

SHORT REVIEW

Application of proteomics to ecology and population biology

TL Karr

Department of Biology and Biochemistry, University of Bath, Bath, UK

Proteomics is a relatively new scientific discipline that merges protein biochemistry, genome biology and bioinformatics to determine the spatial and temporal expression of proteins in cells, tissues and whole organisms. There has been very little application of proteomics to the fields of behavioral genetics, evolution, ecology and population dynamics, and has only recently been effectively applied to the closely allied fields of molecular evolution and genetics. However, there exists considerable potential for proteomics to impact in areas related to functional ecology; this review will introduce the general concepts and methodologies that define the field of proteomics and compare and contrast the advantages and disadvantages with other methods. Examples of how proteomics can aid, complement and indeed extend the study of functional ecology will be discussed including the main tool of ecological studies,

population genetics with an emphasis on metapopulation structure analysis. Because proteomic analyses provide a direct measure of gene expression, it obviates some of the limitations associated with other genomic approaches, such as microarray and EST analyses. Likewise, in conjunction with associated bioinformatics and molecular evolutionary tools, proteomics can provide the foundation of a systems-level integration approach that can enhance ecological studies. It can be envisioned that proteomics will provide important new information on issues specific to metapopulation biology and adaptive processes in nature. A specific example of the application of proteomics to sperm ageing is provided to illustrate the potential utility of the approach.

Heredity (2008) 100, 200–206; doi:10.1038/sj.hdy.6801008; published online 15 August 2007

Keywords: proteomics; ecology; population genetics

Introduction

The term 'proteomics' refers to the biochemical identification of all proteins, including all forms of post-translational modifications, in living cells, tissues, organs or whole organisms. In its fullest implementation, proteomics refers to all protein forms including covalent and non-covalent interactions (for example, phosphorylation, acetylation, and methylation), spatial and temporal distributions and how their temporal dynamic patterns are affected by alterations in the extracellular and intracellular environments. Such wide ranging and all inclusive definitions are easily stated, but virtually impossible to attain using current technologies. However, recent advances in genome sequencing and instrumentation that can handle and analyze complex protein mixtures are providing significant advances in the analysis of cellular proteomes.

The evolutionary and ecological 'value' of proteins

Although the basic blueprint of life is encoded in DNA, the execution of the genetic plan is carried out by the activities of proteins, the complexes they form with lipids, sugars and nucleic acids and the higher order

structures they form. The fabric of life is therefore protein-based and it follows that selection acts on the structures the proteins form. However, with the recent emergence of molecular and bioinformatic analyses of whole genomes, selection at the molecular level is measured almost exclusively by analysis of DNA sequence variation. This is of course understandable, as both theory and practice of the molecular analysis of DNA sequences are mature fields and this approach has been extremely successful. However, there are instances where genomic analyses might not suffice, particularly for studies of the vast majority of species for which sequenced genomes are not yet available.

For example, from a physiological and functional perspective, proteins that comprise skin and other epithelial linings of the body interface directly with the environment, and therefore are prime targets of selection. In this regard, it is reasonable to assume that the tracheal system of the fruit fly *Drosophila*, well understood at the molecular genetic level, might provide the foundation for further studies in field populations where segregating alleles of genes involved in tracheal function might reflect the signature of selection. Given the importance of trachea in the transport of gases to animal tissues in insects (Lubarsky and Krasnow, 2003), selection at this level could influence the evolution of organisms, populations and ecological systems they inhabit. However appealing this possibility appears, a PubMed search for keywords 'Drosophila, trachea and proteomics' returned no references. Indeed, a search using keywords, 'ecology and proteomics' retrieved fewer than 100 entries, thus underscoring the state of proteomics in ecology and population

Correspondence: Dr TL Karr, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK.

E-mail: T.L.Karr@bath.ac.uk

Received 15 September 2006; revised 16 April 2007; accepted 19 April 2007; published online 15 August 2007

studies. The purpose here is to provide an overview of the burgeoning new field of proteomics and to examine the possible application of proteomics approaches to population biology and related studies.

