative Biochemistry and Physiology - C Toxicology and Pharmacology 158, 57-63.

Moreno, Y., Gros, P.-P., Tam, M., Segura, M., Valanparambil, R., Geary, T.G., Stevenson, M.M., 2011. Proteomic Analysis of Excretory-Secretory Products of Heligmosomoides polygyrus Assessed with Next-Generation Sequencing Transcriptomic Information. PLOS Neglected Tropical Diseases 5, e1370.

Mykles, D.L., 2011. Ecdysteroid metabolism in crustaceans. The Journal of steroid biochemistry and molecular biology 127, 196-203.

Nagaraju, G.P.C., 2007. Is methyl farnesoate a crustacean hormone? Aquaculture 272, 39-54.

Nagaraju, G.P.C., Borst, D.W., 2008. Methyl farnesoate couples environmental changes to testicular development in the crab Carcinus maenas. Journal of Experimental Biology 211, 2773-2778.

Neilson, K.A., Ali, N.A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S.C., Haynes, P.A., 2011. Less label, more free: Approaches in label-free quantitative mass spectrometry. Proteomics 11, 535-553.

Nesvizhskii, A.I., 2007. Protein identification by tandem mass spectrometry and sequence database searching. Methods Mol Biol 367, 87-119.

Newman, M.C., 2015. Molecular Effects and Biomarkers, Fundamentals of ecotoxicology : the science of pollution, Fourth edition. ed. CRC Press Taylor and Francis Group, New York, pp. xxv, 654 pages.

Nunes, B., Carvalho, F., Guilhermino, L., 2006. Effects of widely used pharmaceuticals and a detergent on oxidative stress biomarkers of the crustacean Artemia parthenogenetica. Chemosphere 62, 581-594.

Nzoughet, J.K., Hamilton, J.T.G., Botting, C.H., Douglas, A., Devine, L., Nelson, J., Elliott, C.T., 2009. Proteomics identification of azaspiracid toxin biomarkers in blue mussels, Mytilus edulis. Molecular and Cellular Proteomics 8, 1811-1822.

Oehlmann, J., Di Benedetto, P., Tillmann, M., Duft, M., Oetken, M., Schulte-Oehlmann, U., 2007. Endocrine disruption in prosobranch molluscs: Evidence and ecological relevance. Ecotoxicology 16, 29-43.

Oetken, M., Bachmann, J., Schulte-Oehlmann, U., Oehlmann, J., 2004. Evidence for endocrine disruption in invertebrates. International review of cytology 236, 1-44.

Oikari, A., 2006. Caging techniques for field exposures of fish to chemical contaminants. Aquatic Toxicology 78, 370-381.

Olmstead, A.W., Leblanc, G.A., 2002. Juvenoid hormone methyl farnesoate is a sex determinant in the crustacean Daphnia magna. J Exp Zool 293, 736-739.

Olmstead, A.W., LeBlanc, G.A., 2003. Insecticidal juvenile hormone analogs stimulate the production of male offspring in the crustacean Daphnia magna. Environmental health perspectives 111, 919-924.

Olsvik, P.A., Samuelsen, O.B., Agnalt, A.L., Lunestad, B.T., 2015. Transcriptional responses to teflubenzuron exposure in European lobster (Homarus gammarus). Aquatic Toxicology 167, 143-156.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M., 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1, 376-386.

Onisko, B., Dynin, I., Requena, J.R., Silva, C.J., Erickson, M., Carter, J.M., 2007. Mass Spectrometric Detection of Attomole Amounts of the Prion Protein by nanoLC/MS/MS. Journal of the American Society for Mass Spectrometry 18, 1070-1079.

Parkinson, A., B.W. Ogilvie, 2008. Biotransformations of xenobiotics, in: Klaassen, C.D., Ed., McGraw-Hill (Ed.), Casarett & Doull's Toxicology: The Basic Science of Poisons, 7th ed, New York.

Pascoal, S., Carvalho, G., Vasieva, O., Hughes, R., Cossins, A., Fang, Y., Ashelford, K., Olohan, L., Barroso, C., Mendo, S., Creer, S., 2013. Transcriptomics and in vivo tests reveal novel mechanisms underlying endocrine disruption in an ecological sentinel, Nucella lapillus. Molecular Ecology 22, 1589-1608.

