



Applications of proteomics in marine ecology

J. L. López*

Departamento de Genética, Facultad de Biología, Universidad de Santiago de Compostela,
15782 Santiago de Compostela, Spain

ABSTRACT: Proteomics emerged in the beginning of the 1990s due to the need for new methods for protein analysis. Proteomics is a much newer discipline than genomics, and confronts similar challenges to those that genomics researchers faced in the implementation of large-scale sequencing programs. The definition of proteomics as the use of quantitative protein-level measurements of gene expression to characterize biological processes and decipher the mechanisms of gene expression control fits in with any biological approach. In the present study, proteomics is discussed and defined in parallel with genomics, given that many authors integrate proteomics in the context of functional genomics. In this Theme Section, several facets of proteomics in marine ecology were addressed: capacity, or what can be done, utility, the technology possibilities, and how to use the data obtained. As with any new and interesting technology, the expectations often exceed reality. The applications of proteomics, the advantages and disadvantages, as well as a few limitations are discussed.

KEY WORDS: Proteomics · Marine ecology · Two-dimensional electrophoresis · Liquid chromatography · Mass spectrometry

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

The sequencing of complete genomes and the development of large expressed sequence tag (EST) databases have provided us with an understanding of the genomic capacity of many organisms. However, by themselves, these data are of limited use when it comes to fully understanding processes such as development, physiology and environmental adaptation. In order to understand these processes, scientists are now faced with the problem of how to best study the co-expression of large numbers of genes under biologically meaningful conditions. Large-scale gene expression studies can be conducted using either genomic (nucleic acid-based) or proteomic (protein-based) approaches. Genomics has moved into this functional phase through the advent of technologies such as microarrays and gene probes, used to detect gene activity through messenger display (Debouck & Goodfellow 1999). These technologies have been used to produce large-scale data sets that contain information about the messenger RNA (mRNA) molecules or 'transcripts' that are present in a cell, tissue or organism at a particular time (the transcriptome), and have led to a field of study referred to as transcriptomics.

The field of proteomics involves the study of proteomes. The term 'proteome' was originally defined as all the proteins expressed by the genome (Wilkins et al. 1996). However, it is now accepted that the proteome of an organism is more than simply a catalogue of all proteins encoded by the genome because it also includes the dynamic changes within the proteome, such as post-translational modifications that occur in response to various stimuli. An area of study within proteomics is 'expression proteomics', which is defined as the use of quantitative protein-level measurements of gene expression to characterize biological processes and decipher the mechanisms of gene expression control (Anderson & Anderson 1982). Expression proteomics allows researchers to obtain a quantitative description of protein expression and its changes under the influence of biological perturbations, the occurrence of post-translational modifications and the distribution of specific proteins within cells (Anderson & Anderson 1998).

Proteomics is recognized as an extremely important tool in the study of many biological systems. However, to date there has been only limited application of proteomics to address questions in marine ecological research. In the present study, a brief introduction to

*Email: jllopez@usc.es

proteomics is provided and the advantages, disadvantages, and some of the limitations of the field are discussed. The potential for the use of proteomics to address questions in marine ecology is illustrated by our research activities on marine bivalves.

GENOMICS VS. PROTEOMICS

The field of genomics utilizes a variety of technologies to study the information content of cells, i.e. their DNA or RNA. However, the phenotype that the genotype yields is dependent on interactions amongst its genes, the metabolic chemistry of the organism (internal environment) and environmental factors (external environment). Understanding how physiological, environmental and ecological factors (and the time span over which they occur) affect the internal and external environment and ultimately the phenotype of organisms is critical for our understanding of many areas of marine ecology. Proteomics provides us with many necessary tools with which we can improve our understanding of these complex relationships.

The field of proteomics is complementary to genomics in that it provides additional information on gene expression and regulation. Proteomics also enables the analysis of other biological processes that lead to the production of proteins. For example, the analysis of transcription alone provides a limited view of gene expression because it does not take into account regulatory steps at the level of mRNA translation. The poor correlation between the amount of mRNA and their respective proteins in cells was first demonstrated by Anderson & Seilhammer (1997). mRNA is a disposable message that has no other function than to temporarily serve to convey a piece of information, whereas protein measurements relate directly to functional mechanisms. In addition, post-transcriptional changes such as alternative gene splicing and post-translational modifications of proteins, such as glycosylation or phosphorylation, significantly increase the number of different proteins above that predicted by DNA or mRNA analysis alone. With respect to post-translational modifications, it is known that the activity of proteins is regulated by their modification state. Therefore, it is possible that even though the expression of a gene may be the same in 2 situations, differences in the phosphorylation status may result in significant differences in the activity of the proteins produced. The use of transcriptomics alone provides only partial information on such changes.

