High-multiplexed monitoring of protein biomarkers in the sentinel Gammarus fossarum by targeted scout-MRM assay, a new vision for ecotoxicoproteomics

Julien Faugère, Duarte Gouveia, Sophie Ayciriex, Arnaud Chaumot, Christine Almunia, Adeline François, Jean Armengaud, Jérôme Lemoine, Olivier Geffard, Davide Degli Esposti, et al.

High-multiplexed monitoring of protein biomarkers in the sentinel *Gammarus fossarum* by Targeted Scout-MRM assay, a new vision for ecotoxicoproteomics.

Julien Faugere[1]*, Duarte Gouveia[2,3]*, Sophie Ayciriex[1], Arnaud Chaumot[2], Christine Almunia[3], Adeline François[2], Jean Armengaud[3], Jérôme Lemoine[1], Olivier Geffard[2], Davide Degli-Esposti[2], Arnaud Salvador[1#]

* Equal contribution

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

[2,] INRAE, UR RiverLy, Laboratoire d'écotoxicologie, F-69625 Villeurbanne, France.

[3], Laboratoire Innovations Technologiques pour la Détection et le Diagnostic (LI2D), Service de Pharmacologie et Immunoanalyse (SPI), Commissariat à l'Energie Atomique, Bagnols-sur-Cèze, France.

[#]Corresponding Author; E-mail: arnaud.salvador@univ-lyon1.fr

Acknowledgments

The authors thank the French Ministry of Higher Education, Research and Innovation (Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation) for the PhD fellowship of Julien Faugere. This work was supported by the French National Agency Research (Young investigator Grant ANR-18-CE34-0008 and Grant ANR-18-CE34-0013), the "Office Français de la Biodiversité", and the Ecophyto National Action Plan. The authors also benefitted from the French GDR "Aquatic Ecotoxicology" framework which aims at fostering stimulating scientific discussions and collaborations for more integrative approaches. The authors are grateful to Véronique Gouy (INRAE), Matthieu Le Dréau (INRAE), Christine Combe (Water Department of the City of Lons Le Saunier) for their help in the definition of the field deployment study, and to Anaïs Valada-Mennumi for her help on the sample preparation and data analysis.

Declaration of Competing Interest

The authors declare the following financial interests: Sciex company and J.L. filled a patent describing Scout-MRM concept for targeted analysis by mass spectrometry.

Keywords: Mass spectrometry, Protein quantification, Scout-MRM, ecotoxicology, biomarker, multiplex approach.

Abstract

Ecotoxicoproteomics employs mass spectrometry-based approaches centered on proteins of sentinel organisms to assess for instance, chemical toxicity in fresh water. In this study, we combined proteogenomics experiments and a novel targeted proteomics approach free from retention time scheduling called Scout-MRM. This methodology will enable the measurement of simultaneously changes in the relative abundance of multiple proteins involved in key physiological processes and potentially impacted by contaminants in the freshwater sentinel *Gammarus fossarum*. The development and validation of the assay were performed to target 157 protein biomarkers of this non-model organism. We carefully chose and validated the transitions to monitor using conventional parameters (linearity, repeatability, LOD, LOQ). Finally, the potential of the methodology is illustrated by measuring 277-peptide-plex assay (831 transitions) in sentinel animals exposed *in natura* to different agricultural sites potentially exposed to pesticide contamination. Multivariate data analyses highlighted the modulation of several key proteins involved in feeding and molting. This multiplex-targeted proteomics assay paves the way for the discovery and the use of a large panel of novel protein biomarkers in emergent ecotoxicological models for environmental monitoring in the future.

Biological significance

The study contributed to the development of Scout-MRM for the high-throughput quantitation of a large panel of proteins in the *Gammarus fossarum*freshwater sentinel. Increasing the number of markers in ecotoxicoproteomics is of utmost interest to assess the impact of pollutants in freshwater organisms. The development and validation of the assay enabled the monitoring of a large panel of reporter peptides of exposed gammarids. To illustrate the applicability of the methodology, animals from different agricultural sites were analysed. The application of the assay highlighted the modulation of some biomarker proteins involved in key physiological pathways, such as molting, feeding and general stress response. Increasing multiplexing capabilities and field test will provide the development of diagnostic protein biomarkers for emergent ecotoxicological models in future environmental biomonitoring programs.

1 Introduction

Ecotoxicoproteomics is being increasingly used in environmental hazard identification, through the monitoring of protein expression in sentinel organisms exposed to environmental pollutants in both laboratory and field studies [1-7]. Dynamic changes in the molecular machinery of an organism subjected to a toxic stress are the starting points of its physiological response [8]. Molecular biomarkers are therefore able to provide us with early diagnostics of adverse effects in comparison with other higher-level endpoints such as reproductive impairments or other physiological biomarkers.

Among the different mass spectrometry (MS) based strategies available for protein analysis, shotgun proteomics using data-dependent acquisition (DDA) remains the most popular for proteome discovery in ecotoxicological models. In this acquisition mode, a fixed number of precursor ions is selected in the MS1 survey scan, followed by a sequential isolation and fragmentation of the *N* most intense precursors. This acquisition mode is adapted for high-throughput studies for protein discovery but presents some limitations regarding its low reproducibility and inability to identify and quantify low-abundant proteins. DDA-based studies have recently proved to be extremely useful for assessing the impact of several model pollutants in the molecular machinery of sentinel organisms from aquatic ecosystems, identifying exposure fingerprints that inform about the pollutants modes of action, and highlighting potential toxicity biomarkers [1-3, 7, 9-13]. Nevertheless, despite the increasing number of publications proposing new toxicity biomarkers, few are being considered for use in routine environmental biomonitoring. This is mainly due to the lack of high-throughput quantitative assays available for their verification and validation before its implementation in operational monitoring programs.

Alternatively, targeted acquisition methods such as Multiple Reaction Monitoring (MRM) or Parallel Reaction Monitoring (PRM), avoid the lack of reproducibility and limited quantitative power of DDA methods, by focusing the MS/MS scans on a subset of predetermined target peptides. These methods require *a priori* knowledge of the elution time windows of the targeted peptides and the precursor-product ion transitions obtained from a spectral library (subsequent to DDA analysis). MRM assays were proposed in recent years as a promising tool for specific multi-biomarker measurements in environmental biomonitoring [14-16]. However, targeted data acquisition experiments such as MRM or PRM exhibit some restraints. One of the limitations of this approach is the number of transitions monitored per

peptide restricted by the duty cycle to keep an acceptable signal-to-noise ratio. In addition, the development of large multiplexed assays becomes rather complex because of the RT reliance or unwanted retention time shift due to sample matrix effects. More recently, a new MRM-based targeted method, namely Scout-MRM has been proposed to increase the multiplexing capability and the robustness of classic targeted approaches [17-19]. Briefly, Scout-MRM is based on the successive monitoring of complex transition groups triggered by Scout peptide signals distributed along the chromatogram. This method is completely independent from RT and consequently of time scheduling, thereby increasing the multiplexing capability and facilitating the analytical transfer between laboratories.

Herein, we develop for the first time in aquatic ecotoxicology Scout-MRM for the high-throughput quantitation of a list of key proteins resulting from a proteogenomics study in *Gammarus fossarum*, an aquatic model organism used as sentinel to assess freshwater pollution [20, 21]. Increasing multiplexing capabilities is of great importance for the development of biomarkers in ecotoxicology since it allows monitoring a broader list of candidate biomarkers and validating a higher number of reliable surrogate peptide biomarkers for developing absolute quantification assays. If robust enough, the simultaneous quantification of hundreds of peptides and proteins also allows performing shotgun-like protein network and/or co-expression analysis, with the advantage of targeting a sub-proteome covering only the functions of interest. The development and validation of Scout-MRM assay to monitor a large panel of protein's reported peptides from the emergent ecotoxicological model *G. fossarum* is presented. We also demonstrate the interest of the methodology in ecotoxicological studies through an application with active biomonitoring in an agricultural pollution context, followed by a concise discussion around the advantages and innovations of this methodology for environmental monitoring.

2 Materials and Methods

2.1 Reagents and chemicals

Water, acetonitrile (ACN) and methanol (MeOH) were obtained from Fisher Scientific (LC-MS grade, Strasbourg, France). EDTA, triton X-100, iodoacetamide (IMA), dithiothreitol (DTT), formic acid (FA), sodium chloride, aprotinin, ammonium bicarbonate (AMBIC), leupeptin, trypsin (treated TCPK from bovine pancreas), ethylic ether and absolute ethanol were obtained from Sigma Aldrich (St Quentin-Fallavier, France). Isotopically labelled peptides containing either a C-terminal [$^{15}N_2$ and $^{13}C_6$] lysine or arginine were synthesized by Thermo Fisher Scientific (purity > 97 %) and stored at -20°C until use.

2.2 Collection and maintenance of *G. fossarum* organisms

Gammarids were collected by kick sampling from the Pollon river in France, and acclimatized to laboratory conditions, as previously described [15, 16, 22, 23]. This sampling site contains a gammarid population frequently used as a source of organisms for active biomonitoring studies by the laboratory of ecotoxicology in INRAE [23, 24]. Before experiments, organisms were kept for two weeks in 30L tanks continuously supplied with drilled groundwater, which was adjusted with osmotic water to the same conductivity and pH values as the sampling site. The temperature of the water was maintained at 12 ± 1°C, with 16/8h light/dark photoperiods. Organisms were fed *ad libitum* with water-conditioned alder leaves (*Alnus glutinosa*) and tubifex were added once a week.

2.3 Sample preparation for shotgun proteomics

Whole-body of five male and five female gammarids were disrupted in extraction buffer (50 mM Tris-Base, 100 mM NaCl, 1 mM EDTA, 0.1% v/v Triton X-100, 6 M Urea, protease inhibitor cocktail) with a Tissue ruptor device (Qiagen). Homogenates were centrifuged at 10,000g for 7 minutes, and the supernatant transferred to a new tube. 30 µL of homogenate were subjected to a short SDS-PAGE migration at 200 V (roughly 3 minutes). The whole protein content from each sample in the gel was cut and processed with Trypsin Gold (Promega) and 0.011% ProteaseMax surfactant (Promega) as previously described [25].

2.4 Sample preparation for Scout-MRM assay

Total protein content from gammarids was extracted according to previous published works [14, 26]. 10 µL of heavy peptides at 4 µg/mL and 10 µL of Scout peptides at 5 µg/mL were then added. After a SPE clean-up and evaporation, sample was resuspended with 90 µL of a H_2O/ACN mixture (90:10, *v/v*) with 0.5% FA, and centrifuged at 12 000 rpm for 5 min at RT, before the LC-MS/MS analysis [14, 26].

2.5 Standards (SIL peptides, Scout peptides) preparation

Isotopically enriched peptide stock solution was prepared from lyophilized peptides dissolved in water/ACN mixture (50:50, *v/v*) with 0.5% FA to obtain a 400 µg/mL stock solution. Scout peptides used in this study are stable isotope labelled peptides, selected to trigger the analysis of MRM transition groups with Scout-MRM. Solutions containing isotopically enriched peptides were prepared from stock solutions and diluted to obtain the desired concentrations. The list of the labelled peptides and Scout peptides are shown in supplementary information Table S1.

2.6 Nano LC-MS/MS analysis on high resolution mass spectrometry

Extracted and digested proteins were analysed through data-dependent acquisition mode on a Q-Exactive HF mass spectrometer (Thermo Fisher) coupled to an Ultimate 3000 LC system (Dionex-LC Packings). Tryptic peptides were first desalted on a reversed-phase PepMap 100 C_{18} µ-precolumn, and separated on a nanoscale PepMap 100 C_{18} nanoLC column (3 mm, 100 Å, 75 µm i.d. 50 cm, Thermo Fisher Scientific) with a 90 minute gradient of ACN with 0.1% formic acid, and a flow rate of 0.3 µL/min. Full MS were acquired from 350 to 1 800 *m/z* and the 20 most abundant precursor ions were selected for fragmentation with 10 s dynamic exclusion time. Only ions with +2 or +3 charges were selected for MS/MS analysis.