Patterson, S.D., Aebersold, R.H., 2003. Proteomics: the first decade and beyond. Nat Genet 33, 311-323.

Pavlopoulos, A., Kontarakis, Z., Liubicich, D.M., Serano, J.M., Akam, M., Patel, N.H., Averof, M., 2009. Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. Proc Natl Acad Sci U S A 106, 13897-13902.

Pedersen, S.N., Lundebye, A.K., Depledge, M.H., 1997. Field application of metallothionein and stress protein biomarkers in the shore crab (Carcinus maenas) exposed to trace metals. Aquatic Toxicology 37, 183-200.

Pesch, Y.Y., Riedel, D., Behr, M., 2017. Drosophila Chitinase 2 is expressed in chitin producing organs for cuticle formation. Arthropod Structure and Development 46, 4-12.

Petrak, J., Ivanek, R., Toman, O., Cmejla, R., Cmejlova, J., Vyoral, D., Zivny, J., Vulpe, C.D., 2008. Deja vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins. Proteomics 8, 1744-1749.

Picado, A., Bebianno, M.J., Costa, M.H., Ferreira, A., Vale, C., 2007. Biomarkers: a strategic tool in the assessment of environmental quality of coastal waters. Hydrobiologia 587, 79.

Picotti, P., Aebersold, R., 2012. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Meth 9, 555-566.

Piulachs, M.-D., Pagone, V., Bellés, X., 2010. Key roles of the Broad-Complex gene in insect embryogenesis. Insect biochemistry and molecular biology 40, 468-475.

Planelló, R., Martínez-Guitarte, J.L., Morcillo, G., 2008. The endocrine disruptor bisphenol A increases the expression of HSP70 and ecdysone receptor genes in the aquatic larvae of Chironomus riparius. Chemosphere 71, 1870-1876.

Ponton, F., Lefevre, T., Lebarbenchon, C., Thomas, F., Loxdale, H.D., Marché, L., Renault, L., Perrot-Minnot, M.J., Biron, D.G., 2006. Do distantly related parasites rely on the same proximate factors to alter the behaviour of their hosts? Proceedings of the Royal Society B: Biological Sciences 273, 2869-2877.

Poynton, H.C., Varshavsky, J.R., Chang, B., Cavigiolio, G., Chan, S., Holman, P.S., Loguinov, A.V., Bauer, D.J., Komachi, K., Theil, E.C., Perkins, E.J., Hughes, O., Vulpe, C.D., 2007. Daphnia magna ecotoxicogenomics provides mechanistic insights into metal toxicity. Environmental Science and Technology 41, 1044-1050.

Qu, Z., Kenny, N.J., Lam, H.M., Chan, T.F., Chu, K.H., Bendena, W.G., Tobe, S.S., Hui, J.H.L., 2015. How Did Arthropod Sesquiterpenoids and Ecdysteroids Arise? Comparison of Hormonal Pathway Genes in Noninsect Arthropod Genomes. Genome Biology and Evolution 7, 1951-1959.

Radwan, M.A., El-Gendy, K.S., Gad, A.F., 2010. Biomarkers of oxidative stress in the land snail, Theba pisana for assessing ecotoxicological effects of urban metal pollution. Chemosphere 79, 40-46.

Ralston-Hooper, K.J., Turner, M.E., Soderblom, E.J., Villeneuve, D., Ankley, G.T., Moseley, M.A., Hoke, R.A., Ferguson, P.L., 2013. Application of a label-free, gel-free quantitative proteomics method for ecotoxicological studies of small fish species. Environmental Science and Technology 47, 1091-1100.

Reddy, A.N., Venugopal, N.B., Reddy, S.L., 1992. Effect of endosulfan 35 EC on ATPases in the tissues of a freshwater field crab Barytelphusa guerini. Bull Environ Contam Toxicol 48, 216-222.