If proteomics is to contribute to functional ecology, knowledge of power and limitation of this technique will be critical for exploiting it. Judicious use of proteomics could, in principle, identify a large suite of specific protein targets related to the organism of study, thus providing a protein phenotype. This snapshot of protein expression could then serve as a foundation on which to test specific hypotheses and/or further characterize the population of interest. Because many ecological studies involve dioecious populations, one cell type amenable to proteomic studies is sperm. As discussed further below, one sperm proteome has been described in detail providing a nearly complete database for future studies. Our recent work on a proteomic analysis of sperm ageing will then be described and examine its potential use in determining population structure.

Principles of proteomics

The main biochemical tool used in proteomics is mass spectrometry (MS). In many instances two-dimensional gel electrophoresis (2DGel) is a useful adjunct to MS. Each technique has advantages and disadvantages that must be weighed to ensure proper application to a particular biological problem.

2DGel is an established and long-standing biochemical technique. 2DGel separates proteins based on isoelectric point in the first dimension and molecular weight in the second, resulting in a pattern of protein 'spots' (Figure 1). Since its introduction over 30 years ago (O'Farrell, 1975), 2DGel has been the mainstay in the analysis of complex protein mixtures. However, because 2DGel is essentially a separation technique based on biochemical properties of polypeptide chains, amino acid sequences needed to identify proteins cannot be identified using only this technique. 2DGel is most useful when used in conjunction with MS. For example, protein spots excised from 2D gels can be identified using matrix assisted laser desorption/ionization-time of flight (MALDI-TOF), a commonly used MS technique (discussed below). Although limited to the number of protein spots that can be reproducibly resolved on a gel, 2DGel/MALDI-TOF is a rapid and reliable approach successfully applied to a variety of systems.

The application of MS-based techniques for the qualitative and quantitative analysis of cellular proteomes, often from complex mixtures, has become an important tool for the understanding of cellular function across a wide diversity of plant and animal taxa (Fels *et al.*, 2003; Han and Lee, 2003; Ahram and Springer, 2004; Aldred *et al.*, 2004; Banta-Wright and Steiner, 2004; Chen and White, 2004; Baginsky and Gruissem, 2006; Berven *et al.*, 2006; Bowler *et al.*, 2006; Hitchen and Dell, 2006) Historically, MS has been a workhorse analytic tool used by chemists and associated industries, where it is mainly used as a tool for structural determination of organic compounds and quality control assurance in the pharmaceutical and biotech industries. MS is also a prominent tool used by security agencies, for example in rapid detection of explosive materials at airports (Takats *et al.*, 2004, 2005a,b) Over the past two decades, a