2.7 Liquid chromatography and targeted mass spectrometry

LC-MS/MS analysis was performed on an 1290 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. The LC separation of the 20 µL injected sample was carried out on an Xbridge C_{18} column (100 mm ×

2.1 mm, particle size 3.5 µm) with a symmetry C_{18} guard column (2.1 mm x 10 mm, particle size 3.5 µm) from Waters (Milford, MA, USA). Elution was performed at a flow rate of 300 µL/min with water containing 0.1% formic acid as eluent A and acetonitrile containing 0.1% formic acid as eluent B, employing an isocratic gradient from the beginning at 5% B to for 2 min, followed by a linear gradient from 5% B to 35% B in 36 min. Column was washed at 100% B for 5 min and re-equilibrated at 5% B for 5 min. The injection duty cycle was 48 min, considering the column equilibration time. Instrument control, data acquisition, and processing were performed using a modified Analyst 1.6.2 software®. For academic research, Scout-MRM provisional software patch is available on request from Sciex company (contact: yves.leblanc@sciex.com). MS analysis was performed in positive ionization mode using an ion spray voltage of 5500 V. The curtain gas flows and the nebulizer were set at 50 psi using nitrogen. The TurboV™ ion source was set at 550 °C with the auxiliary gas flow (nitrogen) set at 40 psi. The software Skyline v4.1 (MacCoss Lab Software, USA) was used to product a list of suitable SRM transitions. The mass in Q1 and Q3 as well as the collision energy (CE) and the declustering potential (DP) values were predicted using Skyline for endogenous peptides when the associated heavy peptides were not available. The duty cycle is set at 1.52 s in order to attain fifteen points per chromatographic peak for each MRM transition in one of the thirteen Scout groups. For peptides with corresponding associated heavy peptides, their mass parameters have been optimized by direct infusion into the mass spectrometer. From the initial set of candidates SRM transitions, three transitions by peptides were selected for the final assay. Details of parameters are reported in Table S2.

2.8 MS/MS spectra interpretation from GFOSS database

MS/MS spectra were assigned to peptide sequences by searching against a customized RNAseq-derived database GFOSS, which was previously published [27]. This database contains 1,311,444 putative protein sequences and 289,084,257 residues. The algorithm from the MASCOT Daemon v2.3.2 search engine (Matrix Science) was used for database search and spectral matching. The following parameters were used: 5 ppm peptide tolerance and 0.02 Da MS/MS fragment tolerance, +2 or +3 peptide charge, a maximum of two missed cleavages, carbamidomethylation of cysteine as fixed modification, oxidation of methionine as variable modification, and trypsin as proteolytic enzyme. Mascot results were parsed with IRMa 1.31.1c software [28]. Peptide-spectrum matches presenting a MASCOT peptide score with a p-value of <0.05 were filtered and assigned to a protein according to the parsimony

principle. The data have been uploaded to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD017017 and 10.6019/PXD017017 [29].

2.9 Candidate list for Scout-MRM assay development

From the peptide and protein lists obtained in the DDA mode, peptides that did not match the requirements needed for the targeted approach were filtered out of the final list of candidates. At the precursor level, the parameters chosen for selection were the following: idotp >0.92 (isotope dot product), mass error ±4ppm, ≥4 peptides per protein. Only the four most intense precursors (based on MS peak area) per protein were kept. At the MS/MS level, all product ions with *m/z* higher than 1 050 were eliminated, and a maximum of 4 product ions per precursor were kept (the most intense ones). The list was further reduced by removing proteins with whose functional annotations suggested little interest to be used as toxicity biomarkers (mostly housekeeping and orphan proteins). A final list of 2,942 transitions, 919 peptides, and 263 proteins was obtained (Table S3).

2.10 Method validation for Scout-MRM assay

Verification of MRM transitions free from interferences and the correlation of peptide ratios were performed using a correlation study of transition ratios per peptide. The correlation study was carried out on 46 male and 24 female gammarids at different stages of reproduction. A peptide measured with three MRM transitions is considered as non-interfered if at least one of the transition ratios has a RSD < 20%. The remaining analytical parameters were performed from a mixture of male and female *G. fossarum* in order to have samples containing the same amount of protein digest as background matrix. The repeatability of the method including SPE, evaporation under nitrogen flow and MS measurement were evaluated by performing analyses of independent *G. fossarum* samples (n=3) spiked with 500 ng/mL of heavy peptides. The repeatability of the entire analytical protocol including the digestion step for endogenous peptides was evaluated by analysing independent samples (n=10) from the same gammarid pool. Matrix effects, SPE recoveries and evaporation recoveries were evaluated by preparing *G. fossarum* sample spiked with heavy peptides at different protocol steps at a concentration of 500 ng/mL (n=3). The different recoveries (after SPE and evaporation) and matrix effect are calculated according to the following formulas:

SPE recovery (%) = (Heavy peptide area add before SPE / Heavy peptide area add after SPE) x 100

Evaporation recovery (%) = (Heavy peptide area add after SPE / Heavy peptide area add after evaporation) x 100

Matrix effect (%) = ((Heavy peptide area in matrix / Heavy peptide area add in solvent)-1) x 100

In order to carry out the linearity study of heavy peptides, as well as the evaluation of detection limits (LOD) and quantification limits (LOQ), 9 pools of *G. fossarum* peptide solutions were spiked with an increasing quantity of heavy peptides from 1 to 10 000 ng/mL. Each calibration curve has at least 5 levels of concentration (between 1 and 10000 ng/mL, 10 levels of concentrations were investigated) with an accuracy between 80 and 120%. In addition, a weighted least-square linear regression was used ($1/x^2$). To determine the LOD, we measured the signal-to-noise ratio in the detection area of the chromatographic peak and considered the 3:1 ratio as acceptable to define it. The LOQ is defined as being equal to three times the LOD. Protein extraction and digestion kinetics in *G. fossarum* have already been optimized and described in a previous publication [14].

2.11 **Data analysis**

The integration of chromatographic peaks and data reprocessing was performed with MultiQuant™ software (version 2.1.1, Sciex). Peak areas for each peptide were log2 transformed to perform multivariate analyses. In order to identify possible outliers, sample clustering was performed using the hierarchical clustering function implemented in the lumi R package [30]. Principal Component Analysis (PCA) was used to analyse and identify the variables explaining the maximum variance associated to the proteomic data in male gammarids caged in the contaminated or the reference sites. Differential protein abundance analysis was performed using linear models and empirical Bayes methods implemented in the limma R package [31, 32]. Peptides showing differences with a BH adjusted p-value (FDR) <0.1 are considered significantly different. Targeted proteomics data have been uploaded to the PeptideAtlas SRM Experiment Library (PASSEL) under dataset identifier PASS01501.

2.12 **Application of the assay**

2.12.1 Selection and exposure of organisms

For each experiment, organisms were collected from the water tanks at specific reproductive stages. For method validation, mature male and female organisms were sampled based on

visual observation of couples having a female in an advanced stage of the reproductive cycle (well-developed embryos in the brood pouch as described in [20]). For the field studies, couples of *G. fossarum* were sorted in with female in the final molting stages D2 and placed by 7 in punctured polypropylene cylinders with alder leaves as food supply. The day after, four cages of 7 couples were deployed in each study site during three weeks following the reproductive bioassay protocol described in [33]. The organisms were then brought back to the laboratory in water from the study site. Seven males by sites were weighted and directly frozen in liquid nitrogen and stored at -80°C for the proteomic analysis.

2.12.2 Study sites

Organisms were caged in four distinct study sites in the "Jura" region in France for a three week exposure. Three sites (A, I, Av) were localized along the river « La Madeleine » (N 46°42'15"; E 5°31'9") which drains a mixed agricultural watershed (wine, corn and wheat crops) before joining drinking water catchments of the City of Lons-le-Saunier. According to the pesticide program commissioned by the water authorities of Lons-le-Saunier between 2011 and 2017, this watershed is at risk for pesticide contamination. A fourth study site situated upstream of the "Seille" river (N 46°42'58"; E 5°37'54") was used as a reference site. This station is located at less than 10 km of the east of the "Madeleine" watershed and is in a good chemical and ecological status among the National Reference Network implemented for the European Water Framework Directive.

3 Results and discussion

3.1 Selection of proteins of interest from proteogenomics and DDA analysis

In the shotgun experiment, an average of 46,068 MS/MS spectra were recorded per sample, from which 27-29% were attributed to peptide sequences after querying the GFOSS proteogenomics database. These peptide-to-spectrum matches (PSMs) allowed the identification of 5,298 protein sequences, which were functionally annotated through BLAST searches and used as the starting point for the selection of the candidates for the targeted assay. The *in silico* mining of this large list of proteins allowed to select a panel of 263 proteins according to the criteria described in section 2.9. These proteins are involved in key physiological functions such as reproduction, immunity, homeostasis, and detoxification / defence mechanisms, and are therefore susceptible to be disrupted by exposure to toxic compounds. The list of the proteins is shown in Table S3.

3.2 Reporter peptide identification and MS optimization in targeted MS approach

The list of 263 proteins was used to develop the Scout-MRM method on a triple quadrupole mass spectrometer. This type of acquisition method is more suitable for quantification because it provides more robustness. The specificity and sensitivity of targeted-based assays depend on the suitable selection of the proteotypic peptides. For the unambiguous peptide identification in low resolution targeted MS, four MRM transitions corresponding to the most intense MS/MS fragment ion peptides were selected from proteogenomic experiments, which correspond to 3, 676 transitions among 27 MRM methods. The chromatographic conditions of the proteogenomic study and the Scout-MRM method are different. Indeed, nano-LC configuration was not used for the development of Scout-MRM method. It was therefore necessary to redefine the retention times of each peptide. RT information was used as a reference for peptide detection in our system. A retention time was assigned to a peptide when the monitored transitions were perfectly aligned.

As shown in Fig. 1A, the extracted ion chromatograms showed four overlapped and identical MRM transitions for GTLAVIPVQNR and for their corresponding heavy peptide GTLAVIPVQNR*. Fig. 1B shows extracted ion chromatogram of LQQEQVADYK at 8.1

min with the 4 overlapped transitions that confirms the elution time of this peptide. In certain cases, difficulties can be encountered for the determination of the RT of a proxy peptide that exhibits 4 aligned transitions when an interference occurred and the corresponding labelled peptide is not available. As shown in Fig. 1C, the extracted ion chromatogram exhibits several peaks for which the 4 transitions for TDLSITLAER are perfectly aligned. To overcome this problem *i.e.* to correctly identify the RT area of the target peptides, we established a correlation curve between peptide RT obtained in targeted-MRM based assay and high-resolution-MS² based experiment (Fig. 2 and Table S4). Even if peptide separation conditions were very different (nanoLC *versus* microLC configuration, different stationary phase), the RT of the target peptides have been easily identified from other interfering isobaric compounds by using the formula of the regression curve. For 187 peptides, we had also to corroborate our results by performing additional MRM runs by adding supplemental transitions, *i.e.* 1879 transitions. Finally, after the optimization of the precursor and transition selection, 341 peptides proxy of 182 proteins have been perfectly identified on the Q-Trap system.

3.3 Improvement of multiplexing capacity

In a second step, we evaluated the capacity to configure a 182-plex MRM protein assay. This assay requires following 1,155 transitions including 44 heavy labelled peptides (up to 3 peptides per protein, at least 3 transitions per peptides). Considering the large number of transitions, it was not possible to develop a single MRM method. To solve this problem, an acquisition method capable of segmenting the chromatogram is needed to focus the mass spectrometer on specific areas associated with retention time of compounds of interest. For this purpose, we can use a single time fragmentation method (called Scheduled MRM®, Time MRM® or Dynamic MRM® according to the MS supplier).

3.3.1 Principle and limitations of MRM methods with time window.

To guarantee a constant acquisition setting of 10-15 points per chromatographic peak in time window scheduling experiment, MRM transitions of each peptide are centered on their expected retention times. A constant cycle time is maintained with the simultaneous adjustment of the dwell time according to the number of peptides co-eluted in the same window. To ensure highly multiplexed analysis, RT scheduling methods must monitor the transitions across small time segments. However, any unexpected RT shift out of the

scheduled RT window (modification in chromatographic conditions, change of volume delay with distinct pumping systems during a method transfer, injection overload of the column due to samples with different protein contents) can compromise assay robustness and its implementation for large-scale analysis. As shown in Fig. 3A, an artificial modification in HPLC conditions, *i.e.* decrease of the isocratic step at the beginning of the gradient, leads as expected to a RT shift of QFQYTWR peptide during retention time scheduling method. The chromatographic peak is truncated, and the quantification is difficult to achieve. In this case, reducing the RT window detection proves to be detrimental to the multiplexed analysis. In some cases, peptides may no longer be detected because they are outside the detection windows.

3.3.2 Principle and advantages of Scout-MRM

To avoid partial or no peak detection, we introduced a new mode of targeted data acquisition called Scout-MRM in order to rationalize the development of targeted proteomics assay and to facilitate dissemination of ready-to-use methods [18, 34]. Scout-MRM relies on the monitoring of complex transitions successively triggered under the detection of Scout peptides. Compared with scheduled methods, the acquisition was triggered by Scout peptides for each segment, instead of by pre-defined scheduled time windows (here in this case, segment of 1.5 min). No extra adjustment for acquisition windows is needed. As we observed, the peptide is correctly eluted during chromatographic variations (Fig. 3B). In practice, the Scout peptides chosen in this study are 12 stable isotope labelled peptides that can be used for relative quantification (Table S1). Scout peptides triggering groups of transitions (13 groups), are regularly dispatched all along the chromatogram (Fig. 4). When the intensity of the MRM transition of the first Scout exceeds a defined threshold, the monitoring of a transition group is triggered. The follow-up of the group stops when the second Scout peptide is detected and consequently the second group is triggered and so on [18, 34] (Fig. 4). Any incidental RT shift is completely without consequence on the target detection of peptides. To ensure highly multiplexed analysis, RT scheduling method monitors the transitions across large segments generally set to less than 1 min that ensures a compromise between multiplexing capacity and robustness of the method [34]. In Scout-MRM it is sufficient to distribute Scouts with an interval of less than one minute to have a higher gain in multiplexing capacity without loss of robustness. To increase multiplexing capacity for future analysis, it is therefore enough to increase the number of Scout groups in the method.

