Proteomic Analysis of Mussels Exposed to Fresh and Weathered *Prestige*'s Oil

Itxaso Apraiz[1][#], Giulia Leoni[1][#][§], David Lindenstrand[2], Jan-Olov Persson[2], and Susana Cristobal[1][*]

[1]Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden

[2]Department of Mathematics and Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

*Corresponding author: Susana Cristobal, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden, Tel: +46 8 164239; Fax: +46 8 153679; E-mail: Susana.Cristobal@dbb.su.se

[#]These authors contributed equally to this work

[§]Present address: Department of Gene Therapy, National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, United Kingdom

Citation: Apraiz I, Leoni G, David L, Persson JO, Cristobal S (2009) Proteomic Analysis of Mussels Exposed to Fresh and Weathered *Prestige*'s Oil. J Proteomics Bioinform 2: 255-261. doi:10.4172/jpb.1000084

Abstract

Biomonitoring programs that use mussels to assess the water quality around the world could benefit from the use of proteomics techniques. These could be applied to obtain protein expression signatures of exposure to pollution that could be further used for prediction purposes. This would require that a combination of univariate and multivariate statistical analyses of proteomics data were utilized to obtain robust models. We show an application of this approach on mussels exposed to fresh fuel, and weathered fuel in a laboratory experiment that tried to mimic the effects of the *Prestige*'s oil spill. By the combination of those statistical analyses, a set of protein spots were selected that could be used to classify mussels exposed to the two types of fuel oil. As an example of the possibilities that this approach could offer to biomonitoring programs, mussels were collected from ten sampling stations along the NW and NE coasts of the Iberian Peninsula, and their protein expression patterns monitored.

Introduction

Coastal ecosystems are often exposed to diverse sources of pollution such as oil spills, and this has detrimental effects in the biota. The high concern about pollution's effects in the environment apprizes the value of marine biomonitoring programs. The biological effects of the 1989 Exxon Valdez oil spill at different trophic levels have been reported in the years following the disaster. The conclusions from those investigations have already been reviewed somewhere else (Harwell and Gentile, 2006).

Many of the biomonitoring programs use mussels as bioindicators of pollution due to their wide distribution, sessile nature, and filter feeding mechanism (Widdows and Donkin, 1992). Mussels have a reduced biotransformation capacity and can accumulate several xenobiotic compounds that can severely affect metabolic homeostasis. The bioaccumulation capacity is useful in biomonitoring programs because it can

show the actual pollution levels in that environment, and can lead to biomagnification of pollutants higher up in the food chain. Therefore, levels of different cellular and molecular biomarkers can be measured in mussels in order to obtain a picture of their health status (Cajaraville et al., 2000, Guerlet et al., 2007; Zorita et al., 2007).

The *Prestige* tanker's accidental oil spill (November 2002, 42°12.5'N, 12°3'W) resulted in more than 60,000 tons heavy fuel oil overspreading Galician and Bay of Biscay waters in the following months (Albaiges et al., 2006). It has been reported that a year after the *Prestige* oil spill, the incidence of natural oil weathering processes (by evaporation, dissolution, biodegradation, and photo-oxidation) was low, and mainly enhanced in oil stranded on the shoreline (Diez et al., 2007). In the mentioned study, 17% of the analyzed samples did not match the *Prestige* oil fingerprint, and half of these corresponded to a common spill. These results emphasize the need of tools to distinguish the effects that different sources of pollution can have in biota.

Proteomics methods that allow biological data classification and characterization by univariate and multivariate analyses have already been recommended and applied previously (Meleth et al., 2005; Chich et al., 2007; Karp and Lilley, 2007; Karp et al., 2007). In environmental 2-DE proteomics of mussel, Student *t*-test, analysis of variance (ANOVA), principal components analysis (PCA), and hierarchical clustering have been applied to obtain protein expression signatures specific to pollutants, and to a gradient of pollution, but no classification models were built (Apraiz et al., 2006; Mi et al., 2007; Amelina et al., 2007). Monsinjon et al., reported a classification model based on protein peaks obtained by ProteinChip© array technology and surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF)-mass spectrometry (MS), but because of the criteria to guard against overfitting, classification was not successful (Monsinjon et al., 2006). The use of protein expression signatures (PES) to build up statistically verified models that could classify samples exposed to different sources of pollution, could become a powerful tool for biomonitoring programs in the future.