Protein function and the phenotypic traits of a particular genotype depend not only on the proteins present and their possible post-translational modifications, but also on their levels of expression. The use of pro-

teomics to measure changes in the levels of expression at protein level has enabled rapid advances in our understanding of the ecological and environmental adaptations of organisms, as well as of the biogeographical distribution of species (López et al. 2001, 2002, Fuentes et al. 2002). Proteomics provides a higher level of analysis to aid the understanding of gene function in particular and biology in general (López 2005).

PROTEOMICS TOOLS AND MARINE ECOLOGY

A comprehensive description of the proteome of an organism not only provides a catalogue of all proteins encoded by the genome but also data on protein expression under defined conditions (López 2005). For proteomics to be widely adopted, a robust technology must be established that allows the large-scale research needed for a holistic approach to protein science. A fundamental technology in proteomics studies is high-resolution 2-dimensional electrophoresis (2DE), which is a powerful technique used to separate complex mixtures of denatured proteins according to their charge and molecular weight (O'Farrell 1975). Combined with non-specific protein staining, the technique permits the visualization of a very large number of gene products that represent the more abundant proteins in a cell, tissue or organism. In addition, 2DE allows for the detection of some post- and co-translational modifications of proteins, which cannot be predicted from DNA sequences or transcriptomics (Anderson & Anderson 1998).

2DE has been used to generate large amounts of proteomics data for a wide variety of biological systems (Anderson & Anderson 1998, Jungblut et al. 1999, D'Ambrosio et al. 2005). In the marine environment, 2DE has been used to screen organisms for the presence of bioactive compounds and to detect and quantify changes in gene expression at the protein level during development, as well as in response to different physiological and environmental conditions (e.g. López et al. 2001, 2002, 2005, Olsson et al. 2004, Schweder et al. 2005, Barneah et al. 2006, McDonagh et al. 2006). 2DE also has great potential for the study of genetic variability of populations, in that it allows a more representative sample of the genome to be analyzed. However, studies of genetic variability in natural populations of animal species by means of 2DE are relatively scarce. This is because 2DE is technically more difficult and time-consuming than conventional 1-dimensional electrophoresis (1DE); furthermore, early studies that used 2DE revealed substantially less genetic variation than had been estimated by 1DE (Edwards & Hopkinson 1980, Aquadro & Avise 1981,

Neel 1990). Mosquera et al. (2003) successfully used 2DE to determine the degree of genetic variability for loci that encode abundant proteins in the marine mussel *Mytilus galloprovincialis*. In addition to demonstrating that 2DE can be used to study interpopulation genetic variability in *M. galloprovincialis*, Mosquera et al.'s (2003) study also compared the results obtained by 2DE and 1DE and discussed the possible reasons for the differences observed between these 2 approaches. This study was also the first to use 2DE in an attempt to detect linkage disequilibrium between loci that encode abundant proteins. Among a total of 406 two-locus pairs analyzed for the detection of linkage disequilibrium in the population sample, 92 showed statistically significant associations. Proteomics as a tool for genetic mapping needs to be further explored, especially because information on linkage of genetic markers in marine organisms is very scarce (Beumont 1994).

Since its initial development, 2DE has improved significantly. Improvements include: simplification and standardization of the methodology, the ability to load larger amounts of protein, thereby allowing the identification and analysis of less-abundant proteins, and better reproducibility between gels. These changes, along with reduced costs, now make it possible for more laboratories to take advantage of 2DE.

Other protein separation and quantification techniques include liquid chromatography (LC), high pressure liquid chromatography (HPLC), and capillary electrophoresis (CE) (reviewed by Liebler 2002). Compared with 2DE, the amount of sample that can be used with multi-dimensional chromatography (LC/LC-MS/MS) is less restricted, the process is easier to automate, and specific classes of proteins such as very acidic, very basic, and membrane proteins—difficult to detect using 2DE—may be more readily detected. However, these techniques rely on digestion of the proteome into a complex peptide mixture before LC separation. It is questionable whether these techniques retain the ability to study proteolytic processing and post-translation modifications, which can be readily detected by 2DE. Nevertheless, a distinct advantage of the peptide-based techniques is the ability to perform very rigorous relative protein quantification between samples using isotopic labelling techniques such as iTRAQ.