3.4 Scout-MRM method optimisation

3.4.1 Biological validation of MRM transitions

To ensure that the selected MRM transitions were not interfered in the different biological conditions, we verified that transition ratios from a specific peptide are constant. In order to be most relevant several samples need to be analyzed. From the samples available at the laboratory at the time of the study, forty-six adult male sampled from different locations and 24 adult female organisms at distinct reproductive stages were used to include as much variability as possible. MRM transitions were considered as non-interfered when the ratio between 3 transitions remains constant regardless the sample. From our data we have performed transition area ratio calculation for all samples and estimation of relative standard deviation values for each peptide. A threshold above 20% of MRM ratio is considered as interfered. As shown in Fig. 5, the peptide CQLFNDPSDR exhibits different transition ratios between y8/y7, ions with a RSD > 20%, in two different biological conditions (different male gammarids from different sites). On the contrary, the tryptic peptide SLVNLGDVQEGK conserves the transition ratio between the different ions y9/y8, y8/y7 and y9/y7 with a RSD < 20% [35]. After evaluation of all the MRM ratio transitions (341 peptides proxy for 182 proteins), we kept 277 peptides proxy for 157 proteins (Table S2). As a result, some proteins are only reported by one or two peptides in the method. More precisely, 67 proteins with 1 peptide, 42 proteins with 2 peptides and 30 proteins with 3 peptides were followed in male gammarids. For female gammarids, 76 proteins with 1 peptide, 47 protein with 2 peptides and 25 proteins with 3 peptides per protein were monitored.

3.4.2 Biological validation of peptides

A second study between the different peptides of the same protein was carried out in male and female *G. fossarum*. Indeed, peptides from the same protein must have constant area ratios between the different samples analysed, as shown in Fig. 6. This ratio is calculated from the most intense and least interfering transition areas determined in the first correlation study. If the ratios are not respected, this may be due for example to the presence of a modification on one of the peptides of the protein or different matrix effects. Peptide ratios for the same protein are more uncertain than transition ratios of a peptide. Indeed, the different transitions of the same peptide will undergo the same matrix effects because they are eluted at the same retention time while the peptides of the same protein will have different retention times and

can therefore undergo different matrix effects between different samples. This criterion was therefore not used to remove a transition from the method. Rather, it has been used as an indicative value to know which peptides within the protein correlate best with each other. The results obtained for our correlation study between the different peptides of the same protein are summarized in Fig. 7. These results show that for most proteins we have a good correlation between the different peptides.

3.5 Analytical performance evaluation of the assay for quantification

As the multiplexed protein assay will be used to support future ecotoxicology studies, the analytical performances of the assay for quantification were assessed. With our scout-MRM method, an absolute or relative quantification can be considered. Absolute quantitation is achieved with incorporation for each peptide of isotopically labelled synthetic peptide internal standards. Disposing of all the peptides labelled with high purity results in substantial costs that are often incompatible with the financial means of ecotoxicological studies. In a first step, a relative quantification is performed. Only 44 labelled peptides have been synthesized and will be used for performance evaluation. Since relative quantification method is based on sample comparison, it must first be ensured that the analytical protocol is repeatable. Therefore, a repeatability study was carried out.

3.5.1 Repeatability

Intraday repeatability was assessed by spiking heavy peptides to a pool of gammarids at a concentration of 500 ng/mL. All the results are presented in Table S1 as supplementary data. The relative standard deviations (RSD) for this repeatability study is between 1 and 12% that shows good repeatability. However, since we do not have the counterparts marked for each peptide, we have carried out another repeatability study by analysing endogenous peptides from ten independent extractions to consider all the peptides. Furthermore, this study also takes into account the reproducibility of protein digestion into peptides. All the results are presented in Table S5 as supplementary data. Fig. 8 shows that 90.2% of MRM transitions have an RSD of less than 20%. This threshold represents the limit values that have been set to define whether a transition is repeatable. The RSD obtained for all previously selected qualifying MRM transitions are less than or equal to 20%. When the RSD is greater than 20%, this corresponds to the lowest MRM transitions that generate more difficult integrations and are more easily interfered. These results indicate that the digestion step and extraction step are controlled and reproducible.

3.5.2 Extraction recoveries

It was shown in a previous paper that peptides can be lost during evaporation [26]. Therefore, this step was also specifically evaluated in addition to the SPE extraction recovery. To estimate extraction recoveries, digested protein extracts from gammarids were spiked with labelled peptides before and after SPE, and after evaporation at a concentration of 500 ng/mL (n=3) and compared. The results are presented in Table S1. It can be observed that SPE recoveries are between 72 and 122% with an average of 96% for the 44 heavy peptides. Evaporation recovery range from 67 to 106% with an average of 94% for the 44 heavy peptides. These results showed a good extraction recovery.

3.5.3 Matrix effect evaluation

Matrices effects in MS correspond to the influence of coeluted compounds during their simultaneous introduction into the source due to ionization competition. This can result in an increase or a decrease of their intensity. To measure these matrix effects, sample extracts were spiked with heavy peptides added just before injection at a concentration of 500 ng/mL (n=3) which were compared to samples containing the same quantity of heavy peptides spiked in reconstitution solvent (n=3). The results are presented in Table S1. The 44 heavy peptides undergo matrix effects that induce a loss or gain of the signal between -67% and 51% with an average loss of 28%. As expected, matrix effects occur. However, the case presented here illustrates the most extreme case where a very complex mixture is compared to compounds in solution. The analysis of heavy peptides spiked at the same concentration in biologically different *G. fossarum* extracts showed that even if matrix effects affect the peptides analysed, these effects are comparable from one sample to another. Indeed, the results presented in Fig.S1. obtained from the data in table S6 show that for 93% of heavy peptides the RSD are less than 20%. To correct different matrix effects between different samples, internal standards (heavy peptides) similar to endogenous peptides must be used, which will undergo the same matrix effects as the compounds of interest. In our case, we do not have all the heavy peptides corresponding to the 277 peptides selected in the method. Our study aims to highlight trends by identifying potential biomarkers. Once the candidates have been found, it will be necessary to have their counterparts marked to carry out the most accurate quantification possible, considering the matrix effects. This correction of matrix effects will also allow to get peptide ratios for the same protein as precise as possible.

3.5.4 Linearity and limit of quantification (LOQ)

Although this study is based on a relative quantification, we have determined the linearity domains and LOQ of our heavy peptides in order to evaluate the performance of the method. A weighted least-square linear regression ($1/x^2$) was used for the calibration. All calibration curves corresponding to heavy peptides spiked into the *G. fossarum* matrix at different levels of concentration confirmed the high degree of linearity ($0.991 < r^2 < 0.999$) (Table S1 and Fig. S2). The LOD was determined to be 3 times greater than the background noise (S/N=3) and the LOQ is determined as 3 times the LOD. The LOQ are between 1.12 pmol/mL and 113 pmol/mL (Table S1). Finally, the method developed with Scout-MRM enabled to monitor 157 proteins in male and female species of *G. fossarum* following the validation criteria (157 proteins, 277 peptides, and 831 MRM transitions) among the targeted 263 proteins considered initially.

3.6 Application of Scout-MRM assay: comparative proteomic analysis using active biomonitoring in an agriculture watershed

We applied SCOUT-MRM method to assess its interest and feasibility in the context of active biomonitoring in freshwater streams. Twenty-eight calibrated male gammarids (7 from a non-contaminated reference site, 7 from each of the three contaminated sites) were used for the proteomics analysis. Two hundred sixty-five peptides were detected for each male organism.

To compare the global proteomes of the organisms caged in the contaminated sites or the reference site, we performed clustering and principal component analyses. The cluster analysis identified three organisms as potential outliers, based on their distance with the rest of the samples (Figure 9A). These outliers (an organism in the reference site, R3; and two organisms in two different contaminated sites, A CE3 and AV CE10) were excluded from the following analyses. The PCA analysis showed that organisms are quite dispersed, even though there is a tendency of the organisms caged in the reference site to cluster more closely than those clustered in the contaminated sites (Figure 9B).

In order to identify peptides whose abundance is able to distinguish the organisms caged in the different type of sites, and thus could be identified as potential biomarkers of exposure to agriculture pollutants, a differential analysis was performed. The differential analysis identified 10 peptides as differently expressed (FDR<0.1) (Figure 9C). The peptide DIDAAFLVGAMPR, annotated as cytosolic malate dehydrogenase or malic enzyme (ME)

was the most upregulated in organisms exposed in the contaminated sites (logFC=1.38, FDR=0.038) (Figure 9D). The Drosophila ME activity can be induced by the juvenile hormone (JH) by both a direct effect on the enzyme in the short term and the activation of its gene (Men) transcription [36]. Since JH analogs are among the most used insecticides in agriculture, the increased expression of a homolog of ME in *G. fossarum* might suggest a response of these non-target organisms to JH analogs contamination in the observed sites. Among the downregulated peptides, the most significant was QIDNPDYK (logFC=-0.86, FDR=7.9*10-5) (Figure 9D), belonging to a homolog of calreticulin. Calreticulin is a highly conserved endoplasmic reticulum protein of the lectin family. It is involved in osmoregulation, molecular chaperoning and immune response in crustaceans, with most data available from decapods models [37-39]. Calreticulin has been reported as a general stress biomarker in in vitro and vertebrate models [40] and its downregulation may thus suggest a stress effects in amphipods caged in the contaminated sites.

Among the downregulated peptides found in organisms caged in the contaminated site, we found 2 peptides belonging to an homolog of endochitinase (EAFDTVGR, logFC=-0.61, FDR=0.059; LVLGIPFYGR, logFC=-0.55, FDR=0.072) and 2 peptides belonging to a protein annotated as endoglucanase (or cellulase A). (SAMNVALIAFK logFC=-0.76, FDR=0.063; AADLGLDSTNNR, logFC=-1.28, FDR=0.06).Chitinases are key enzymes for successfully complete molting cycle in arthropods [41]. We have previously observed its decreased detection in *G. fossarum* exposed to contaminated sites [15]. Similarly, other crustacean chitinases have been reported as sensitive biomarker to insecticides and fungicides [41]. Cellulases are essential for digestion in *G. fossarum* and their decreased activity and/or abundances may affect its reproductive capacity [39]. Moreover, cellulase activity were reported to be very sensitive to insecticides and fungicides [39]. Overall, these results suggest that multiple protein biomarkers in *G. fossarum* are useful to detect a biological response to a contamination of agricultural origin. The limited number of modulated peptides in the contaminated sites might suggest a certain specificity of the impact of the chemicals present in the aquatic environment to which the gammarids were exposed. However, due to the limited number of investigated sites, it will be crucial to extend the use of these multiplexed protein biomarker approach to a larger spatial scale with contrasted contamination profiles.

4 Conclusion

A robust LC-MS/MS method for the simultaneous analysis of many potential protein biomarkers in the sentinel species *G. fossarum* has been successfully developed and validated. We applied a new acquisition mode called Scout-MRM recently developed in our laboratory to significantly increase the multiplexing capacity through the implementation of a 157-protein multiplex (277 peptides, 831 MRM transitions) in adult gammarids. Scout-MRM provides a more robust method than acquisition mode using time window scheduling. Scout-MRM is free from retention times, thus limiting the loss of information due to potential RT shifts. The RT independency of Scout-MRM opens up the perspective to build large multiplex by adding more scout peptides and MRM transitions if one wants to follow more proteins of interest in the future. Indeed, since *G. fossarum* genome is still not fully characterized, new discovered proteins can be implemented.

The first application of our method for biomonitoring key proteins to assess freshwater pollution from different agricultural sites demonstrated the potential value of this methodology in ecotoxicology studies. Indeed, the detection of protein reporter peptides with modulation in response to stress shows that the Scout-MRM method is a relevant method to detect biological responses due to contamination. Further studies will allow to identify reference values for the investigated peptides and eventually fine-tuning the choice of more specific reporter peptides of proteins related with different mode of actions involved in the adverse outcome observed in the ecotoxicological bioassays used in the field [42, 43]. In conclusion, Scout-MRM streamlines the development of targeted proteomics method in an ecotoxicology study and simplifies dissemination of ready-to-use assays as they are easily transferable from one laboratory to another.

In conclusion, this Scout-MRM method streamlines the development of targeted proteomics method in an ecotoxicology study and simplifies dissemination of ready-to-use assays as they are easily transferable from one laboratory to another.