Therefore, a laboratory experiment was set where mussels, *Mytilus galloprovincialis*, were exposed to fresh and weathered *Prestige*-like fuel oil for two and sixteen days. A control group was kept in parallel. Mussel digestive glands were subjected to a simple cellular prefractionation and liquid chromatography (LC) coupled with two-dimensional electrophoresis (2-DE) method previously developed by our group, and that has been successful in separating four stations along a pollution gradient around the harbor of Gothenburg (Amelina et al., 2007). Here, we performed

ANOVA, and false discovery rate (FDR) procedures to extract protein spots composing a PES that were further analyzed by principal components analysis (PCA). These spots were successful in separating the exposed groups. Furthermore, samples from ten sampling sites along the Galician (NW) and Bay of Biscay (NE) coasts were also processed by LC coupled with 2-DE, and we showed how the previously obtained PES could be used to classify the sampling sites.

Materials and Methods

Animal Collection and Experimental Procedure

Mussels, *M. galloprovincialis*, 3.5 to 4.5 cm shell length and of undetermined sex were collected at low tide from ten different sampling sites in the NW and NE of the Iberian Peninsula in July 2005 for the field studies, and in a NE location in September 2005 for the laboratory experiments. Sampling sites in the NW were Sao Bartolomeu do Mar (41°34'36''N, 8°48'2''W) (from now on referred as Sao Bartolomeu), Aguiño (42°31'13''N, 9°0'36''W), Caldebarcos (42°50'48''N, 9°7'52''W), Camelle (43°11'38''N, 9°5'48''W), and Segaño (43°27'21''N, 8°18'34''W). Sampling sites in the NE were Muskiz (43°21'32''N, 3°6'40''W), Arrigunaga (43°21'17''N, 3°1'11''W), Gorliz (43°25'7''N, 2°56'51''W), Mundaka (43°24'16''N, 2°41'43''W), and Hondarribia (43°22'40''N, 1°47'24''W). Mussels for the laboratory experiment were collected from Mundaka (43°24'16''N, 2°41'43''W), a relatively clean location in the mouth of the Biosphere Reserve of Urdaibai estuary (Orbea and Cajaraville, 2006). Sampling sites are summarized in Figure 1.

Mussels collected from Mundaka for the laboratory experiment were acclimatized in the laboratory for 15 days and afterwards divided in three high-density polyethylene tanks at a mussel density of one mussel per three liters of seawater. Water temperature was kept at 20 °C, salinity at
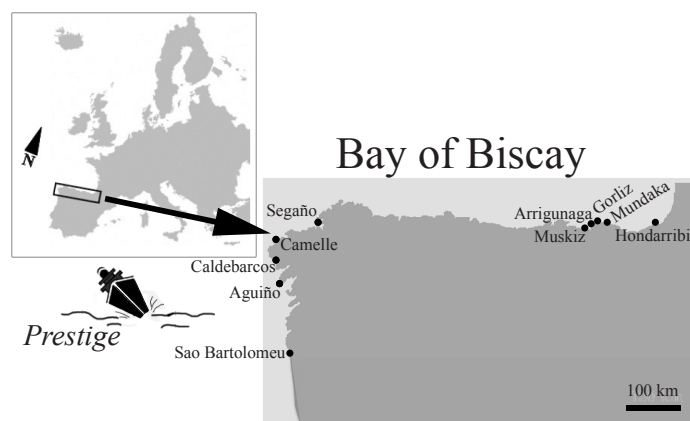

Figure 1: Map showing sampling sites along the NW and NE coasts in the Iberian Peninsula.

35‰ and oxygen levels above 6 mg/L by constant aeration. A photoperiod of 11 hours was set and commercial food (JBL KorallFluid, JBL BmgH & Co. KG, Neuhofen, Germany) provided every day. The heavy fuel oil that is similar to that spilled by the *Prestige* (IFO 380, marine fuel RMG 35-ISO 8217) was provided by the Vigo Technical Office Against Accidental Marine Spills (University of Vigo, Spain). Oily sediments were prepared by mixing 150 mL oil with 5 kg gravel, and 6 kg sand, and placed on the bottom of the tanks. Weathered fuel oil (WF) was obtained by letting the sediment stand in a water-filled tank during two and a half months. Fresh fuel oil (FF) was obtained by adding the sediment to a water filled tank precisely before the exposure started. Exposure to FF tried to mimic the situation in the most affected areas in the NW immediately after the *Prestige*'s oil spill, whereas exposure to WF would mimic the situation in any of the sampling sites months after the spill. Mussels were also kept in a control tank where no oil was added.