Regardless of the technique used for separation, proteins are identified by mass spectrometry (MS) and bioinformatics analysis. MS allows protein identification and characterization with speed and accuracy (Aebersold 1993). It is mandatory for rapid proteomics development and plays a central role in proteome research today (Shevchenko et al. 1996). Several types of MS techniques can be used to identify proteins, e.g. peptide mass fingerprinting and partial sequencing by tandem MS, but a detailed review of these techniques

is beyond the scope of this study. In addition to the quantification and identification of proteins, recent advances in MS now enable studies of post-translational changes in proteins (Figeys & Aebersold 1997, Carr et al. 2005).

An important consideration at the onset of any proteomics study is that the protein separation method must be able to produce polypeptides in a form that is compatible with the MS technique to be used. For example, if one selects the nano-electrospray ionisation principle (Mann & Wilm 1995), the sample pre-fractionation should terminate in a liquid form and should be separated at the end of the procedure by micro-LC, HPLC or CE. If one prefers matrix-assisted laser desorption ionisation (MALDI) (Patterson & Aebersold 1995), then the polypeptide of interest needs to be in a form that can be deposited on a solid target. For this and other reasons, researchers who wish to apply proteomics to their research are advised to consult with proteomics experts during the design of their studies. Fortunately, many universities and research centers now have laboratories or services that support proteomics research. These facilities have the knowledge and equipment necessary to deal with protein samples from diverse biological sources.

BIOINFORMATICS SUPPORT

Proteomics and genomics research generates large data sets, which must be organized, stored, and made accessible in logical ways. One of the key components of genome and proteome research is bioinformatics. Several categories of bioinformatic tools are required for proteome analysis (Haoudi & Bensmail 2006). Briefly these can be classified into those used for quantification and those used for protein sequence analysis. With respect to quantification, a variety of software-based image analysis tools are available to monitor and quantify proteins separated by 2DE or to facilitate the quantification of isotope-labelled peptides. Protein sequence analysis depends upon a variety of analytical tools in order to search databases for peptide and protein matches, as well as to predict structure and function.

High quality and well-annotated genomics and protein databases are the core of proteome research. In most instances, the characterization and identification of proteins by a proteomic approach is dependent on the existence of genomic resources for the organism of interest, or at least for closely related organisms. When working with samples from many marine species, we are limited by the availability of genomics and proteomics resources for those species and, in many instances, even closely related species. For example,

López et al. (2001) compared differences in protein expression between intertidal and cultured populations of *Mytilus galloprovincialis* using high resolution 2DE. Over 750 proteins that were consistently expressed in foot tissues were observed in that study. From these, 92 proteins were selected for additional analysis and statistically significant differences in protein abundance for almost 50% of these proteins were identified.

In another study, a proteomic approach was used to generate proteomics reference maps and subsequently to detect, quantify, and compare the global protein expression between 2 related species of marine mussel, *Mytilus edulis* and *M. galloprovincialis*, growing in their respective geographical habitats (López et al. 2002). A comparative study of the protein profiles generated from analytical 2DE gels was performed, and changes in protein expression were analyzed quantitatively by computer analysis. On average 1278 proteins were detected per gel and, of these, 420 proteins were selected for quantification. Of these, 15 proteins showed higher expression in *M. edulis* and 22 proteins in *M. galloprovincialis*. The technique of peptide mass fingerprinting using MALDI-TOF (matrix assisted laser desorption ionization-time of flight) and/or nanoelectrospray MS/MS was then applied to identify these differentially expressed proteins. We were able to unambiguously identify only 15 of these 37 proteins using these techniques. Our results demonstrated the sensitivity of 2DE when detecting differences in protein expression. However, our ability to only identify 41% of these differentially expressed proteins revealed an important limitation with respect to protein identification using peptide mass fingerprinting, which is that proteins can only be identified if their sequence (or a sequence of the same protein from a closely related species) is available for comparison. The poor characterization of *Mytilus* spp. and other mollusk species at both the genomic and proteomic levels is responsible for our limited ability to identify these differentially expressed proteins. With the development of new analytical methods that enable de novo sequencing (nano-ESI [nanoelectrospray time of flight], Q-TOF [quadrupole time of flight] etc.), the application of proteomics to marine organisms will become more routine (López et al. 2002, 2005).