Figure captions

Fig.1. Identification of endogenous peptides from *G. fossarum* by MRM and selection of transitions. (A) Detection of endogenous peptides when labelled peptides are available. Retention times and transition ratios must be identical. (B) Detection of endogenous peptides when isotopically enriched peptides are not available. The four MRM transitions must be aligned. (C) Detection and selection of peptide with an isobaric interference also having an alignment of the four MRM transitions. The RT area in MRM mode was predicted based on proteogenomics data.

Fig. 2. Correlation curve plotting the retention time of peptides identified by proteogenomics to the peptide retention time obtained in MRM. Retention times of 66 peptides obtained in nanoLC-MS/MS (reversed-phase PepMap 100 C18 µ-precolumn, nanoscale PepMap 100 C18 nanoLC column, 3 mm, 100 Å, 75 µm i.d. 50 cm and Q-Exactive HF) and LC-MS/MS (symmetry C18 guard column, 2.1 mm x 10 mm, particle size 3.5 µm, Xbridge C18 LC column, 2.1 mm, 10 cm, particle size 3.5 µm and Qtrap 5500) have been reported. (Table S4).

Fig. 3. Effects of RT shifts for targeted peptides monitored by Scheduled-MRM® *versus* Scout-MRM method. (A) When the chromatographic conditions change (decrease in the isocratic level at the beginning of the gradient) a RT shift occurs and induces an offsetting of the chromatographic peak from the detection window. (B) If any RT shift occurs in Scout-MRM method, scout peptides automatically realigns the detection window.

Fig. 4. Scout-MRM concept. This acquisition mode consists in the detection of a first compound called Scout. When the intensity of the Scout MRM transition exceeds a threshold defined by the operator, the monitoring of an unordered transition group as well as the monitoring of a new scout transition and so on until the last group is triggered by scout n. The triggering of a new group ends the follow-up of the previous one.

Fig. 5. Correlation study between 3 MRM transitions of the same peptide. (A) Non-constant transition ratio obtained for the 3 MRM transitions of CQLFNDPSDR peptide under two different biological conditions. (B) Constant transition ratio obtained for 3 MRM transitions the peptide SLVNLGDVQEGK under two different biological conditions. (C) RSD obtained for the mean transition ratios of CQLFNDPSDR and SLVNLGDVQEGK

559 peptides in 46 different male *G.fossarum* samples. Only the 3 transitions of the peptide
560 SLVNLGDVQEGK are considered as non-interfered (RSD < 20%)

561 **Fig. 6. Area ratio between 3 peptides of the same protein.** Peptide area ratio between
562 different peptides of the same protein, aspartate aminotransferase and cytosolic malate
563 dehydrogenase, in 46 different male *G. fossarum* samples. For the aspartate aminotransferase,
564 the correlation of the 3 peptides shows good results (RSD < 20%) and for the cytosolic malate
565 dehydrogenase protein, the peptide area ratios show a lower correlation (RSD > 40%).

566 **Fig. 7. Assessment of the correlation study of peptide ratios for the same protein.** For
567 each gender are represented the percentages of protein that have at least one peptide ratio with
568 the indicated correlation coefficient. The red and blue curves represent the cumulative protein
569 percentages in function of peptide area ratios

570 **Fig. 8. Repeatability of endogenous peptides.** Average intensity of the MRM transitions
571 presents in the Scout-MRM method according to their repeatability. Results obtained after
572 independent extraction from the same pool of male and female *G. fossarum* (n=10). The
573 repeatability validation criterion is set at 20% which corresponds to 90.2% of the MRM
574 transitions of the method. The entire protocol method is therefore well repeatable. When the
575 criterion is not satisfied, this corresponds to the least intense MRM transition not used as a
576 qualifying transition (more easily interferable and more difficult to integrate) (Table S5)

577 **Fig. 9. Comparative proteomic analysis using active biomonitoring in an agriculture**
578 **watershed**. Analyses were performed using log2 transformed peak areas for each peptide
579 identified in caged organisms. (A) Hierarchical clustering Samples over the red line show a
580 higher distance compared with the rest of the organisms, thus they were excluded from the
581 datasets. (B) Principal Component Analysis. Samples caged in the reference site (green) tend
582 to cluster together. (C) Heatmap of differentially expressed proteins in male gammarids caged
583 for three weeks in a reference site (R) and three sites under agriculture pressure (A, I, Av).
584 Colours represent abundance fold changes (orange over detection and blue down
585 regulation).(D) Boxplot representing the differential expression of the most deregulated
586 peptides (DIDAAFLVGAMPR; QIDNPDYK).

References

[1] A. Lafontaine, D. Baiwir, C. Joaquim-Justo, E. De Pauw, S. Lemoine, C. Boulange-Lecomte, J. Forget-Leray, J.P. Thome, E. Gismondi, Proteomic response of Macrobrachium rosenbergii hepatopancreas exposed to chlordecone: Identification of endocrine disruption biomarkers?, Ecotoxicol Environ Saf 141 (2017) 306-314.

[2] W. Schmidt, L.C. Rainville, G. McEneff, D. Sheehan, B. Quinn, 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 Test Anal 6(3) (2014) 210-9.

[3] J. Trapp, J. Armengaud, O. Pible, J.C. Gaillard, K. Abbaci, Y. Habtoul, A. Chaumot, O. Geffard, Proteomic investigation of male Gammarus fossarum, a freshwater crustacean, in response to endocrine disruptors, J Proteome Res 14(1) (2015) 292-303.

[4] V.A. Dowling, D. Sheehan, Proteomics as a route to identification of toxicity targets in environmental toxicology, Proteomics 6(20) (2006) 5597-604.

[5] M.F. Lemos, A.C. Esteves, B. Samyn, I. Timperman, J. van Beeumen, A. Correia, C.A. van Gestel, A.M. Soares, Protein differential expression induced by endocrine disrupting compounds in a terrestrial isopod, Chemosphere 79(5) (2010) 570-6.

[6] D. Gouveia, C. Almunia, Y. Cogne, O. Pible, D. Degli-Esposti, A. Salvador, S. Cristobal, D. Sheehan, A. Chaumot, O. Geffard, J. Armengaud, Ecotoxicoproteomics: A decade of progress in our understanding of anthropogenic impact on the environment, J Proteomics 198 (2019) 66-77.

[7] F. Silvestre, J.F. Dierick, V. Dumont, M. Dieu, M. Raes, P. Devos, Differential protein expression profiles in anterior gills of Eriocheir sinensis during acclimation to cadmium, Aquat Toxicol 76(1) (2006) 46-58.

[8] T.E. Allen, J.M. Goodman, S. Gutsell, P.J. Russell, Defining molecular initiating events in the adverse outcome pathway framework for risk assessment, Chem Res Toxicol 27(12) (2014) 2100-12.

[9] M. Borgatta, C. Hernandez, L.A. Decosterd, N. Chevre, P. Waridel, Shotgun ecotoxicoproteomics of Daphnia pulex: biochemical effects of the anticancer drug tamoxifen, J Proteome Res 14(1) (2015) 279-91.

[10] E. Gismondi, J.P. Thome, N. Urien, E. Uher, D. Baiwir, G. Mazzucchelli, E. De Pauw, L.C. Fechner, J.D. Lebrun, Ecotoxicoproteomic assessment of the functional alterations caused by chronic metallic exposures in gammarids, Environ Pollut 225 (2017) 428-438.

[11] W. Hu, S. Tedesco, B. McDonagh, D. Sheehan, Shotgun redox proteomics in sub-proteomes trapped on functionalised beads: Identification of proteins targeted by oxidative stress, Mar Environ Res 69 Suppl (2010) S25-7.

[12] M. Lepretre, C. Almunia, J. Armengaud, A. Salvador, A. Geffard, M. Palos-Ladeiro, The immune system of the freshwater zebra mussel, Dreissena polymorpha, decrypted by proteogenomics of hemocytes and plasma compartments, J Proteomics 202 (2019) 103366.

[13] C.J. Martyniuk, K.J. Kroll, N.J. Doperalski, D.S. Barber, N.D. Denslow, Environmentally relevant exposure to 17alpha-ethinylestradiol affects the telencephalic proteome of male fathead minnows, Aquat Toxicol 98(4) (2010) 344-53.

[14] A. Charnot, D. Gouveia, J. Armengaud, C. Almunia, A. Chaumot, J. Lemoine, O. Geffard, A. Salvador, Multiplexed assay for protein quantitation in the invertebrate Gammarus fossarum by liquid chromatography coupled to tandem mass spectrometry, Anal Bioanal Chem 409(16) (2017) 3969-3991.

[15] D. Gouveia, A. Chaumot, A. Charnot, C. Almunia, A. Francois, L. Navarro, J. Armengaud, A. Salvador, O. Geffard, Ecotoxico-Proteomics for Aquatic Environmental Monitoring: First in Situ Application of a New Proteomics-Based Multibiomarker Assay Using Caged Amphipods, Environ Sci Technol 51(22) (2017) 13417-13426.

[16] G. Jubeaux, R. Simon, A. Salvador, C. Lopes, E. Lacaze, H. Queau, A. Chaumot, O. Geffard, Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: inconclusive experience, Aquat Toxicol 122-123 (2012) 9-18.

[17] S. Ayciriex, R. Carriere, C. Bardet, J. Blanc, A. Salvador, T. Fortin, J. Lemoine, Streamlined Development of Targeted Mass Spectrometry-Based Method Combining Scout-MRM and a Web-Based Tool Indexed with Scout Peptides, Proteomics 20(2) (2020) e1900254.

[18] B. Rougemont, S. Bontemps Gallo, S. Ayciriex, R. Carriere, H. Hondermarck, J.M. Lacroix, J.C. Le Blanc, J. Lemoine, Scout-MRM: Multiplexed Targeted Mass Spectrometry-Based Assay without Retention Time Scheduling Exemplified by Dickeya dadantii Proteomic Analysis during Plant Infection, Anal Chem 89(3) (2017) 1421-1426.

[19] A. Salvador, R. Carriere, S. Ayciriex, C. Margoum, Y. Leblanc, J. Lemoine, Scout-multiple reaction monitoring: A liquid chromatography tandem mass spectrometry approach for multi-residue pesticide analysis without time scheduling, J Chromatogr A 1621 (2020) 461046.

[20] O. Geffard, B. Xuereb, A. Chaumot, A. Geffard, S. Biagianti, C. Noel, K. Abbaci, J. Garric, G. Charmantier, M. Charmantier-Daures, Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment, Environ Toxicol Chem 29(10) (2010) 2249-59.

[21] P.Y. Kunz, C. Kienle, A. Gerhardt, Gammarus spp. in aquatic ecotoxicology and water quality assessment: toward integrated multilevel tests, Rev Environ Contam Toxicol 205 (2010) 1-76.

[22] J.P. Besse, M. Coquery, C. Lopes, A. Chaumot, H. Budzinski, P. Labadie, O. Geffard, 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 Res 47(2) (2013) 650-60.

[23] A. Ciliberti, A. Chaumot, R. Recoura-Massaquant, A. Chandesris, A. Francois, M. Coquery, M. Ferreol, O. Geffard, Caged Gammarus as biomonitors identifying thresholds of toxic metal bioavailability that affect gammarid densities at the French national scale, Water Res 118 (2017) 131-140.

[24] B. Alric, O. Geffard, A. Chandesris, M. Ferreol, A. Francois, O. Perceval, J. Piffady, B. Villeneuve, A. Chaumot, Multisubstance Indicators Based on Caged Gammarus Bioaccumulation Reveal the Influence of Chemical Contamination on Stream Macroinvertebrate Abundances across France, Environ Sci Technol 53(10) (2019) 5906-5915.

[25] E.M. Hartmann, F. Allain, J.C. Gaillard, O. Pible, J. Armengaud, Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria, Methods Mol Biol 1197 (2014) 275-85.

[26] D.G. Aurore Charnot, Sophie Ayciriex, Jérôme Lemoine, Jean Armengaud, Christine Almunia, Arnaud Chaumot, Olivier Geffard, Arnaud Salvador, On-Line Solid Phase Extraction Liquid Chromatography-Mass Spectrometry Method for Multiplexed Proteins Quantitation in an Ecotoxicology Test Specie: Gammarus fossarum, Journal of Applied Bioanalysis 4(3) (2018).

[27] J. Trapp, O. Geffard, G. Imbert, J.C. Gaillard, A.H. Davin, A. Chaumot, J. Armengaud, Proteogenomics of Gammarus fossarum to document the reproductive system of amphipods, Mol Cell Proteomics 13(12) (2014) 3612-25.

[28] V. Dupierris, C. Masselon, M. Court, S. Kieffer-Jaquinod, C. Bruley, A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa, Bioinformatics 25(15) (2009) 1980-1.

[29] J.A. Vizcaino, R.G. Cote, A. Csordas, J.A. Dianes, A. Fabregat, J.M. Foster, J. Griss, E. Alpi, M. Birim, J. Contell, G. O'Kelly, A. Schoenegger, D. Ovelleiro, Y. Perez-Riverol, F. Reisinger, D. Rios, R. Wang, H. Hermjakob, The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013, Nucleic Acids Res 41(Database issue) (2013) D1063-9.