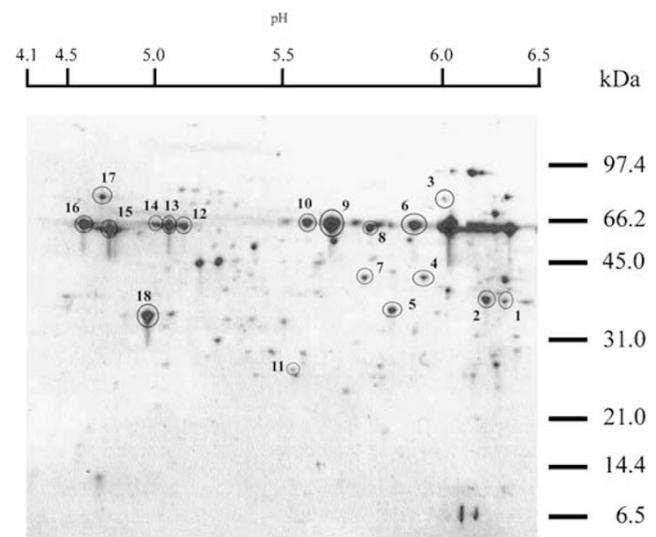


Figure 1 Two-dimensional gel electrophoresis of *D. melanogaster* sperm. A typical 2D gel of purified sperm run on a pH 4–7 first dimension isoelectric focusing strip and subsequently run on a 12.5% SDS-PAGE gel. Circles indicate those spots excised and subjected to MALDI-TOF/MS. Peptide mass fingerprinting was used to assign peptide masses to *in silico* predicted peptides in the *D. melanogaster* genome. Spots 9, 10, 12, and 13 were identified by GO ontology as leucylaminopeptidases. Spot 18 was identified as *D. melanogaster* Porin 2 (see Figure 4, below).

number of major advances in MS technologies applied to the analysis of proteins and peptides has elevated MS to a central role in proteomics, including new methods for the introduction of complex protein and peptide mixtures into the mass spectrometer in such a manner as to allow for the *de novo* sequencing of small peptides (Lin *et al.*, 2003; Yates, 2004; Gingras *et al.*, 2005; Guerrero and Kleiner, 2005; Domon and Aebersold, 2006) Combined with the ever-expanding availability of whole, or partial, genome sequences and bioinformatic tools for their analysis, MS has assumed a primary role in proteomic analyses. In this regard, MS, genomic sequencing and bioinformatics exist in symbiosis, depending and benefiting from advances in each.

Genomic sequencing methods and their application to evolution and ecology are well known and will not be further discussed. Bioinformatics analysis of MS data is a new and burgeoning field that has been central to genomics and proteomics (Cristoni and Bernardi, 2004). Bioinformatic tools and methods for data management, data mining and data interpretation are continuing to develop and interface with computer-based tools for automated MS data analysis. Below a brief review of these topics will be provided before discussion of their application to the areas of functional ecology and population biology. There are a number of excellent reviews on recent developments in MS and allied bioinformatic analyses, and their application to proteomics to which the reader is referred for additional information (Aebersold and Mann, 2003; Yates, 2004).

The basic principles of mass spectrometry in proteomics is a fairly straightforward proposition- identify as many peptide fragments derived from the proteins of interest and use this information to identify the genes that encode them. Generally, strategies have been developed that allow high-throughput identification of

proteins (via their peptide fragments, see below) from highly complex mixtures. These strategies include, among others, reliable ways to generate proteins digests and deliver them to the MS machine for analysis (Nagele *et al.*, 2004). MS relies almost exclusively on its ability to very accurately measure peptide masses, something that a new generation of instrumentation can do with remarkable precision (Zimmer *et al.*, 2006).

Mass spectrometers basically measure the m/z ratio (mass/charge) and their utility in proteomics comes from their ability to measure m/z with astounding accuracy (<1 ppm in the best case scenario). This level of accuracy allows assignment of individual amino acid composition to the peptide of interest, thus making possible an assignment to a unique gene that contains this peptide. The two most commonly used techniques are MALDI and electrospray ionization (ESI). Both methods use so-called 'soft' ionization techniques, but differ in important ways for which the reader can acquire additional information in a variety of recent excellent reviews (Aebersold and Mann, 2003; Yates, 2004; Gingras *et al.*, 2005; Guerrero and Kleiner, 2005; Doman and Aebersold, 2006) Suffice it to say that both techniques produce ionized peptides in the gas phase, which enter a mass analyzer that measures the m/z ratio.

For most proteomics studies, four types of mass analyzers are currently in use: ion trap, quadrupole, TOF and FT-ICR (Fourier transform ion cyclotron resonance). Although all four differ in accuracy, sensitivity and resolution, they are all suitable for ecological and evolutionary studies. One advantage of the ion trap instrument is the ability to not only determine the mass of a given peptide, but also to obtain sequence information. Ions, selected for specific m/z ratios, can be fragmented by collision with an inert gas, a process called collision-induced fragmentation (CID). The fragment masses are then measured and they then represent a so-called tandem mass (MS/MS) spectrum and the sequence of the original peptide can be deduced if enough high-quality fragment MS/MS spectra are obtained. Using this system, the sequence information containing MS/MS spectra is compared against comprehensive protein sequence databases using search algorithms designed to match MS spectral data to a database generated *in silico* by computation from known sequence databases (Pappin *et al.*, 1993; Perkins *et al.*, 1999).