For our experiments, four mussels were collected from each sampling site and eight from each tank: four after two days of exposure, and four after 16 days had passed. Digestive glands were immediately dissected out and frozen in liquid nitrogen *in situ* in all the cases, and kept at - 80 °C until the proteomics analysis.

Proteomics Analysis

Digestive glands were processed following a protocol for sample prefractionation and 2-DE protein separation already described (Amelina et al., 2007). Briefly, digestive glands were homogenized with the aid of a pestle and AG®501-X8 Resin glass beads (BioRad Laboratories, Inc., Hercules, CA, USA) in a homogenization buffer containing a protease inhibitor cocktail. Following homogenization, a three-step centrifugation was applied and an organelle-enriched fraction therefore obtained. Low-abundant proteins were then obtained by an anion-exchange chromatography in batch using Q-sepharose™ Fast Flow (Amersham Biosciences AB, GE Healthcare, Uppsala, Sweden).

Proteins from the eluted fractions were then precipitated by the addition of 20% trichloroacetic acid in 100% cold acetone containing 0.07% β-mercaptoethanol, and the precipitate was washed with 100% cold acetone containing 0.07% β-mercaptoethanol. Precipitated proteins were solubilized in a solubilization buffer described by Rabilloud with some modifications (Rabilloud, 1998; Amelina et al., 2007), alkylated with 30 mM iodoacetamide (IAA) in darkness and mixed with a rehydration buffer previous to the isoelectrofocusing (IEF) step. Proteins (300 µg) were loaded in the PROTEAN® IEF Cell (BioRad Laboratories) tray and IPG strips (11 cm, pH range of 4-7, BioRad Laboratories) placed on top. The following program was followed: passive rehydration for 12 h at 50 V and 20 °C, 250 V for 15 min, rapid voltage ramping to obtain 8,000 V and a final focusing at 8,000 V until 35,000 V·h were achieved. The focusing was held at 500 V until strips were removed from the tray. In all the steps, a maximum current limit of 50 µA per strip was established. IPG strips were first reduced (1% dithiothreitol (w/v)), and then alkylated (4% IAA (w/v)) in an equilibration buffer (Amelina et al., 2007) previous to SDS-PAGE.

Equilibrated IPG strips were laid on top of homogeneous 12.5% Tris-HCl Criterion™ Precast Gels (BioRad Laboratories) and SDS-PAGE run at 120 V. 2- DE gels were fixed and stained with CBB G-250 for 12-18 h. Distained 2-DE gels were scanned in a UMAX Image Scanner (Amersham Biosciences) and analyzed by ImageMaster™ 2D Platinum 6.0 (Amersham Biosciences). 2-DE gel images were cropped, spots automatically detected, wrong detections manually corrected and finally the volume % (vol%) of each spot calculated based on the total spot volume in each gel. A master gel was chosen for each sampling site and exposure group. Spots from the rest of the three gels inside each sampling site/group were then matched to the master gel. Higher-level match-sets were constructed between master gels. Image analyses of the field study and the laboratory exposure were separately performed, but their highest-level master images were finally matched between them.

Statistical Analysis

Vol% data was exported to SAS® 9.1.9 (SAS Institute Inc., Cary, NC, USA) and MATLAB® 7.5.0 (The MathWorks, Inc., Natick, MA, USA) for the statistical analyses. In total, 468 spots were obtained in the match set from the laboratory exposure experiment. Missing values in the data set came from spots with intensities lower than the detection limit of the image analyzer, or from spots absent in the 2-DE gels, but not from an incorrect matching. Therefore, zero values were input. In the few cases where the missing value happened in a group with relatively high values, the mean value of the three replicates from the group was input.