[30] D. Degli Esposti, C. Almunia, M.A. Guery, N. Koenig, J. Armengaud, A. Chaumot, O. Geffard, Co-expression network analysis identifies gonad- and embryo-associated protein modules in the sentinel species Gammarus fossarum, Sci Rep 9(1) (2019) 7862.

[31] K. Kammers, R.N. Cole, C. Tiengwe, I. Ruczinski, Detecting Significant Changes in Protein Abundance, EuPA Open Proteom 7 (2015) 11-19.

[32] G.K. Smyth, Linear models and empirical bayes methods for assessing differential expression in microarray experiments, Stat Appl Genet Mol Biol 3 (2004) Article3.

[33] C. Lopes, A. Chaumot, B. Xuereb, R. Coulaud, G. Jubeaux, H. Queau, A. Francois, O. Geffard, In situ reproductive bioassay with caged Gammarus fossarum (Crustacea): Part 2 – evaluating the relevance of using a molt cycle temperature-dependent model as reference to assess toxicity in freshwater monitoring Environmental Toxicology and Chemistry (2020).

[34] S. Ayciriex, R. Carriere, C. Bardet, J.C.Y. Blanc, A. Salvador, T. Fortin, J. Lemoine, Streamlined Development of Targeted Mass Spectrometry-Based Method Combining Scout-MRM and a Web-Based Tool Indexed with Scout Peptides, Proteomics 20(2) (2020) e1900254.

[35] A.J. Percy, A.G. Chambers, J. Yang, C.H. Borchers, Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma, Proteomics 13(14) (2013) 2202-15.

[36] R. Farkas, P. Danis, L. Medved'ova, B.M. Mechler, J. Knopp, Regulation of cytosolic malate dehydrogenase by juvenile hormone in Drosophila melanogaster, Cell Biochem Biophys 37(1) (2002) 37-52.

[37] Y. Duan, P. Liu, J. Li, Y. Wang, J. Li, P. Chen, Molecular responses of calreticulin gene to Vibrio anguillarum and WSSV challenge in the ridgetail white prawn Exopalaemon carinicauda, Fish Shellfish Immunol 36(1) (2014) 164-71.

[38] Y. Huang, K. Hui, M. Jin, S. Yin, W. Wang, Q. Ren, Two endoplasmic reticulum proteins (calnexin and calreticulin) are involved in innate immunity in Chinese mitten crab (Eriocheir sinensis), Sci Rep 6 (2016) 27578.

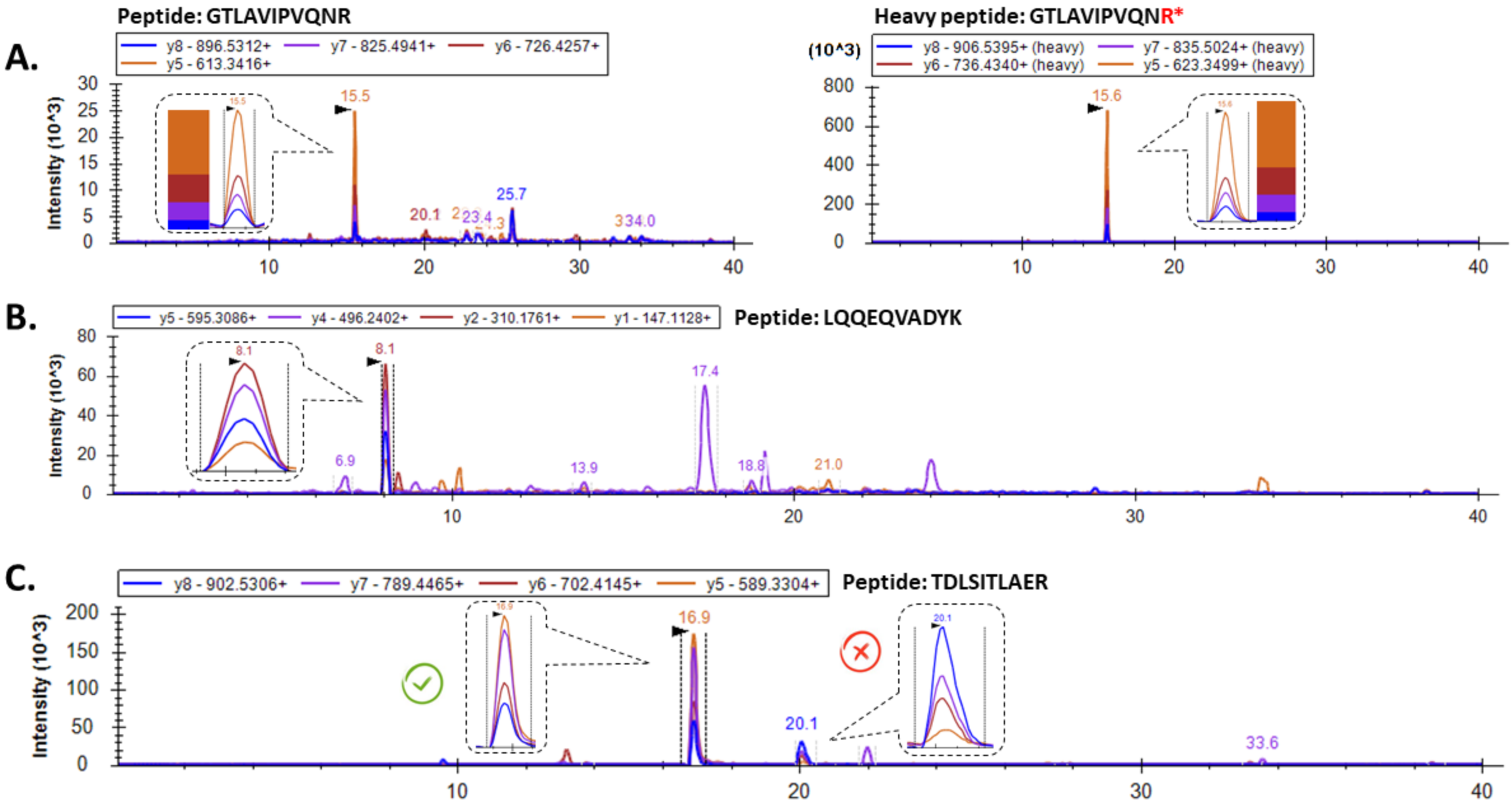
[39] A. Moshtaghi, M.L. Rahi, P.B. Mather, D.A. Hurwood, Understanding the Genomic Basis of Adaptive Response to Variable Osmotic Niches in Freshwater Prawns: A Comparative Intraspecific RNA-Seq Analysis of Macrobrachium australiense, J Hered 108(5) (2017) 544-552.

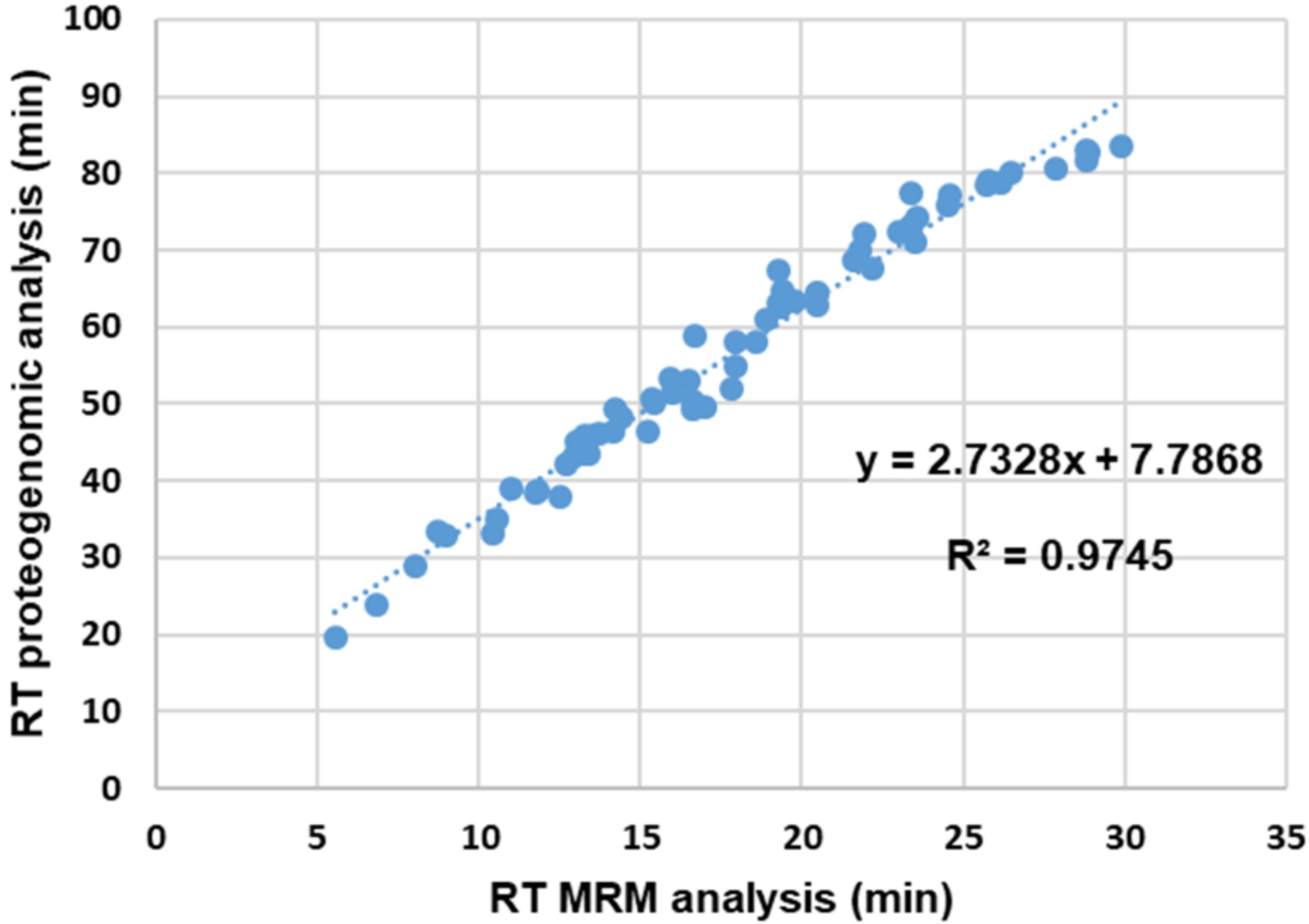
[40] K. Groebe, K. Hayess, M. Klemm-Manns, G. Schwall, W. Wozny, M. Steemans, A.K. Peters, C. Sastri, P. Jaeckel, W. Stegmann, H. Zengerling, R. Schopf, S. Poznanovic, T.C. Stummann, A. Seiler, H. Spielmann, A. Schrattenholz, Protein biomarkers for in vitro testing of embryotoxicity, J Proteome Res 9(11) (2010) 5727-38.

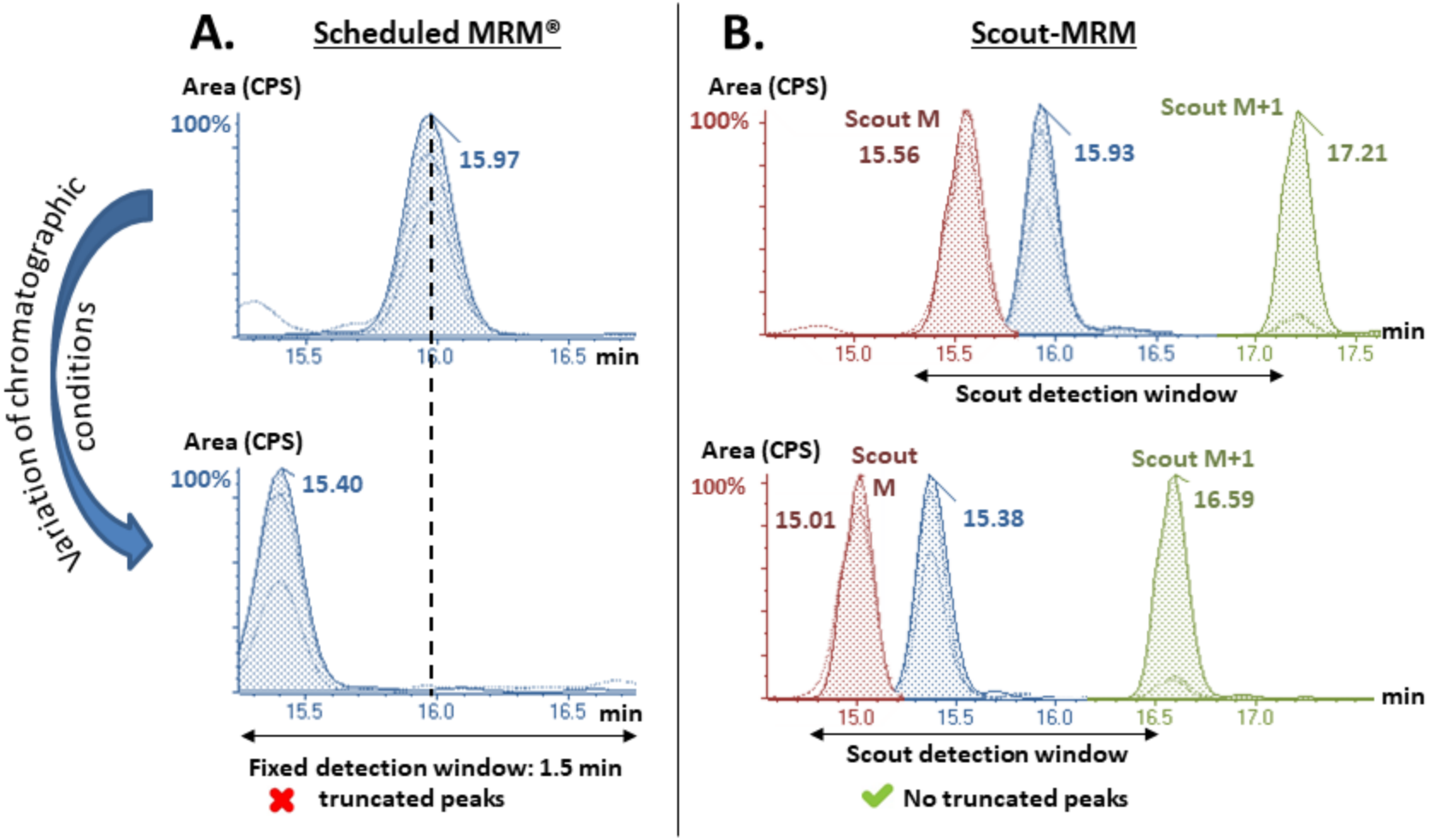
[41] S. Muthukrishnan, H. Merzendorfer, Y. Arakane, Q. Yang, Chitin Organizing and Modifying Enzymes and Proteins Involved In Remodeling of the Insect Cuticle, Adv Exp Med Biol 1142 (2019) 83-114.