MALDI-TOF is one of the most widely used MS systems for proteomic studies (Henzel *et al.*, 2003; Marvin *et al.*, 2003; Ragoussis *et al.*, 2006). This approach, although somewhat limited due to the analysis of relatively pure samples (for example, spots extracted from 2D gels), is highly advantageous due to its relatively low cost, simplicity in design and sensitivity. Unlike the ion-trap/CID systems, MALDI-TOF is unable to directly provide sequence information but instead identifies peptide mass fingerprints (PMFs) that can match a computer database of predicted peptide masses (Pappin *et al.*, 1993).

Application of proteomics to phylogenetics and population biology

Sperm proteomics and population biology

Recently, MS and 2DGel was used to elucidate the *Drosophila* sperm proteome (DmSP) (Dorus *et al.*, 2006).

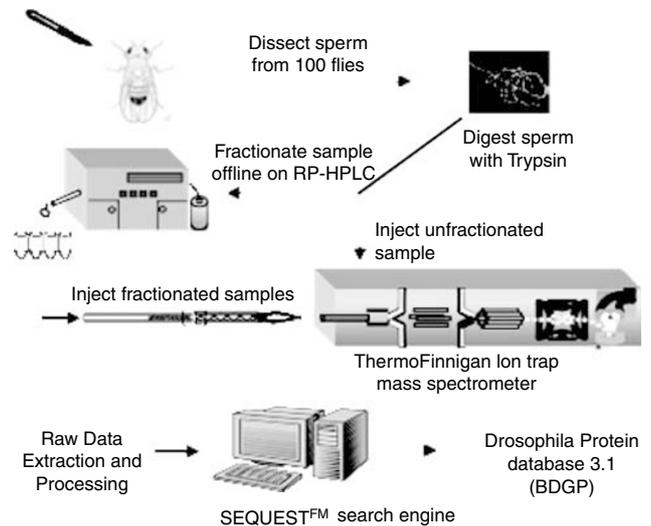


Figure 2 General approach for the identification of the *D. melanogaster* sperm proteome using electrospray MS/MS. Purified sperm are subjected to trypsin cleavage that generates peptides with defined cleavage sites (Arg/Lys). Samples are then either directly injected into the nanoelectro nozzle and volatilized peptides of defined molecular weight ranges chosen and then their m/z values determined. Alternatively, for greater coverage, trypsinized peptides can be fractionated by high-pressure liquid chromatography before introduction to the mass spectrometer.

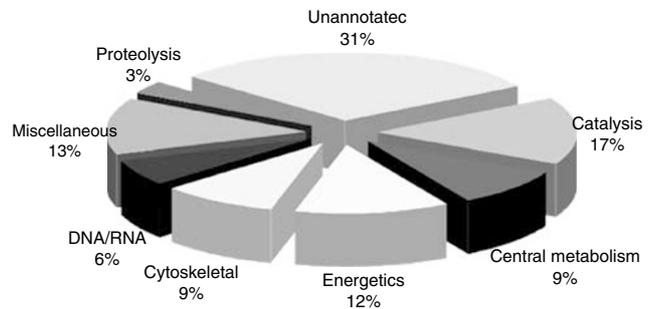


Figure 3 Functional categories of the DmSP based on Gene Ontology annotation. Eight functional categories were found by querying FlyBase for DmSP proteins. Of the seven categories, the single largest category of proteins (31%) has no predicted function.