Reddy, P.R., Nagaraju, G.P.C., Reddy, P.S., 2004. Involvement of Methyl Farnesoate in the Regulation of Molting and Reproduction in the Freshwater Crab Oziotelphusa senex senex. Journal of Crustacean Biology 24, 511-515.

Regier, J.C., Shultz, J.W., Kambic, R.E., 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proceedings of the Royal Society B: Biological Sciences 272, 395-401.

Richardson, N., Gordon, A.K., Muller, W.J., Whitfield, A.K., 2011. A weight-of-evidence approach to determine estuarine fish health using indicators from multiple levels of biological organization. Aquatic Conservation: Marine and Freshwater Ecosystems 21, 423-432.

Righetti, P.G., Boschetti, E., 2013. Chapter 1 - Introducing Low-Abundance Species in Proteome Analysis, Low-Abundance Proteome Discovery. Elsevier, Boston, pp. 1-11.

Roat, T.C., dos Santos-Pinto, J.R.A., dos Santos, L.D., Santos, K.S., Malaspina, O., Palma, M.S., 2014. Modification of the brain proteome of Africanized honeybees (Apis mellifera) exposed to a sub-lethal doses of the insecticide fipronil. Ecotoxicology 23, 1659-1670.

Rocha, J., Garcia-Carreño, F.L., Muhlia-Almazán, A., Peregrino-Uriarte, A.B., Yépiz-Plascencia, G., Córdova-Murueta, J.H., 2012. Cuticular chitin synthase and chitinase mRNA of whiteleg shrimp Litopenaeus vannamei during the molting cycle. Aquaculture 330–333, 111-115.

Rodriguez, E.M., Medesani, D.A., Fingerman, M., 2007. Endocrine disruption in crustaceans due to pollutants: a review. Comp Biochem Physiol A Mol Integr Physiol 146, 661-671.

Rodríguez, E.M., Medesani, D.A., Fingerman, M., 2007. Endocrine disruption in crustaceans due to pollutants: A review. Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology 146, 661-671.

Roland, K., Kestemont, P., Dieu, M., Raes, M., Silvestre, F., 2016. Using a novel "Integrated Biomarker Proteomic" index to assess the effects of freshwater pollutants in European eel peripheral blood mononuclear cells. Journal of Proteomics 137, 83-96.

Roland, K., Kestemont, P., Loos, R., Tavazzi, S., Paracchini, B., Belpaire, C., Dieu, M., Raes, M., Silvestre, F., 2014. Looking for protein expression signatures in European eel peripheral blood mononuclear cells after in vivo exposure to perfluorooctane sulfonate and a real world field study. Science of the Total Environment 468-469, 958-967.

Rougemont, B., Bontemps Gallo, S., Ayciriex, S., Carrière, R., Hondermarck, H., Lacroix, J.M., Le Blanc, J.C.Y., Lemoine, J., 2017. Scout-MRM: Multiplexed Targeted Mass Spectrometry-Based Assay without Retention Time Scheduling Exemplified by Dickeya dadantii Proteomic Analysis during Plant Infection. Analytical chemistry 89, 1421-1426.

Russom, C.L., LaLone, C.A., Villeneuve, D.L., Ankley, G.T., 2014. Development of an adverse outcome pathway for acetylcholinesterase inhibition leading to acute mortality. Environ Toxicol Chem 33, 2157-2169.

Sagi, A., Silkovsy, J., Fleisher-Berkovich, S., Danon, A., Chayoth, R., 1995. Prostaglandin E<inf>2</inf> in Previtellogeic Ovaries of the Prawn Macrobrachium rosenbergii: Synthesis and Effect on the Level of cAMP. General and Comparative Endocrinology 100, 308-313.

Sarkar, A., Ray, D., Shrivastava, A.N., Sarker, S., 2006. Molecular Biomarkers: Their significance and application in marine pollution monitoring. Ecotoxicology 15, 333-340.

Schmidt, W., Rainville, L.-C., McEneff, G., Sheehan, D., Quinn, B., 2014. A proteomic evaluation of the effects of the pharmaceuticals diclofenac and gemfibrozil on marine mussels (Mytilus spp.): evidence for chronic sublethal effects on stress-response proteins. Drug testing and analysis 6, 210-219.