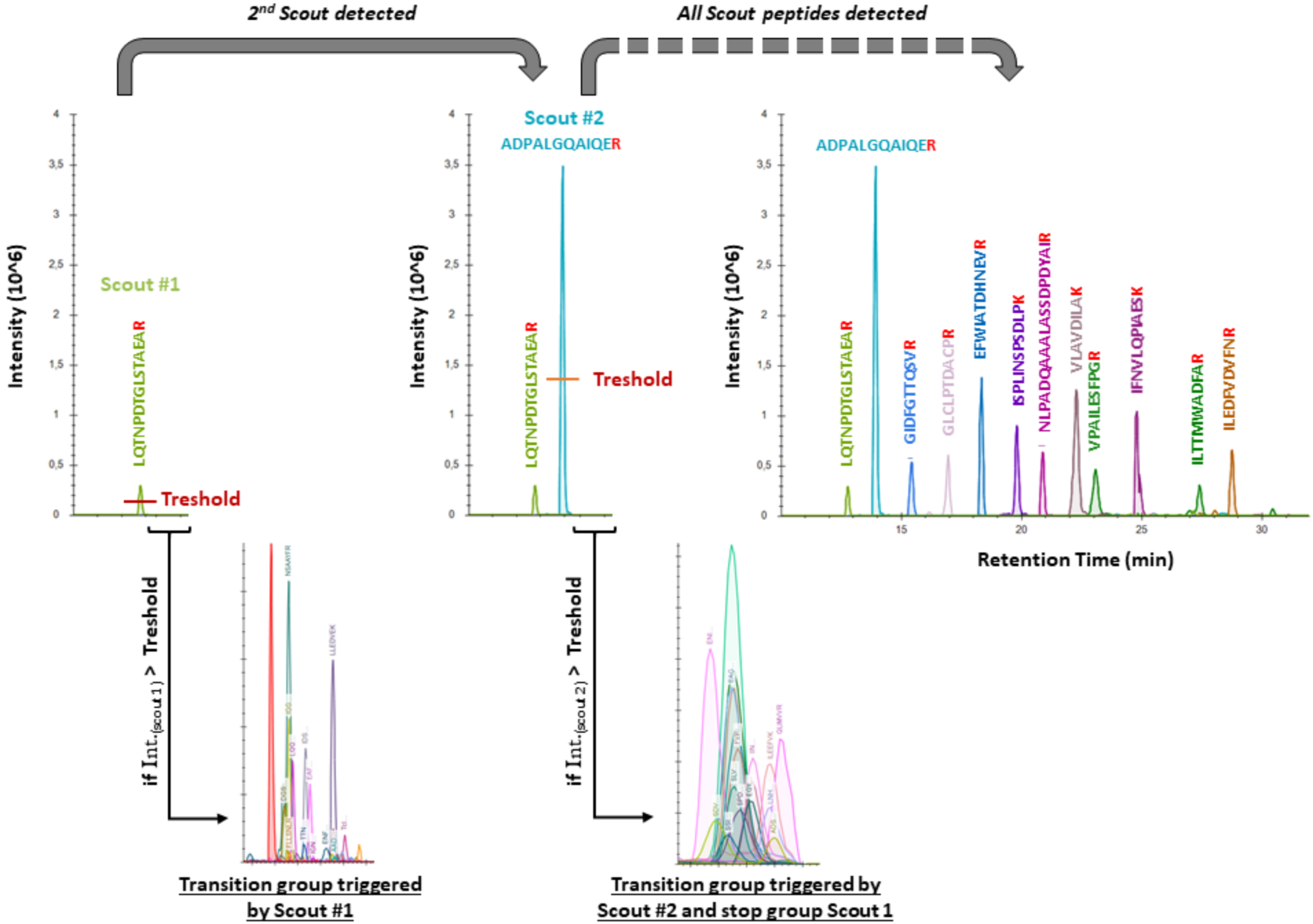
[42] R. Coulaud, O. Geffard, B. Xuereb, E. Lacaze, H. Queau, J. Garric, S. Charles, A. Chaumot, In situ feeding assay with Gammarus fossarum (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring, Water Res 45(19) (2011) 6417-29.

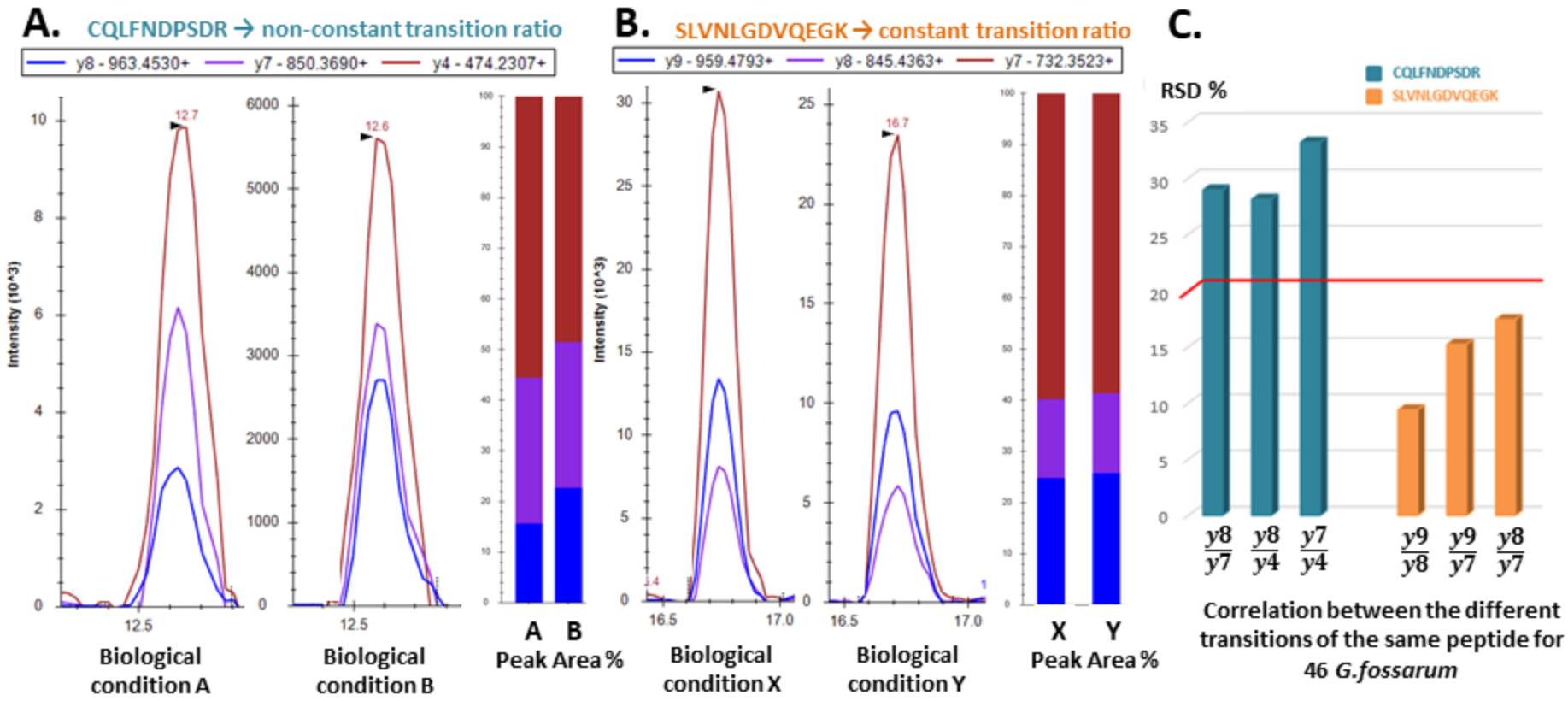
[43] C. Lopes, A. Chaumot, B. Xuereb, R. Coulaud, G. Jubeaux, H. Queau, A. Francois, O. Geffard, In Situ Reproductive Bioassay with Caged Gammarus fossarum (Crustacea): Part 2-Evaluating the Relevance of Using a Molt Cycle Temperature-Dependent Model as a Reference to Assess Toxicity in Freshwater Monitoring, Environ Toxicol Chem 39(3) (2020) 678-691.


A.
Peptide: GTLAVIPVQNR
y8 - 896.5312+
y7 - 825.4941+
y6 - 726.4257+
y5 - 613.3416+
Intensity (10^3)
15.5
20.1
25.7
34.0
Heavy peptide: GTLAVIPVQNR*
(10^3)
y8 - 906.5395+ (heavy)
y7 - 835.5024+ (heavy)
y6 - 736.4340+ (heavy)
y5 - 623.3499+ (heavy)
15.6
B.
y5 - 595.3086+
y4 - 496.2402+
y2 - 310.1761+
y1 - 147.1128+
Peptide: LQQEQVADYK
Intensity (10^3)
8.1
6.9
13.9
17.4
18.8
21.0
C.
y8 - 902.5306+
y7 - 789.4465+
y6 - 702.4145+
y5 - 589.3304+
Peptide: TDLSITLAER
Intensity (10^3)
16.9
20.1
33.6

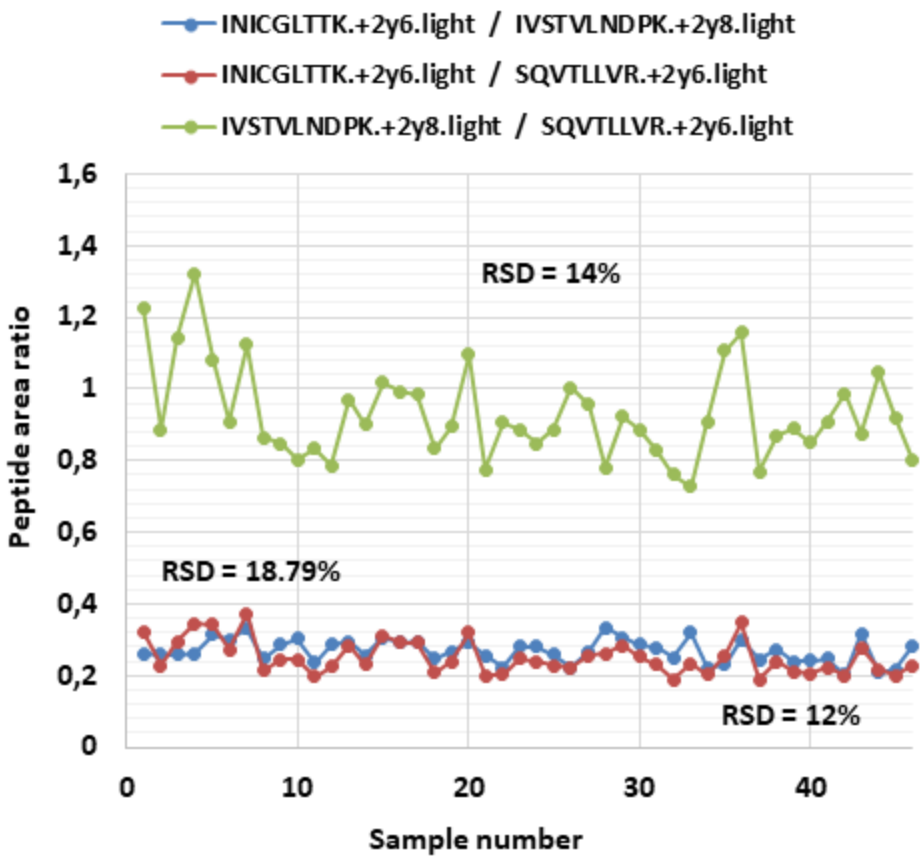

RT proteogenomic analysis (min)
RT MRM analysis (min)
y = 2.7328x + 7.7868
R² = 0.9745
0
10
20
30
40
50
60
70
80
90
100
0
5
10
15
20
25
30
35