This study identified over 380 integral sperm proteins spanning numerous, and in some cases, unexpected, functional categories (Figures 2 and 3). For example, 3% of the DmSP encodes putative proteases/peptidases, including a family of recently evolved aminopeptidases (S Dorus and TL Karr, unpublished results). Sperm aminopeptidases were identified from 2D gel spots (Figure 1). MALDI-TOF identified aminopeptidases in four major spots (Figure 1, spots 9, 10, 12 and 13), which, by comparison to the α - and β -tubulins (spots 15 and 16), comprise a significant proportion of sperm protein composition (U Gerike and TL Karr, unpublished results). Thus, 2DGel/MALDI-TOF can provide both qualitative and quantitative information on cellular proteomes.

Estimates using 2DGel suggest that more than 90% of the sperm proteome were identified by MS, suggesting a near-completed proteome (U Gerike and TL Karr, unpublished results), thus paving the way for whole cell proteomic studies of spermatozoa. Although knowledge

of the sperm proteome will undoubtedly yield valuable new insights into sperm biology and evolution, one immediate and striking characteristic of the DmSP is its relative 'simplicity' (Figure 1). Exhaustive counts from numerous 2D gels run under a variety of conditions consistently result in no greater than 500 spots, and as is often the case, spot counts generally overestimate the number of unique polypeptides present in the sample (Gorg *et al.*, 2004). The restricted number of proteins present in sperm is in accord with its highly specialized function as an engine carrying DNA cargo with a very specific mission to find an egg and fuse with it. However, the proteomic complexity of other metazoan spermatozoa remains to be determined, and others have estimated mammalian sperm proteomes at greater than 1000 proteins.

Sperm are well known for their capacity to interact with male seminal fluids and with factors present in the female reproductive tract (Setchell *et al.*, 1993; Lung *et al.*, 2002; Heifetz and Wolfner, 2004; Gage, 2005; Robertson, 2005; Troedsson *et al.*, 2005). Many of these studies reveal a myriad of sperm/male accessory gland interactions and in some cases hundreds of potential proteins identified, particularly in mammalian systems (Pilch and Mann, 2006; Walker *et al.*, 2006). Indeed, proteomic studies of sperm interactions during transit through the female oviduct has begun to reveal the inductive effects of spermatozoa on female gene expression (Fazeli *et al.*, 2004). These studies will all benefit from specific knowledge of the core sperm proteomes, and in the case of *Drosophila*, knowledge of the DmSP opens up numerous avenues of research into the molecular details of sperm-male accessory gland, sperm female and sperm egg interactions.

Phylogenetic considerations

Proteomics has been used sparingly in phylogenetic studies. However, it is becoming increasingly clear that proteomics will become a major player in these areas as it continues to mature as a science and as technological improvements bring down costs. While there are vanishingly few examples of the application of proteomics to population biology studies, the ones that have recently appeared serve as exemplars of the potential power of the approach. Proteomics has been successfully applied to phylogenetic analysis of plants (Thiellement *et al.*, 1999, 2002).

Phylogenetic analyses are often frustrated by issues relating to mono-, para- and polyphyly (Navas and Albar, 2004). Proteomics could be a valuable tool in establishing phylogenetic relationships as analysis of protein variation can be applied to any organism where access to tissue is available (Biron *et al.*, 2006). Phylogenetic analyses based on protein sequences have decided advantages over DNA-based methods, but has not been widely utilized due to experimental difficulties of protein extraction, isolation and purification. Protein-based phylogenies also, ironically, rely upon genomic information in related taxa for meaningful comparisons. For example, knowledge of a sperm proteome from *D. melanogaster*, made by direct MS approaches (Dorus *et al.*, 2006) is being used to datamine all orthologous sequences from the genomes of other closely related *Drosophila* species (Consortium, 2003). However, applica-

tion of MS to define the proteomes in these species will also provide valuable new information about the protein composition in each species, something that BLAST is unable to accomplish directly. This 'differential MS' approach could, in principle, immediately identify protein differences related to population-wide processes such as population fragmentation, reproductive isolation or speciation.