Scott, A.P., 2012. Do mollusks use vertebrate sex steroids as reproductive hormones? Part I: Critical appraisal of the evidence for the presence, biosynthesis and uptake of steroids. Steroids 77, 1450-1468.

Scott, A.P., 2013. Do mollusks use vertebrate sex steroids as reproductive hormones? II. Critical review of the evidence that steroids have biological effects. Steroids 78, 268-281.

Scott, A.P., Sanders, M., Stentiford, G.D., Reese, R.A., Katsiadaki, I., 2007. Evidence for estrogenic endocrine disruption in an offshore flatfish, the dab (Limanda limanda L.). Mar Environ Res 64, 128-148.

Serafim, A., Lopes, B., Company, R., Cravo, A., Gomes, T., Sousa, V., Bebianno, M.J., 2011. A multi-biomarker approach in cross-transplanted mussels Mytilus galloprovincialis. Ecotoxicology 20, 1959-1974.

Shariati, F., Shariati, S., 2011. Review on Methods for Determination of Metallothioneins in Aquatic Organisms. Biological Trace Element Research 141, 340-366.

Shen, H., Hu, Y., Ma, Y., Zhou, X., Xu, Z., Shui, Y., Li, C., Xu, P., Sun, X., 2014. In-depth transcriptome analysis of the red swamp crayfish Procambarus clarkii. PloS one 9, e110548-e110548.

Shepard, J.L., Olsson, B., Tedengren, M., Bradley, B.P., 2000. Protein expression signatures identified in Mytilus edulis exposed to PCBs, copper and salinity stress. Marine environmental research 50, 337-340.

Shi, X., Zhang, S., Pang, Q., 2006. Vitellogenin is a novel player in defense reactions. Fish & Shellfish Immunology 20, 769-772.

Short, S., Yang, G., Guler, Y., Green Etxabe, A., Kille, P., Ford, A.T., 2014a. Crustacean Intersexuality Is Feminization without Demasculinization: Implications for Environmental Toxicology. Environmental Science & Technology 48, 13520-13529.

Short, S., Yang, G., Kille, P., Ford, A.T., 2014b. Vitellogenin is not an appropriate biomarker of feminisation in a crustacean. Aquatic toxicology (Amsterdam, Netherlands) 153, 89-97.

Silvestre, F., Dierick, J.F., Dumont, V., Dieu, M., Raes, M., Devos, P., 2006. Differential protein expression profiles in anterior gills of Eriocheir sinensis during acclimation to cadmium. Aquat Toxicol 76, 46-58.

Simmons, D.B.D., Benskin, J.P., Cosgrove, J.R., Duncker, B.P., Ekman, D.R., Martyniuk, C.J., Sherry, J.P., 2015. Omics for aquatic ecotoxicology: Control of extraneous variability to enhance the analysis of environmental effects. Environmental Toxicology And Chemistry 34, 1693-1704.

Simmons, D.B.D., Sherry, J.P., 2016. Plasma proteome profiles of White Sucker (Catostomus commersonii) from the Athabasca River within the oil sands deposit. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 19, 181-189.

Simon, R., Jubeaux, G., Chaumot, A., Lemoine, J.r.m., Geffard, O., Salvador, A., 2010. Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: Application to vitellogenin in Crustacea (Gammarus fossarum). Journal of Chromatography A 1217, 5109-5115.

Sin, Y.W., Kenny, N.J., Qu, Z., Chan, K.W., Chan, K.W.S., Cheong, S.P.S., Leung, R.W.T., Chan, T.F., Bendena, W.G., Chu, K.H., Tobe, S.S., Hui, J.H.L., 2014. Identification of putative ecdysteroid and juvenile hormone pathway genes in the shrimp Neocaridina denticulata. General and comparative endocrinology.

Smit, M.G.D., Bechmann, R.K., Hendriks, A.J., Skadsheim, A., Larsen, B.K., Baussant, T., Bamber, S., Sanni, S., 2009. Relating biomarkers to whole-organism effects using species sensitivity distributions: A pilot study for marine species exposed to oil. Environmental Toxicology And Chemistry 28, 1104-1109.