A.
Scheduled MRM®
Area (CPS)
100%
15.97
15.5
16.0
16.5
min
Variation of chromatographic conditions
Area (CPS)
100%
15.40
15.5
16.0
16.5
min
Fixed detection window: 1.5 min
truncated peaks
B.
Scout-MRM
Area (CPS)
100%
Scout M
15.56
15.93
Scout M+1
17.21
15.0
15.5
16.0
16.5
17.0
17.5
min
Scout detection window
Area (CPS)
100%
Scout M
15.01
15.38
Scout M+1
16.59
15.0
15.5
16.0
16.5
17.0
min
Scout detection window
No truncated peaks

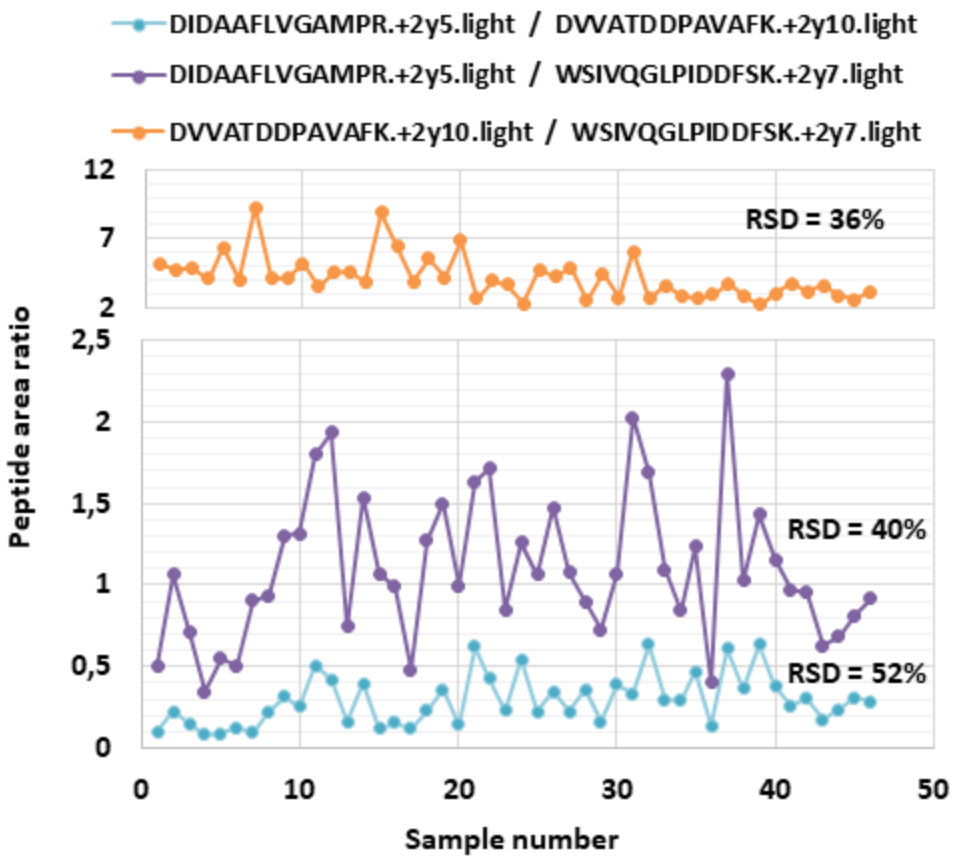

2nd Scout detected
All Scout peptides detected
Intensity (10^6)
Scout #1
LQTNPDTGLSTAEAR
Treshold
Scout #2
ADPALGQAIQER
LQTNPDTGLSTAEAR
Treshold
ADPALGQAIQER
LQTNPDTGLSTAEAR
GIDFGTTQSVR
GLCLPTDACPR
EFWATDHNEVR
ISPLINSPSDLPK
NLPADQAAALASSDPDYAIR
VLAVDILAK
VPAILESFPGR
IFNVLQPIAESK
ILTTMWADFAR
ILEDFVDVFNR
Retention Time (min)
if Int.(scout 1) > Treshold
if Int.(scout 2) > Treshold
Transition group triggered by Scout #1
Transition group triggered by Scout #2 and stop group Scout 1

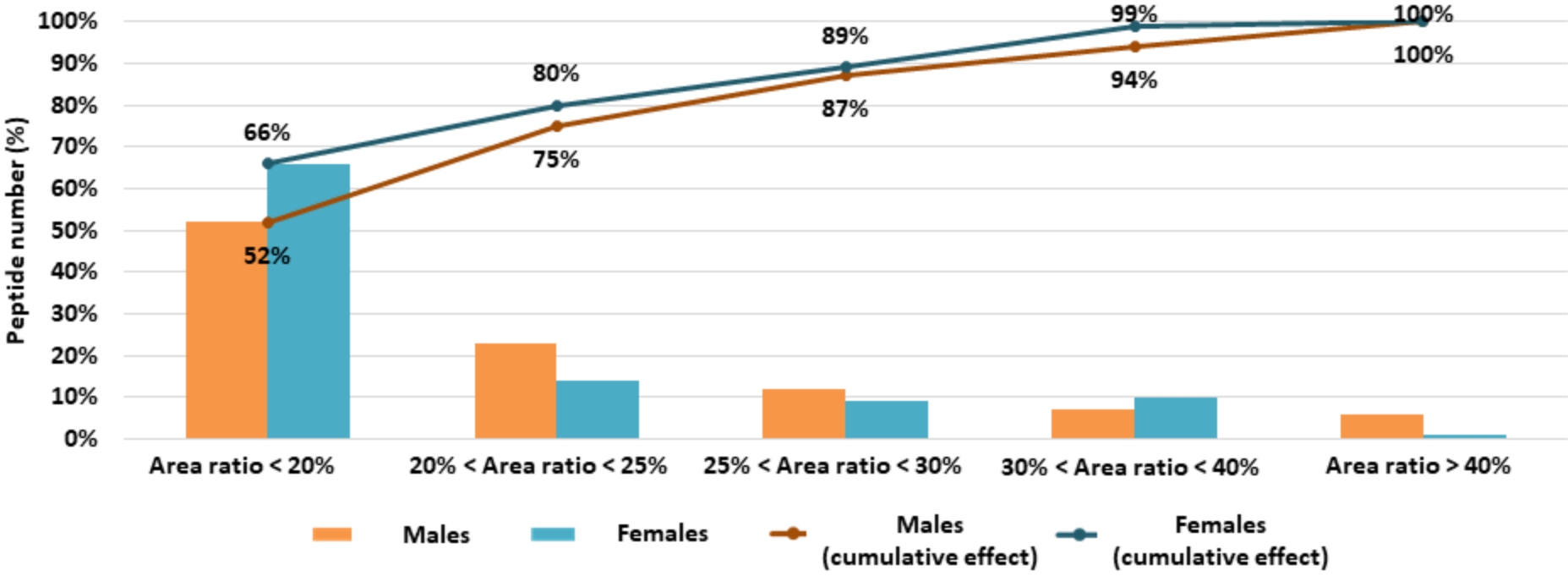

A.
CQLFNDPSDR → non-constant transition ratio
y8 - 963.4530+
y7 - 850.3690+
y4 - 474.2307+
Intensity (10^3)
12.7
12.6
12.5
12.5
Biological condition A
Biological condition B
A
B
Peak Area %
B.
SLVNLGDVQEGK → constant transition ratio
y9 - 959.4793+
y8 - 845.4363+
y7 - 732.3523+
Intensity (10^3)
16.7
16.5
17.0
16.5
17.0
Biological condition X
Biological condition Y
X
Y
Peak Area %
C.
RSD %
CQLFNDPSDR
SLVNLGDVQEGK
y8/y7
y8/y4
y7/y4
y9/y8
y9/y7
y8/y7
Correlation between the different transitions of the same peptide for 46 G.fossarum