Two-way ANOVA was performed on each spot separately to extract those spots that differed among the groups, based on the following linear effects model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

i = 1,2; j = 1, 2, 3; and k = 1, 2, 3, 4, where α is the time effect average over treatments, β is the treatment effect average over time, γ is the interaction effect, and ε is the

variation within each group of 4 replicates, $\varepsilon_{ijk} \sim N(0, \sigma_{ij})$. The response variable *y* is the value of the specific spot. On account of performing multiple tests, there will necessarily be a number of false positives. By use of the False Discovery Rate (FDR) procedure (Hochberg and Benjamini, 1995) we can protect against too many false positives. FDR was set to 5 %.

PCA is a multivariate statistics technique that takes into account a group of variables instead of focusing in one variable at a time, as is the case for univariate analyses. PCA was used to find out if there was any structure in the data selected after the ANOVA and FRD analyses that could explain differences among the exposure groups. A covariance matrix where each spot was set as a variable and each gel as an observation was used to extract the principal components. In order to improve the PCA outcome, several spots were removed from the dataset.

Finally, the vol% of the selected spots was obtained from the field experiment data. A putative group membership for the different sampling sites was obtained based on the new variable's proximity to the experimental variables that were separated by the PCA.

Results and Discussion

In this work, organelle-enriched fractions were obtained from mussel digestive glands and separated into 2-DE gels. Four biological replicates were run per sampling site and experimental group. In total 40 gels were obtained from the field experiment and 24 gels from the laboratory exposure. Automatic spot detection parameters were adjusted so that approximately 400 spots were detected per gel. In general the gels were alike each other, although several high-abundance spots and gel areas were found in particular cases.

Laboratory Exposure

Differences between K, FF and WF samples were analyzed after 2- and 16-day exposures. First, we conducted a PCA analysis using the whole dataset comprised of 468 spots. The results did not give a satisfactory separation of the six groups, and the variance explained by the first two components was of 54%. Therefore, the two-way ANOVA was applied to the whole dataset and 178 spots were separated for which the ANOVA model separated the six groups on a 1% significance level. In other words, there were 178 spots for which the null hypothesis that all groups were equal on a 1% significance level could be rejected. Applying FDR at a 5% rate, a set of 148 spots was selected, about eight of which (5% of 148) were expected to be false discoveries.

Including both 2- and 16-days exposure data in the PCA,

a clear separation of groups could not be obtained. Taking the 2- and 16-day exposures separately, only the 16-day exposure data gave a clear separation of groups. Therefore, the 2-days exposure data was excluded, and analysis proceeded with the 16-days data only. It was hypothesized that this data would provide the analysis with a more realistic picture of the mechanisms of response to the pollution at a molecular level. To further improve the separation of the PCA, seven spots that showed high variation within one of the groups, were removed, and a neat separation of the three exposure groups was obtained with the selected 141 spots forming the PES. The first principal component separated the K, WF and FF groups form each other, and the second component separated the WF from the K and FF groups, indicating that the selected PES may be used to classify mussels according to exposure to the different sources of oil under study. The PCA score plots for the 2- and 16-day exposures are shown in Figure 2. At this point one could obtain a different selection of spots with a one-way ANOVA procedure on the 16-day exposure data subset. Although this was a possibility, the exposure groups' separation with the current selection of PES was satisfactory, and therefore they were kept for the following analyses.

Questions we cannot answer in a qualitative way from the present small experiment are how the mussels are affected by the concentration of oil, the age of the oil, and the
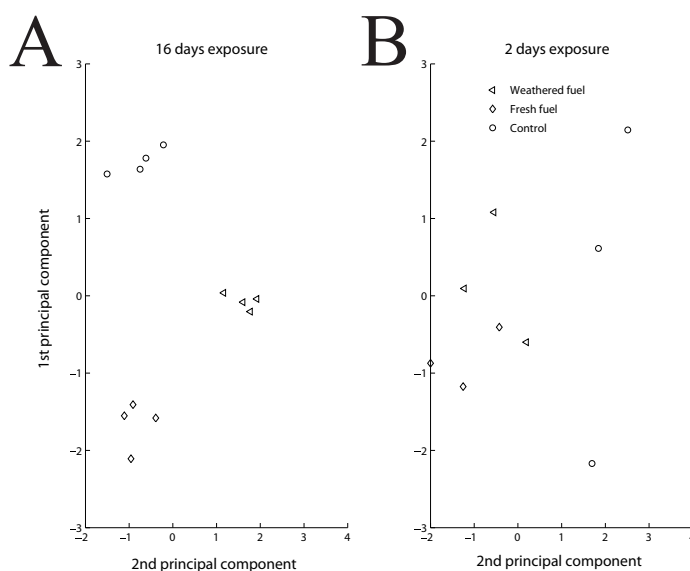