Soin, T., Smagghe, G., 2007. Endocrine disruption in aquatic insects: A review. Ecotoxicology 16, 83-93.

Son, J., Lee, S.-E., Park, B.-S., Jung, J., Park, H.S., Bang, J.Y., Kang, G.-Y., Cho, K., 2011. Biomarker discovery and proteomic evaluation of cadmium toxicity on a collembolan species, Paronychiurus kimi (Lee). Proteomics 11, 2294-2307.

Song, Y., Villeneuve, D.L., Toyota, K., Iguchi, T., Tollefsen, K.E., 2017. Ecdysone Receptor Agonism Leading to Lethal Molting Disruption in Arthropods: Review and Adverse Outcome Pathway Development. Environmental Science and Technology 51, 4142-4157.

Sornom, P., Felten, V., Médoc, V., Sroda, S., Rousselle, P., Beisel, J.-N., 2010. Effect of gender on physiological and behavioural responses of Gammarus roeseli (Crustacea Amphipoda) to salinity and temperature. Environmental Pollution 158, 1288-1295.

Spaziani, E.P., Hinsch, G.W., Edwards, S.C., 1995. The effect of prostaglandin E<inf>2</inf> and prostaglandin F<inf>2α</inf> on ovarian tissue in the Florida crayfish P rocambarus paeninsulanus. Prostaglandins 50, 189-200.

Sroda, S., Cossu-Leguille, C., 2011. Effects of sublethal copper exposure on two gammarid species: Which is the best competitor? Ecotoxicology 20, 264-273.

Stay, B., Tobe, S.S., 2007. The role of allatostatins in juvenile hormone synthesis in insects and crustaceans. Annu Rev Entomol 52, 277-299.

Subramoniam, T., 2000. Crustacean ecdysteriods in reproduction and embryogenesis. Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology 125, 135-156.

Sumpownon, C., Engsusophon, A., Siangcham, T., Sugiyama, E., Soonklang, N., Meeratana, P., Wanichanon, C., Hanna, P.J., Setou, M., Sobhon, P., 2015. Variation of prostaglandin E<inf>2</inf> concentrations in ovaries and its effects on ovarian maturation and oocyte proliferation in the giant fresh water prawn, Macrobrachium rosenbergii. General and Comparative Endocrinology 223, 129-138.

Sumpter, J.P., Jobling, S., 1995. Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. Environ Health Perspect 7, 173-178.

Tamone, S.L., Chang, E.S., 1993. Methyl farnesoate stimulates ecdysteroid secretion from crab Y-organs in vitro. Gen Comp Endocrinol 89, 425-432.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., Matsuo, T., 1988. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 2, 151-153.

Tao, T., Xie, X., Liu, M., Jiang, Q., Zhu, D., 2017. Cloning of two carboxylesterase cDNAs from the swimming crab Portunus trituberculatus: Molecular evidences for their putative roles in methyl farnesotae degradation. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 203, 100-107.

Thompson, E.L., Taylor, D.A., Nair, S.V., Birch, G., Hose, G.C., Raftos, D.A., 2012. Proteomic analysis of Sydney Rock oysters (Saccostrea glomerata) exposed to metal contamination in the field. Environmental Pollution 170, 102-112.

Timbrell, J.A., 1998. Biomarkers in toxicology. Toxicology 129, 1-12.

Tiu, S.H.K., Hult, E.F., Yagi, K.J., Tobe, S.S., 2012. Farnesoic acid and methyl farnesoate production during lobster reproduction: Possible functional correlation with retinoid X receptor expression. General and comparative endocrinology 175, 259-269.

Tkatcheva, V., Poirier, D., Chong-Kit, R., Furdui, V.I., Burr, C., Leger, R., Parmar, J., Switzer, T., Maedler, S., Reiner, E.J., Sherry, J.P., Simmons, D.B.D., 2015. Lithium an emerging contaminant: Bioavailability, effects on protein expression, and homeostasis disruption in short-term exposure of rainbow trout. Aquatic Toxicology 161, 85-93.