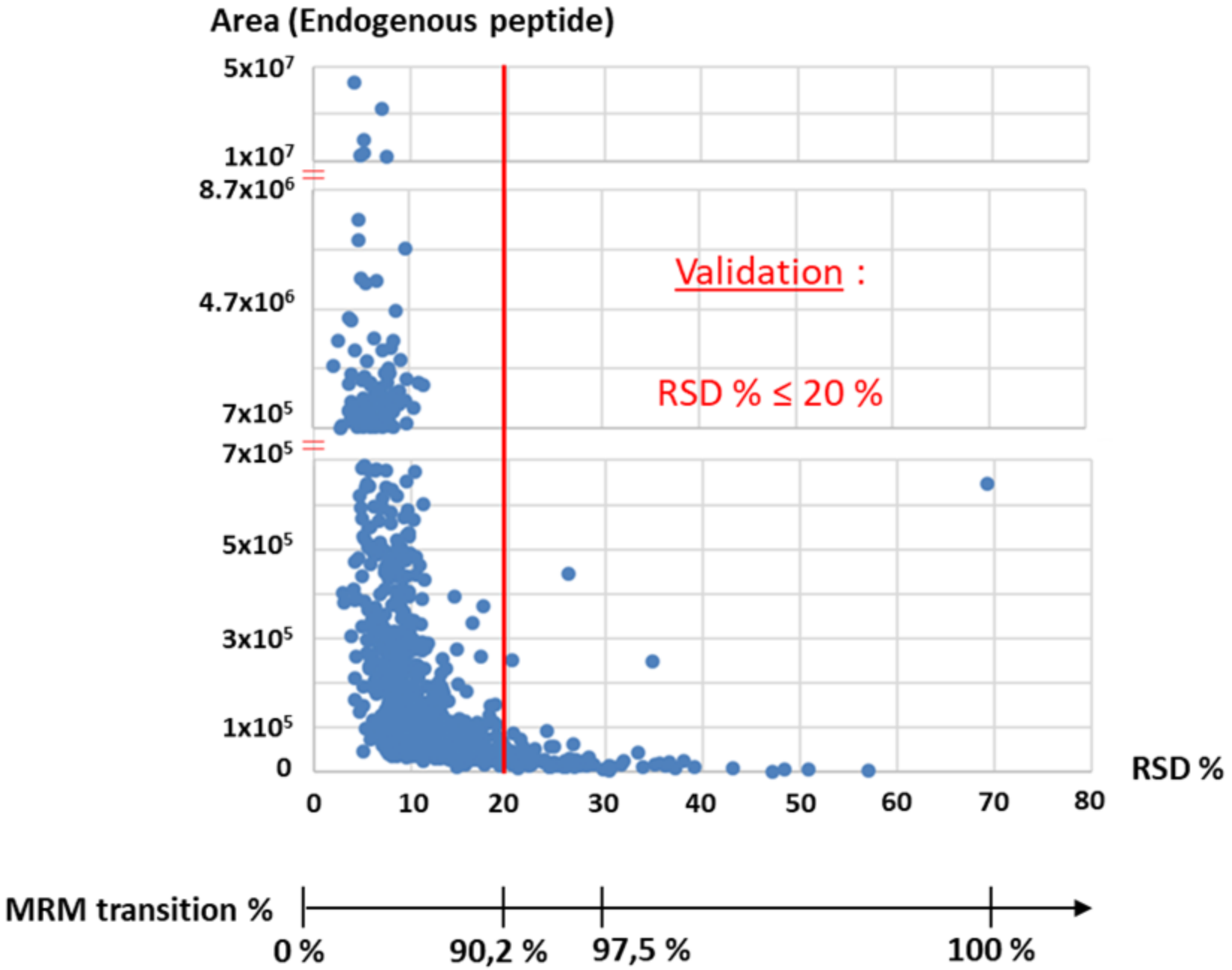
Aspartate aminotransferase

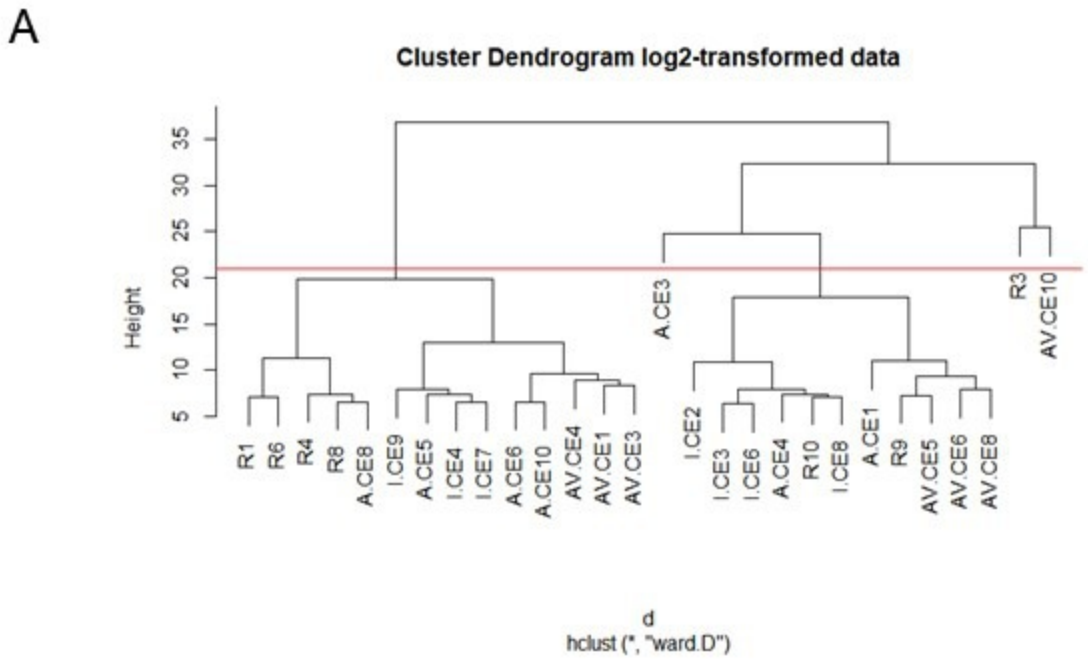


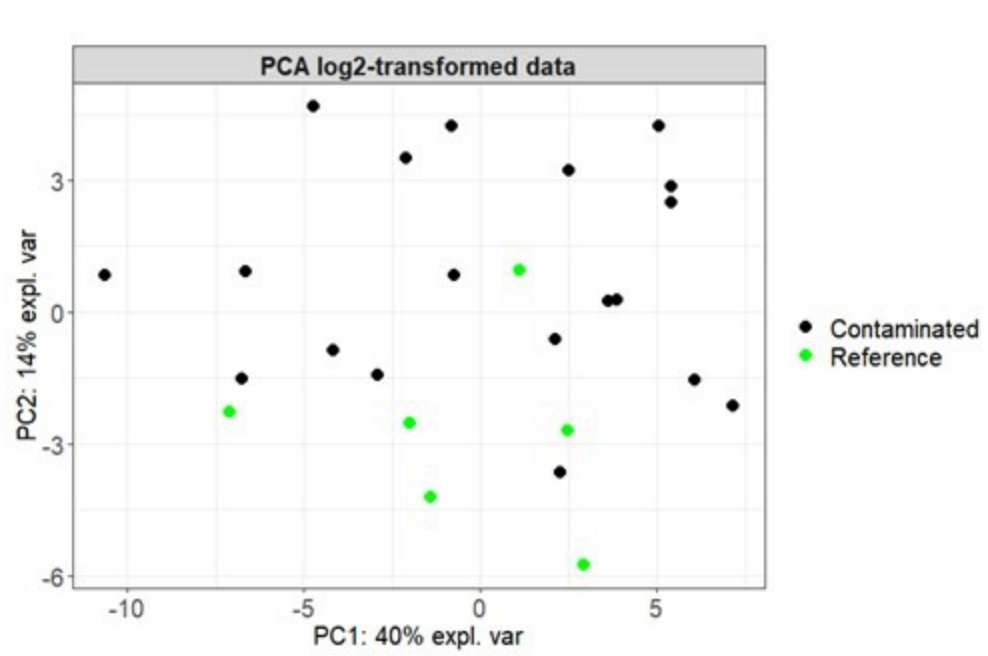
Cytosolic malate dehydrogenase

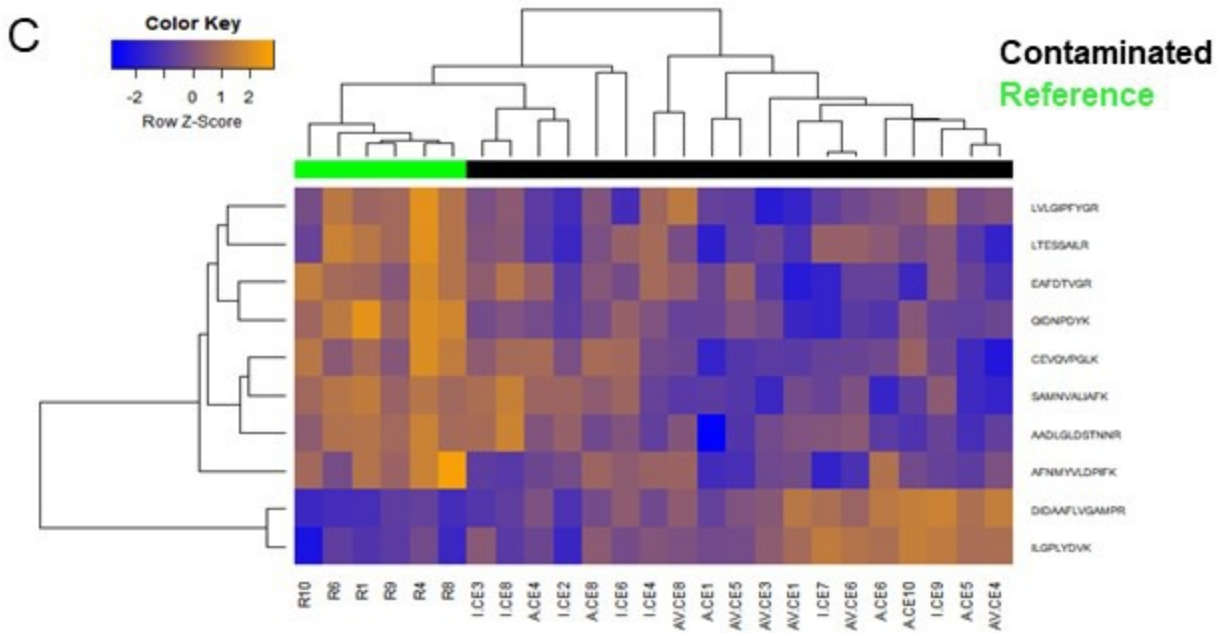


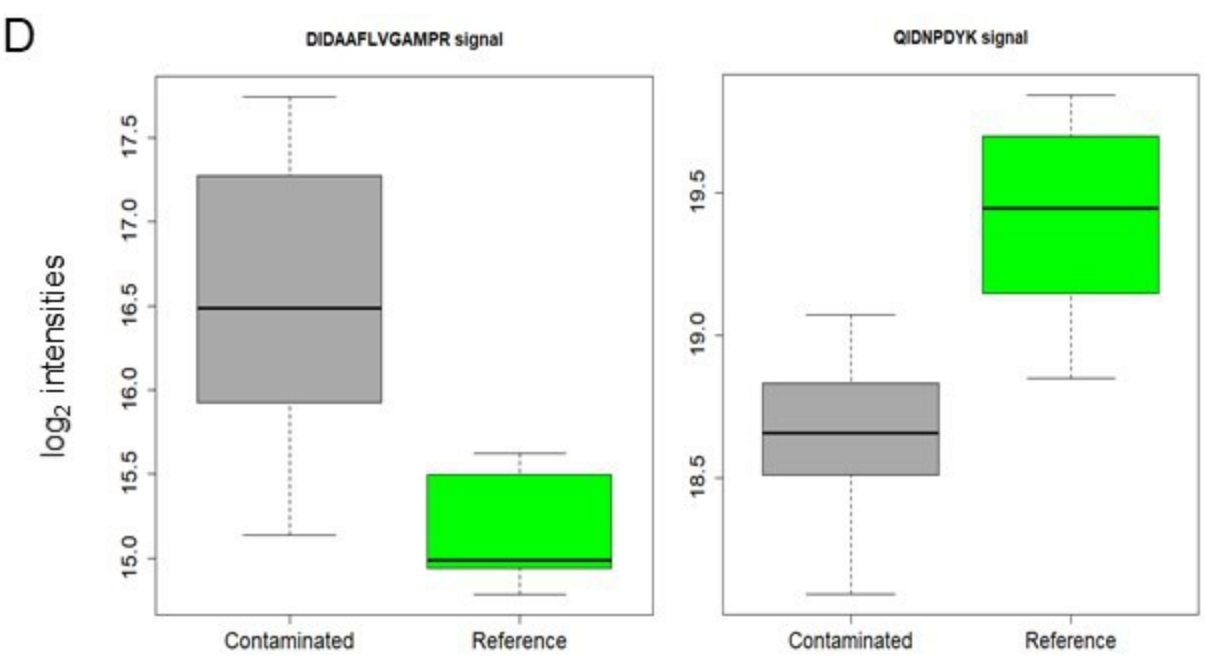

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Peptide number (%)
66%
80%
89%
99%
100%
52%
75%
87%
94%
100%
Area ratio < 20%
20% < Area ratio < 25%
25% < Area ratio < 30%
30% < Area ratio < 40%
Area ratio > 40%
Males
Females
Males (cumulative effect)
Females (cumulative effect)


Area (Endogenous peptide)
5x10⁷
1x10⁷
8.7x10⁶
4.7x10⁶
7x10⁵
7x10⁵
5x10⁵
3x10⁵
1x10⁵
0
Validation :
RSD % ≤ 20 %
0
10
20
30
40
50
60
70
80
RSD %
MRM transition %
0 %
90,2 %
97,5 %
100 %


A
Cluster Dendrogram log2-transformed data
Height
5
10
15
20
25
30
35
R1
R6
R4
R8
A.CE8
I.CE9
A.CE5
I.CE4
I.CE7
A.CE6
A.CE10
AV.CE4
AV.CE1
AV.CE3
A.CE3
I.CE2
I.CE3
I.CE6
A.CE4
R10
I.CE8
A.CE1
R9
AV.CE5
AV.CE6
AV.CE8
R3
AV.CE10
d
hclust (*, "ward.D")


B
PCA log2-transformed data
PC2: 14% expl. var
3
0
-3
-6
-10
-5
0
5
PC1: 40% expl. var
Contaminated
Reference


C
Color Key
-2
0
1
2
Row Z-Score
Contaminated
Reference
LVLGIPFYGR
LTESSAILR
EAFDTVGR
QIDNPDYK
CEVDVPGLK
SAMNVALIAFK
AADLGLDSTNNR
AFNMYVLDPIFK
DIDAAFLVGAMPR
ILGPLYDVK
R10
R6
R1
R9
R4
R8
I.CE3
I.CE8
A.CE4
I.CE2
A.CE8
I.CE6
I.CE4
AV.CE8
A.CE1
AV.CE5
AV.CE3
AV.CE1
I.CE7
AV.CE6
A.CE6
A.CE10
I.CE9
A.CE5
AV.CE4


D
DIDAAFLVGAMPR signal
log2 intensities
17.5
17.0
16.5
16.0
15.5
15.0
Contaminated
Reference
QIDNPDYK signal
19.5
19.0
18.5
Contaminated
Reference

Protein discovery

(DDA)

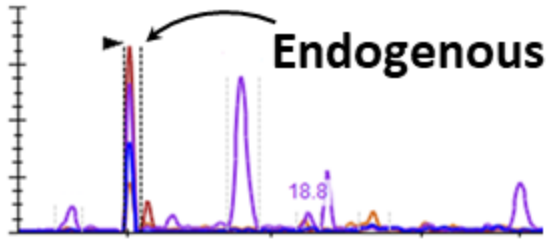
Nano-LC-HRMS

Proteogenomic study

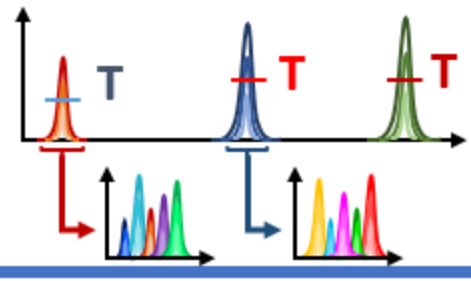
Protein data base

Targeted proteomics

LC-MS (MRM)



Scout-MRM development



Validation

- Linearity
- LOD /LOQ
- Repeatability
- Interference free

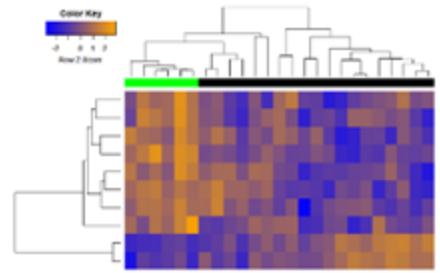
Application in biomonitoring

Reference site

VS

Contaminated site

Differential analysis



Sentinel species : *Gammarus. fossarum*

1 311 444 protein sequences

263 proteins selected for the study (919 peptides)

Detected by LC-MS (MRM)

182 proteins
341 peptides

157 proteins
277 peptides
831 transitions

10 peptides with variations in expression