Figure 2: PCA score plot obtained by analyzing 141 spots (variables) and 12 samples (observations) for the 2- and 16-day exposures. Only samples have been plotted. **A**: Samples from the 16-day exposure, where the first two components explained 71.9% of the variability in the data. **B**: Samples from the 2-day exposure, where the first two components explained 64.3% of the variability in the data.
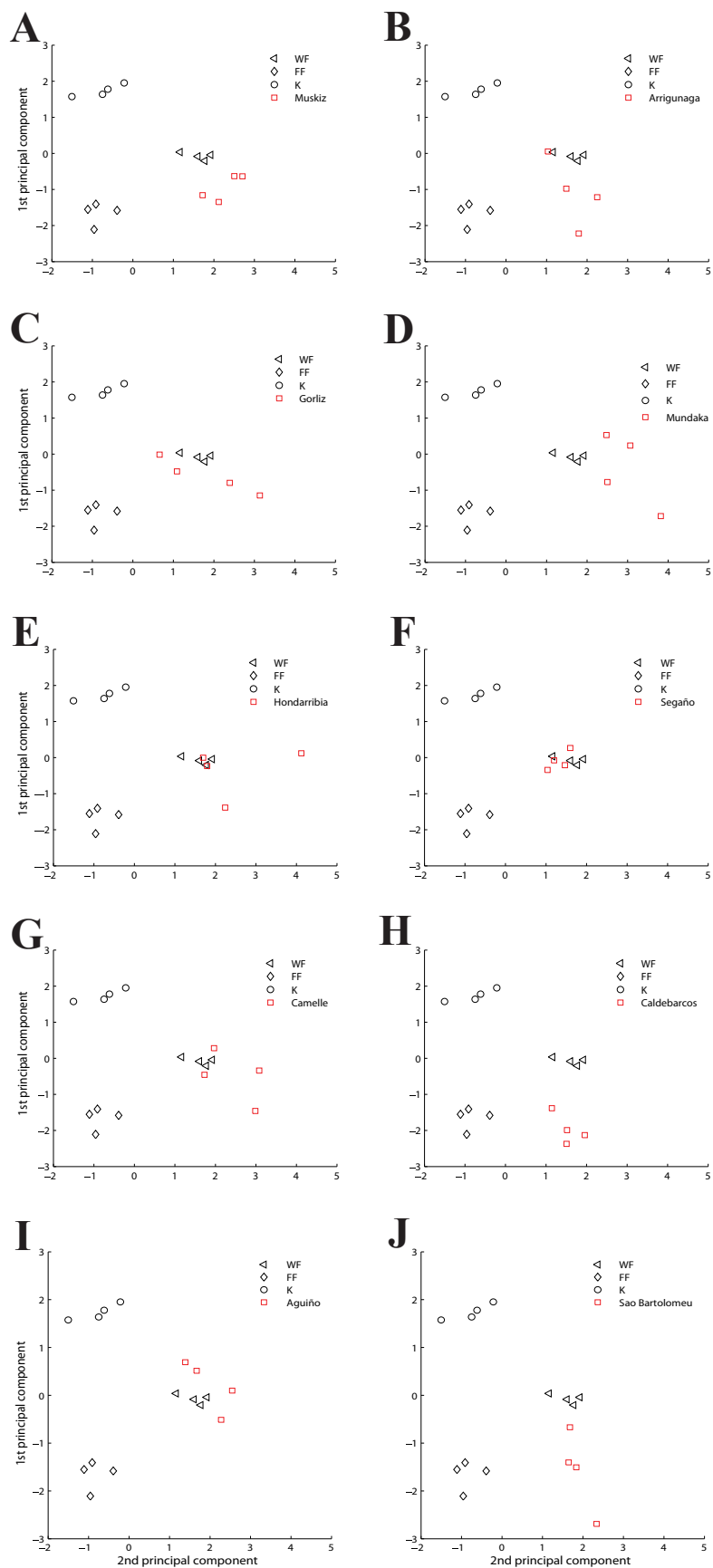
Figure 3: PCA score plot obtained by analyzing 141 spots (variables) and all the samples (observations) after 16 days of exposure. Samples corresponding to each experimental group (K, WF, FF) are plotted in black, and those corresponding to the sampling site groups are plotted in red as follows: **A**: Muskiz; **B**: Arrigunaga; **C**: Gorliz; **D**: Mundaka; **E**: Hondarribia; **F**: Segaño; **G**: Camelle; **H**: Caldebarcos; **I**: Aguiño; and **J**: Sao Bartolomeu.