Trapp, J., Armengaud, J., Gaillard, J.C., Pible, O., Chaumot, A., Geffard, O., 2016. High-throughput proteome dynamics for discovery of key proteins in sentinel species: Unsuspected vitellogenins diversity in the crustacean Gammarus fossarum. J Proteomics 146, 207-214.

Trapp, J., Armengaud, J., Pible, O., Gaillard, J.C., Abbaci, K., Habtoul, Y., Chaumot, A., Geffard, O., 2015. Proteomic investigation of male Gammarus fossarum, a freshwater crustacean, in response to endocrine disruptors. Journal of proteome research 14, 292-303.

Trapp, J., Armengaud, J., Salvador, A., Chaumot, A., Geffard, O., 2014a. Next-generation proteomics: Toward customized biomarkers for environmental biomonitoring. Environmental Science and Technology 48, 13560-13572.

Trapp, J., Geffard, O., Imbert, G., Gaillard, J.C., Davin, A.H., Chaumot, A., Armengaud, J., 2014b. Proteogenomics of gammarus fossarum to document the reproductive system of amphipods. Molecular and Cellular Proteomics 13, 3612-3625.

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., Minh, B.Q., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44, W232-W235.

Truebano, M., Tills, O., Spicer, J.I., 2016. Embryonic transcriptome of the brackishwater amphipod Gammarus chevreuxi. Marine genomics 28, 5-6.

Turja, R., Guimarães, L., Nevala, A., Kankaanpää, H., Korpinen, S., Lehtonen, K.K., 2014. Cumulative effects of exposure to cyanobacteria bloom extracts and benzo[a]pyrene on antioxidant defence biomarkers in Gammarus oceanicus (Crustacea: Amphipoda). Toxicon : official journal of the International Society on Toxinology 78, 68-77.

Tyler, C.R., Jobling, S., Sumpter, J.P., 1998. Endocrine Disruption in Wildlife: A Critical Review of the Evidence. Critical Reviews in Toxicology 28, 319-361.

Tyler, C.R., van der Eerden, B., Jobling, S., Panter, G., Sumpter, J.P., 1996. Measurement of vitellogenin, a biomarker for exposure to oestrogenic chemicals, in a wide variety of cyprinid fish. Journal of Comparative Physiology B 166, 418-426.

Vellinger, C., Sohm, B., Parant, M., Immel, F., Usseglio-Polatera, P., 2016. Investigating the emerging role of comparative proteomics in the search for new biomarkers of metal contamination under varying abiotic conditions. Science of the Total Environment.

Verhaegen, Y., Parmentier, K., Swevers, L., Renders, E., Rougé, P., De Coen, W., Cooreman, K., Smagghe, G., 2011. The heterodimeric ecdysteroid receptor complex in the brown shrimp Crangon crangon: EcR and RXR isoform characteristics and sensitivity towards the marine pollutant tributyltin. General and comparative endocrinology 172, 158-169.

Vidova, V., Spacil, Z., 2017. A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. Analytica Chimica Acta 964, 7-23.

Wang, M., You, J.-S., Bemis, K.G., Fitzpatrick, D.P.G., 2008. Label-Free Mass Spectrometry-Based Protein Quantification Technologies in Protein Biomarker Discovery, in: Wang, F. (Ed.), Biomarker Methods in Drug Discovery and Development. Humana Press, Totowa, NJ, pp. 211-230.

Wang, Y.H., Wang, G., LeBlanc, G.A., 2007. Cloning and characterization of the retinoid X receptor from a primitive crustacean Daphnia magna. General and Comparative Endocrinology 150, 309-318.

Ward, D.G., Wei, W., Cheng, Y., Billingham, L.J., Martin, A., Johnson, P.J., Lyons, B.P., Feist, S.W., Stentiford, G.D., 2006. Plasma Proteome Analysis Reveals the Geographical Origin and Liver Tumor Status of Dab (Limanda limanda) from UK Marine Waters. Environmental science & technology 40, 4031-4036.

Wetmore, B.A., Merrick, B.A., 2004. Toxicoproteomics: proteomics applied to toxicology and pathology. Toxicol Pathol 32, 619-642.