Proteomics also shows great promise with respect to the identification of protein markers that would allow for precise and rapid species identification (López et al. 2005). This would be especially beneficial for the identification of marine species that are difficult to identify using morphological characteristics or are ambiguous with respect to their taxonomic status (López et al. 2002, 2005). In previous studies, López et al. (2002, 2005) demonstrated how proteomics can be

used for the routine identification of species-specific peptides. Although we cannot foresee how many proteins are, in general, needed for the identification of closely related organisms, the high throughput and speed of analysis of the modern MALDI-TOF mass spectrometers would allow the extension of this kind of comparative study to include hundreds and even thousands of proteins from a large number of individuals, making the identification of species-specific peptide markers highly likely. Studies of this sort have the advantage that no information from genomic or proteomic databases is needed. Once potential species-specific peptides are identified, lower throughput MS techniques such as nano-ESI-IT MS could be used to perform a more detailed characterization of these markers. HPLC-tandem MS, focused on the peptide markers, may then be used for a fast and highly accurate confirmation of specific identification. Since this last technique is strictly quantitative, it might also be used as a routine technique for species identification. In addition, this technique could aid in the development of antibodies against species-specific peptides, which would allow the identification of these peptides in crude tissue extracts. Furthermore, this procedure (described by López et al. 2002) is also suitable for phylogenetic studies.

FUTURE CHALLENGES IN MARINE PROTEOMICS

One of the principal challenges in proteomics is to achieve a level of understanding of protein expression, post-translational modification and interaction that is similar in scope to what genomics has provided us for genes. This is a more difficult task with proteins than with nucleic acids because genes are approximately equimolecular in genomic DNA, whereas proteins may span 7 or 8 orders of magnitude in terms of functional abundance within a cell type, i.e. a functioning protein may be much less concentrated than other functioning proteins. In addition, there is difficulty in resolving very hydrophobic, very basic, or very large proteins using current 2DE systems. In proteomics, important discoveries will be made through quantitative observations of a limited (but large) number of protein gene products once the protein database is rich enough.

In response to technical challenges, we are likely to see the emergence of fully automated 2DE systems. Furthermore, continued development of non-gel-based alternative technologies that use combinations of capillary electrophoresis or multidimensional-HPLC coupled to MS will make proteomics data acquisition even more routine and possibly cost effective.

The major obstacle in the application of proteomics to many fields is generally considered to be data analy-

sis. However, with regard to the marine environment, a lack of genomic and proteomic resources for species of interest are often the major obstacle. Although it is possible to use data from related species, there are relatively few marine organisms for which sufficient genomic or proteomic data exist. This situation will likely improve in the near future owing to significant reductions in the time required for, and costs associated with, large-scale sequencing and proteomics studies. Such changes will make it economically feasible to begin to study a wider variety of organisms, including those from marine environments. The many genome projects that are planned or underway for numerous marine species will provide genomic resources that will greatly improve our ability to apply proteomics to the study of marine ecology. As mentioned previously, the development of reliable software tools that allow for the identification of proteins not represented in databases will greatly accelerate the rate at which proteomics is applied in marine science.