amount of time exposed to oil. The reasons for that are the scarce amount of data (12 observations), and the lack of additional data that could be used to validate a potential model with. Hence, we only attempted to find out if there were spots forming a PES that may be used to separate mussels into groups of exposed to oil spill, from unexposed ones.

Field Experiment

The proteome profiles of ten sampling sites in the NW and NE coasts of the Iberian Peninsula were analyzed after two and a half years after the *Prestige*'s oil spill, and the values of the PES selected by ANOVA, FDR, and PCA recorded.

Therefore, spots in the master gel from the field experiment were matched to the master gel from the laboratory exposure group. Furthermore, the 40 gels from the field experiment were manually checked. If any of the 141 selected spots had not been matched previously, the matching was performed. Vol% values of these 141 spots from each sampling site were plotted in the PCA (Figure 3). It was observed that all the stations were placed closer to the WF group than to the K or FF groups. In particular, following the separation of groups by the first component, several groups were found closer to the FF: three samples from Arrigunaga (Figure 3B), two samples from Gorliz and Mundaka (Figure 3C and D), one sample from Camelle (Figure 3G), all the samples from Caldebarcos (Figure 3H), and three samples from Sao Bartolomeu (Figure 3J). None of the groups was closer to the K group in the first component separation. Moreover, all the groups were closer to the WF following the separation by the second component. It is worth mentioning that mussels for the laboratory experiment were collected from Mundaka in September 2005. Mundaka is considered a relatively clean sampling site (Orbea and Cajaraville, 2006). But our data showed that samples collected in Mundaka in July 2005 were clustered around the samples exposed to the WF. As it was mentioned before, owing to the scarce amount of data, no strong model was obtained in this study, so it could not be concluded whether Mundaka was polluted or not. Nevertheless, with this study it was meant to show that, in the hypothetical case when a strong model was obtained from laboratory exposure experiments, that model could be used to classify the data from field experiments, and thereby, give information about the health status of mussels.

As a conclusion, applying our proteomics approach to the study of mussels exposed to WF and FF, and to non-exposed mussels, these groups were separated by PCA based on a set of spots forming a PES obtained by ANOVA and FDR analyses. In this study, we did not try to obtain a model

that can predict sources of fuel oil pollution since our data set was too small, and we did not have an external data set for cross-validating a possible model. But, in the future, that set of 141 spots could be used to build and validate a robust model to use it with classification purposes. As an example of how this model could be used in the future, the same set of protein spots was used to group samples collected at ten sampling sites along the NW and NE coasts of the Iberian Peninsula two and a half years after the *Prestige*'s oil spill. These samples were grouped closer to the WF, rather than to the K or FF. This application would be valuable for classifying data based on an oil pollution model, but it would not detect other sources of pollution; for that purpose, models for different pollutants or mixtures of them will have to be built based on a combination of univariate and multivariate analyses. These kinds of models would take into account the orchestrated changes among proteins, and not fluctuations in individual proteins, as is the case when univariate analyses alone are applied. We believe that the development and validation of models that can predict sources of pollution based on protein expression signatures will be an important step towards robust methods for marine pollution biomonitoring in the near future. Moreover, these protein expression signatures will not be affected by biotic and abiotic factors as much as single parameter biomarkers could be influenced. The characteristics of the method hereby applied are the simplicity of the experimental procedure, the possibility to high-throughput, the low experimental and ecological (number of samples needed) costs, and the possibility of, at a glance, screening the global response to pollution.