The enormous complexity of proteomes and the highly technical aspects and costs of MS has hindered broader application of these techniques and technologies. It is also clear that much work must be done in both the practical and theoretical arenas before proteomics can be successfully applied to large-scale phylogenetic analyses based on proteomes. The entire proteomics field is struggling to standardize methodologies used, and the reporting of results (Orchard *et al.*, 2004). MS is a highly sensitive technique and can generate enormous volumes of spectral data that must be managed and analyzed properly. As the field grapples with, and overcomes, these technical issues, application of proteomics to the analysis of protein composition of related taxa should become more common place. For example, application of bioinformatic and molecular evolutionary analysis of MS data using known amino acid substitution patterns can now identify putative homologous proteins from organisms lacking a sequenced genome (Liska and Shevchenko, 2003; Dworzanski *et al.*, 2006).

Population biology considerations

A population of organisms at any given moment reflects the summed selective forces acting upon them, and population structure is determined by a complex array of factors including birth/death rates, growth rates and population movements (influx and efflux). As such, populations can reflect the overall history of adaptations and evolutionary forces that have shaped them, and is often the starting point for functional ecological studies. Thus, measurement of genetic variation and the estimation of fitness differences between genes and genotypes in populations are the main tools for studying the spatial structure of populations and the level of gene flow between them.

Measurement of natural variation within and between populations is an essential tool used by population and conservation biologists. A good example of the application of proteomics to this area was a recent 2DGel and MALDI-TOF/MS study of natural variation among eight *Arabidopsis* ecotypes (Chevalier *et al.*, 2004). This study identified 49 proteins of *Arabidopsis* root cells from 2D gels. Analysis by MALDI-TOF/MS revealed clear ecotype differences in both relative protein abundance and isoelectric points. Interestingly, high variability in the major spots (indicating the more highly expressed proteins) was observed, suggesting that ecotype variation is pervasive and highly complex at the protein level. Similarly, 2D gels were used to study the adaptive responses of 11 wheat populations growing in France to differing environmental conditions (Thiellement *et al.*, 1999). This pioneering work established the validity of the technique and showed that population variation among closely related strains could be identified.

Population proteomics in the analysis of genetic polymorphisms has been limited generally as it can only

determine qualitative trait differences (presence/absence). However, newer technologies and advancements in protein quantification of 2D gels and MS spectral data can allow for the detection of even minor protein differences (Blonder *et al.*, 2004; Whitelegge, 2004; Aggarwal *et al.*, 2006; Fricker *et al.*, 2006). However, as these studies demonstrate, additional statistical and technological tools are needed before the full potential of 2DGel and MS technologies are realized and can provide an objective and reliable approach to the study of genetic divergence. Advances in these areas are crucial for proteomic studies due to the enormous complexity exhibited by most cellular proteomes. Until data management tools and appropriate software are developed, proteomic applications to population biology, functional ecology and evolution will be most effective in those systems tractable using current technologies, such as those used in the above cited cases of *Arabidopsis* and wheat, that examined relatively 'simple' systems, where extensive data of the experimental system was readily available. Another such system which should be amenable to many ecological and field biology studies is described below along with an example of potential application to the study of population age structure.

A specific example of the application of sperm proteomics to the study of population age structure