Wheelock, C.E., Goss, V.M., Balgoma, D., Nicholas, B., Brandsma, J., Skipp, P.J., Snowden, S., Burg, D., D'Amico, A., Horvath, I., Chaiboonchoe, A., Ahmed, H., Ballereau, S., Rossios, C., Chung, K.F., Montuschi, P., Fowler, S.J., Adcock, I.M., Postle, A.D., Dahleń, S.E., Rowe, A., Sterk, P.J., Auffray, C., Djukanović, R., 2013. Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. European Respiratory Journal 42, 802-825.

Whyte, J.J., Jung, R.E., Schmitt, C.J., Tillitt, D.E., 2000. Ethoxyresorufin-O-deethylase (EROD) Activity in Fish as a Biomarker of Chemical Exposure. Critical Reviews in Toxicology 30, 347-570.

Wilson, K., Cahill, V., Ballment, E., Benzie, J., 2000. The complete sequence of the mitochondrial genome of the crustacean Penaeus monodon: Are malacostracan crustaceans more closely related to insects than to branchiopods? Molecular Biology and Evolution 17, 863-874.

Xuereb, B., 2009. Développement de marqueurs de neurotoxicité et de perturbations endocrines chez l'amphipode d'eau douce Gammarus fossarum.

Xuereb, B., Bezin, L., Chaumot, A., Budzinski, H., Augagneur, S., Tutundjian, R., Garric, J., Geffard, O., 2011. Vitellogenin-like gene expression in freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization in females and potential for use as an endocrine disruption biomarker in males. Ecotoxicology (London, England) 20, 1286-1299.

Xuereb, B., Lefèvre, E., Garric, J., Geffard, O., 2009a. Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda): Linking AChE inhibition and behavioural alteration. Aquatic Toxicology 94, 114-122.

Xuereb, B.t., Chaumot, A., Mons, R., Garric, J., Geffard, O., 2009b. Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda). Intrinsic variability, reference levels, and a reliable tool for field surveys. Aquatic Toxicology 93, 225-233.

Yang, C., Lin, Y., Liu, H., Shen, G., Luo, J., Zhang, H., Peng, Z., Chen, E., Xing, R., Han, C., Xia, Q., 2014. The Broad Complex isoform 2 (BrC-Z2) transcriptional factor plays a critical role in vitellogenin transcription in the silkworm Bombyx mori. Biochimica et Biophysica Acta (BBA) - General Subjects 1840, 2674-2684.

Yang, J., Liu, D., Dahms, H.U., Wang, L., 2015. Cadmium inhibits the vitellogenesis of freshwater crab Sinopotamon henanense. Environ Toxicol Chem 34, 1609-1616.

Yang, M., Wang, Y., Jiang, F., Song, T., Wang, H., Liu, Q., Zhang, J., Zhang, J., Kang, L., 2016. miR-71 and miR-263 Jointly Regulate Target Genes Chitin synthase and Chitinase to Control Locust Molting. PLoS Genetics 12.

Yeh, M.S., Lai, C.Y., Liu, C.H., Kuo, C.M., Cheng, W., 2009. A second proPO present in white shrimp Litopenaeus vannamei and expression of the proPOs during a Vibrio alginolyticus injection, molt stage, and oral sodium alginate ingestion. Fish and Shellfish Immunology 26, 49-55.

Yeom, D.H., Adams, S.M., 2007. Assessing effects of stress across levels of biological organization using an aquatic ecosystem health index. Ecotoxicology and Environmental Safety 67, 286-295.

Yuan, H.X., Xu, X., Sima, Y.H., Xu, S.Q., 2013. Reproductive toxicity effects of 4-nonylphenol with known endocrine disrupting effects and induction of vitellogenin gene expression in silkworm, Bombyx mori. Chemosphere 93, 263-268.

Zhang, S., Sun, Y., Pang, Q., Shi, X., 2005. Hemagglutinating and antibacterial activities of vitellogenin. Fish Shellfish Immunol 19, 93-95.