Acknowledgements

References

1. Albaigés J, Morales-Nin B, Vilas F (2006) The Prestige

oil spill: a scientific response. Mar Pollut Bull 53: 205-207. » CrossRef » Pubmed » Google Scholar

2. Apraiz I, Mi J, Cristobal S (2006) Identification of proteomic signatures of exposure to marine pollutants in mussels (Mytilus edulis). Mol Cell Proteomics 5: 1274-1285. » CrossRef » Pubmed » Google Scholar

3. Amelina H, Apraiz I, Sun W, Cristobal S (2007) Proteomics-based method for the assessment of marine pollution using liquid chromatography coupled with two-dimensional electrophoresis. J Proteome Res 6: 2094-2104. » CrossRef » Pubmed » Google Scholar

4. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat 57: 289-300. » CrossRef » Google Scholar

5. Cajaraville MP, Bebianno MJ, Blasco J, Porte C, Sarasquete C, et al. (2000) The use of biomarkers to assess the impact of pollution in coastal environments of the Iberian Peninsula: a practical approach. Sci Total Environ 247: 295-311. » CrossRef » Pubmed » Google Scholar

6. Chich JF, David O, Villers F, Schaeffer B, Lutomski D, et al. (2007) Statistics for proteomics: experimental design and 2-DE differential analysis. J Chromatogr B Analyt Technol Biomed Life Sci 849: 261-272. » CrossRef » Pubmed » Google Scholar

7. Díez S, Jover E, Bayona JM, Albaigés J (2007) Prestige oil spill. III. Fate of a heavy oil in the marine environment. Environ Sci Technol 41: 3075-3082. » CrossRef » Pubmed » Google Scholar

8. Guerlet E, Ledy K, Meyer A, Giambérini L (2007) Towards a validation of a cellular biomarker suite in native and transplanted zebra mussels: a 2-year integrative field study of seasonal and pollution-induced variations. Aquat Toxicol 81: 377-388. » CrossRef » Pubmed » Google Scholar

9. Harwell MA, Gentile JH (2006) Ecological significance of residual exposures and effects from the Exxon Valdez oil spill. Integr Environ Assess Manag 2: 204-246. » CrossRef » Pubmed » Google Scholar

10. Karp NA, Lilley KS (2007) Design and analysis issues in quantitative proteomics studies. Proteomics 7S1: 42-50. » CrossRef » Pubmed » Google Scholar

11. Karp NA, McCormick PS, Russell MR, Lilley KS (2007) Experimental and statistical considerations to avoid false conclusions in proteomics studies using differential in-gel electrophoresis. Mol Cell Proteomics 6: 1354-1364.

12. Meleth S, Deshane J, Kim H (2005) The case for well-conducted experiments to validate statistical protocols for 2D gels: different pre-processing = different lists of significant proteins. BMC Biotechnol 5: 7. » CrossRef » Pubmed » Google Scholar

13. Mi J, Apraiz I, Cristobal S (2007) Peroxisomal proteomic approach for protein profiling in blue mussels (Mytilus edulis) exposed to crude oil. Biomarkers 12: 47-60. » CrossRef » Pubmed » Google Scholar

14. Monsinjon T, Andersen OK, Leboulenger F, Knigge T (2006) Data processing and classification analysis of proteomic changes: a case study of oil pollution in the mussel, Mytilus edulis. Proteome Sci 4: 17. » CrossRef » Pubmed » Google Scholar

15. Orbea A, Cajaraville MP (2006) Peroxisome proliferation and antioxidant enzymes in transplanted mussels of four basque estuaries with different levels of polycyclic aromatic hydrocarbon and polychlorinated biphenyl pollution. Environ Toxicol Chem 25: 1616-1626. » CrossRef » Pubmed » Google Scholar

16. Rabilloud, T (1998) Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. Electrophoresis 19: 758-760. » CrossRef » Pubmed » Google Scholar

17. Widdows J, Donkin P (1992) Mussels and environmental contaminants: bioaccumulation and physiological aspects. In: Gosling, E (ed) The Mussel 'Mytilus': Ecology, Physiology, Genetics, and Culture, Elservier, Amsterdam, pp383-424. » CrossRef » Google Scholar

18. Zorita I, Apraiz I, Ortiz-Zarragoitia M, Orbea A, Cancio I, et al. (2007) Assessment of biological effects of environmental pollution along the NW Mediterranean Sea using mussels as sentinel organisms. Environ Pollut 148: 236-250. » CrossRef » Pubmed » Google Scholar