Both intrinsic and extrinsic factors affect male fertility. The effect of male age has been well studied in numerous systems, with a variety of factors involved that affect male fertility (Spencer *et al.*, 2003; Jones and Elgar, 2004; Hauser, 2006; Hellstrom *et al.*, 2006). Genes that affect the ageing process in a variety of laboratory animals have been identified over the past decade (Guarente and Kenyon, 2000). However, studies of the effect of male age on the sperm proteome have not yet been reported. We therefore undertook an examination of the 2DGel patterns of *D. melanogaster* sperm proteins during ageing (Figure 4). As expected, similar patterns on 2D gels were observed at different ages (Figures 4a and b). However, further examination clearly identified one prominent spot showing a consistent increase in expression levels during ageing (circled in Figure 4c). Spot intensities are determined using software-based programs that analyze scanned images using a suitably sensitive digital camera of high resolution and bit depth (Figure 4d). Owing to technical limitations of 2DGel, particularly reproducibility in the solubilization and isoelectric focusing steps, comparison of between-experiment spot intensities can vary significantly. Thus, it is common practice to perform statistical tests (for example, *t*-test) of scanned spots from multiple gels to establish statistical confidence in any differences detected.

The spot identified (Figure 4c) following excision from the gel, trypsin digestion and MALDI-TOF analysis as the *Drosophila* gene *Porin 2* (FBgn0051722). *Porin 2* is an outer mitochondrial membrane voltage-gated ion channel protein (Guarino *et al.*, 2006) and homologues have been described in mosquitos (Sardiello *et al.*, 2003). Quantitative analysis suggests that *Porin 2* protein expression in sperm increases approximately 2 to 3-fold over this time period and that *Porin 2* levels begin to measurably increase by 6–7 days (K Steeds and TL Karr, unpublished results). These Coomassie-stained gels also show that *Porin 2* is a major component of sperm

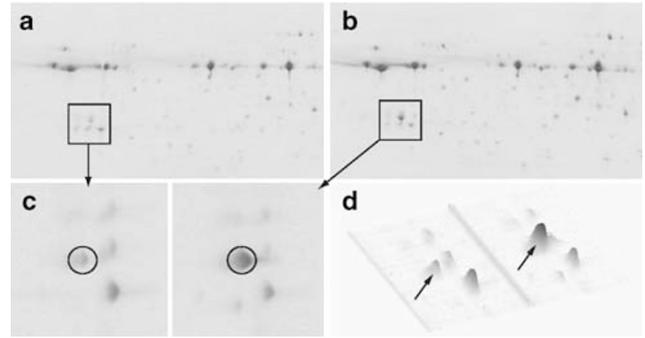


Figure 4 Age-related changes in sperm proteome. Coomassie-stained 2D gels of total sperm proteins from 6 day old (a) and 30 day old (b). Boxed regions were further compared as indicated suggesting that the levels of one major protein had increased during ageing (c, circles). Note that the boxed regions in (c) are rotated 90° counterclockwise. (d) Quantitative changes in protein levels were assessed by comparing surface scans of this region using NIH Image or other software designed to measure intensity differences. Measurement of peak heights (arrows) indicated an approximately 3-fold increase in *Porin 2* levels.

therefore raising the possibility that its increased expression during ageing may be related to its function in older sperm. *Porin 2* has been identified on 2D gels from numerous other *Drosophila* species (K Steeds and TL Karr, unpublished) and it will be of interest to see if increased expression of this protein occurs across these lineages. If so, *Porin 2* may be a prime candidate for a number of population biology and ecological studies in parallel with further functional studies in *D. melanogaster*. This single example serves to illustrate the potential inherent in a thorough analysis of all proteins in the proteome to reveal additional differences either within or between species useful for future studies of population structure and/or reproductive fitness factors operating in natural populations.

Future directions

The study of the interrelationships between genetics and environmental factors is the foundation of population biology and functional ecology. How proteomics will influence future studies in these areas depends critically on the continued development of analytical MS and 2DGel equipment and in the area of equipment costs. For example, standards regarding the analysis and reporting of high-throughput MS proteomics database must be established (reference). Likewise web-based tools tailored to comparative proteomics at the population level are not yet available, limiting the range and types of studies possible. For example, current ongoing large-scale studies of human population variation at the protein level would greatly benefit by integration with the HapMap project (reference), which contains a very large and growing database of human polymorphisms. Until then, most population biology and functional ecological studies at the protein level will most likely involve target analysis of smaller subsets of protein differences such as described above for the *Porin 2* case. Given that population biologists are funded by the NSF or USDA generally, which can offer only limited financial support, 2DGel/MALDI-TOF will most likely remain the

method of choice for many studies. It is expected that, as with PCR and DNA sequencing, costs incurred in high-throughput proteomics studies involving ion trap and Q-TOF will eventually drop and be followed by a concomitant increase in the use of these powerful technologies.