Zhang, X., Zheng, S., 2017. 20-hydroxyecdysone enhances the expression of the chitinase 5 via Broad-Complex Zinc-Finger 4 during metamorphosis in silkworm, Bombyx mori. Insect Molecular Biology 26, 243-253.

Zhou, J., Cai, Z.H., Li, L., Gao, Y.F., Hutchinson, T.H., 2010. A proteomics based approach to assessing the toxicity of bisphenol A and diallyl phthalate to the abalone (Haliotis diversicolor supertexta). Chemosphere 79, 595-604.

Zhu, J.-Y., Huang, H.-Q., Bao, X.-D., Lin, Q.-M., Cai, Z., 2006. Acute toxicity profile of cadmium revealed by proteomics in brain tissue of Paralichthys olivaceus: Potential role of transferrin in cadmium toxicity. Aquatic Toxicology 78, 127-135.

Abstract

The routine use of biomarkers in environmental biomonitoring faces several drawbacks, especially in invertebrates. Among these limitations, the lack of validated species-specific biomarkers and high-throughput multibiomarker methodologies constitute major constraints. Based on gene and protein catalogs obtained from proteogenomics experiments previously performed on the ecotoxicological-relevant species *Gammarus fossarum*, this thesis aimed at identifying and validating molecular biomarkers for the diagnostic of toxic perturbations in this species.
Following recent methodologies for the development of biomarkers in human disease diagnosis, we implemented a fast, specific, quantitative multiplexed targeted proteomics assay (using Selected Reaction Monitoring mass spectrometry) to study dozens of protein biomarker candidates simultaneously. The assay was applied to assess the physiological significance of protein candidate modulations, and to assess their pertinence as biomarkers after laboratory and field exposures to chemical contamination.
The second part of this thesis aimed at the development of specific endocrine disruption biomarkers, through two complementary approaches. The first approach, based on a comparative shotgun proteomic analysis using a known endocrine disruptor for arthropods, allowed exhaustively increasing the whole-proteome analysis through the detection of roughly 4000 proteins (53 modulated by the exposure). The second comprised sequence homology searches, phylogenetic analyses, and gene expression studies for proposing new biomarkers among candidates identified from literature searches.

Keywords: biomarkers, proteomics, gene candidate, endocrine disruption, *Gammarus fossarum*

Résumé

L'utilisation en routine de biomarqueurs pour la biosurveillance environnementale présente plusieurs limitations, en particulier chez les invertébrés, notamment le manque de biomarqueurs spécifiques et de méthodologies permettant l'acquisition haut débit de données multibiomarqueurs. En se basant sur les catalogues de gènes et protéines obtenus à partir des études proteogénomiques antérieures chez l'espèce clé en écotoxicologie *Gammarus fossarum*, cette thèse a visé l'identification et la validation de biomarqueurs moléculaires pour le diagnostic de perturbations toxiques chez ce crustacé.
S'appuyant sur des méthodologies récentes utilisées en santé humaine utilisant la spectrométrie de masse ciblée, nous avons mis en place un dosage multiplexé rapide et spécifique pour quantifier simultanément des dizaines de candidats biomarqueurs protéiques. L'essai a été appliqué pour mettre en lien modulation de la concentration des protéines suivies et processus physiologiques, ainsi que leur pertinence comme biomarqueurs lors d'expositions à des contaminants au laboratoire et sur le terrain.
Le deuxième axe de cette thèse a visé le développement de biomarqueurs spécifiques d'une perturbation endocrinienne via deux approches. La première approche, basée sur un étude de protéomique comparative de la réponse à un insecticide perturbateur endocrinien, a permis d'augmenter l'exhaustivité dans le suivi du protéome avec la détection d'environ 4000 protéines (dont 53 modulées par l'exposition). Une deuxième approche gène candidat a été mise en œuvre en s'appuyant sur des recherches d'homologies de séquences, des analyses phylogénétiques, et l'étude d'expression de trois gènes identifiés comme impliqué dans la régulation endocrine de la mue et de la reproduction.

Mots clés : biomarqueurs, protéomique, gène candidat, perturbation endocrinienne, *Gammarus fossarum*