LITERATURE CITED

- Aebersold R (1993) Mass spectrometry of proteins and peptides in biotechnology. *Curr Opin Biotechnol* 4:412–419
- Anderson L, Seilhamer (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18:533–537
- Anderson NG, Anderson L (1982) The human protein index. *Clin Chem* 28:739–748
- Anderson NG, Anderson L (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 19:1835–1861
- Aquadro CF, Avise JC (1981) Genetic divergence between rodent species assessed by using two-dimensional electrophoresis. *Proc Natl Acad Sci USA* 78:3784–3788
- Barneah O, Benayahu Y, Weiss VM (2006) Comparative proteomics of symbiotic and aposymbiotic juvenile soft corals. *Mar Biotechnol* 8:11–16
- Beaumont AR (1994) Linkage studies in *Mytilus edulis*, the mussel. *Heredity* 72:557–562
- Carr SA, Annan RS, Huddleston MJ (2005) Mapping post-translational modifications of proteins by MS-based selective detection: application to phosphoproteomics. *Methods Enzymol* 405:82–115
- D'Ambrosio C, Arena S, Talamo F, Ledda L, Renzone G, Ferrara L, Scaloni A (2005) Comparative proteomic analysis of mammalian animal tissues and body fluids: bovine proteome database. *J Chromatogr B* 815:157–168
- Debouck C, Goodfellow PN (1999) DNA microarrays in drug discovery and development. *Nature Genet* 21:48–50
- Edwards Y, Hopkinson DA (1980) Are abundant proteins less variable? *Nature* 284:511–512
- Figey D, Aebersold R (1997) High sensitivity identification of proteins by electrospray ionization tandem mass spectrometry: initial comparison between an ion trap mass spectrometer and a triple quadrupole mass spectrometer. *Electrophoresis* 18:360–368
- Fuentes J, López JL, Mosquera E, Vázquez J, Villalba A, Álvarez G (2002) Growth, mortality, pathological conditions and protein expression of *Mytilus edulis* and *M. galloprovincialis* cosses cultured in the Ría de Arousa (NW of Spain). *Aquaculture* 213:233–251
- Haoudi A, Bensmail H (2006) Bioinformatics and data mining in proteomics. *Expert Rev Proteomics* 3:333–343
- Jungblut PR, Zimny-Arndt U, Zeindl-Eberhart E, Stulik J and 8 others (1999) Proteomics in human disease: cancer, heart and infectious diseases. *Electrophoresis* 8:1217–1242
- Liebler DC (2002) Introduction to proteomics. Tools for a new biology. Humana Press, Totowa, NJ
- López JL (2005) Role of proteomics in taxonomy: the *Mytilus* complex as a model of study. *J Chromatogr B* 815:261–274
- López JL, Mosquera E, Fuentes J, Marina A, Vázquez J, Álvarez G (2001) Two-dimensional gel electrophoresis of *Mytilus galloprovincialis*: differences in protein expression between intertidal and cultured mussels. *Mar Ecol Prog Ser* 224:149–156
- López JL, Marina A, Vázquez J, Álvarez G (2002) A proteomic approach to the study of the marine mussels *Mytilus edulis* and *M. galloprovincialis*. *Mar Biol* 141:217–223
- López JL, Lorenzo S, Fuentes J (2005) Proteomic approach to probe for larval proteins of the mussel *Mytilus galloprovincialis*. *Mar Biotechnol* 7:396–404
- Mann M, Wilm M (1995) Electrospray mass spectrometry for protein characterization. *Trends Biochem Sci* 20:219–224
- McDonagh B, Tyther R, Sheehan D (2006) Redox proteomics in the mussel, *Mytilus edulis*. *Mar Environ Res* 62: S101–S104
- Mosquera E, López JL, Álvarez G (2003) Genetic variability of the marine mussel *Mytilus galloprovincialis* assessed using two-dimensional electrophoresis. *Heredity* 90: 432–442
- Neel JV (1990) Average locus differences in mutability related to protein 'class': a hypothesis. *Proc Natl Acad Sci USA* 87:2062–2066
- O'Farrell (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250:4007–4021
- Olsson B, Bradley BP, Gilek M, Reiner O, Shepard JL, Tedengren M (2004) Physiological and proteomic responses in *Mytilus edulis* exposed to PCBs and PHAs extracted from Baltic Sea sediments. *Hydrobiologia* 514:15–27
- Patterson SD, Aebersold R (1995) Mass spectrometry approaches for the identification of gel-separated proteins. *Electrophoresis* 16:1791–1814
- Schweder T, Lindequist U, Lalk M (2005) Screening for new metabolites from marine microorganisms. In: Le Gal Y, Ulber R (eds) *Advances in biochemical engineering/biotechnology*, Vol 96. Springer, Berlin, p 1–23
- Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F and 5 others (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from 2 dimensional gels. *Proc Natl Acad Sci USA* 93: 14440–14445
- Wilkins MR, Sánchez JC, Gooley AA, Appel RD, Humphrey-Smith I, Hochstrasser DF, Williams KL (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13:19–50

Editorial responsibility: Howard Browman (Associate Editor-in-Chief), Storebø, Norway

Submitted: April 17, 2006; Accepted: November 6, 2006
Proofs received from author(s): February 5, 2007