A recent study (Schmidt *et al.*, 2005) provides one example of an experimental system that should be amenable to proteomic analysis. This study examined the effects of diapause on reproduction in *D. melanogaster*. This form of reproductive quiescence is in response to reduced temperatures and/or shortened photoperiods, and natural populations expression of reproductive diapause was correlated with their latitudinal origin. The potential fitness trade-offs between diapause and non-diapause populations (or individuals within a population) in the observed clines were reflected by distinct life history traits (for example, life span, age-specific mortality rates, fecundity profiles and lipid content). The nature of the genetic variances and covariances between these lines is unknown. One possible way to extend these studies would be an examination of the proteomic profiles of sperm (or other suitable tissues). Variation in either quality or quantity of the sperm proteomes in lines that display diapause versus those that do not might provide genetic clues to the nature of the variations present in these populations.

Although the above study focused on *Drosophila*, it should be noted that MS can be applied to any organism for which suitable samples of interest can be obtained. Of course, until more whole genome sequences become available, model systems such as *Drosophila*, *Caenorhabditis elegans*, *Arabidopsis*, mouse and humans will remain systems in which proteomics is most conveniently applied. However, whole genome sequences are not an absolute necessity for proteomic studies. For example, EST libraries of specific tissue types (from any organism) can be an important starting point for proteomic analyses. Further, there are currently many genome projects underway for previously unsequenced species and new technologies are further enabling the rapid acquisition of additional genome sequences. Thus, non-model systems no longer represent an impenetrable barrier to population-level proteome analyses, and MS will continue to develop as an important tool for systems-level studies of population structure and functional ecology.

Acknowledgements

I am grateful to Dr U Gerike and K Steeds for providing expert technical assistance. This work was funded in part by grants to TLK from the Royal Society (Wolfson Merit Award) and the Biotechnology and Biological Sciences Research Council (BB/C007670/1).

References

- Aebersold R, Mann M (2003). Mass spectrometry-based proteomics. *Nature* **422**: 198–207.
- Aggarwal K, Choe LH, Lee KH (2006). Shotgun proteomics using the iTRAQ isobaric tags. *Brief Funct Genomic Proteomic* **5**: 112–120.
- Ahram M, Springer DL (2004). Large-scale proteomic analysis of membrane proteins. *Expert Rev Proteomics* **1**: 293–302.
- Aldred S, Grant MM, Griffiths HR (2004). The use of proteomics for the assessment of clinical samples in research. *Clin Biochem* **37**: 943–952.
- Baginsky S, Gruissem W (2006). Arabidopsis thaliana proteomics: from proteome to genome. *J Exp Bot* **57**: 1485–1491.
- Banta-Wright SA, Steiner RD (2004). Tandem mass spectrometry in newborn screening: a primer for neonatal and perinatal nurses. *J Perinat Neonatal Nurs* **18**: 41–58; quiz 59–60.
- Berven FS, Flikka K, Berle M, Vedeler C, Ulvik RJ (2006). Proteomic-based biomarker discovery with emphasis on cerebrospinal fluid and multiple sclerosis. *Curr Pharm Biotechnol* **7**: 147–158.
- Biron DG, Loxdale HD, Ponton F, Moura H, Marche L, Brugidou C *et al.* (2006). Population proteomics: an emerging discipline to study metapopulation ecology. *Proteomics* **6**: 1712–1715.
- Blonder J, Conrads TP, Veenstra TD (2004). Characterization and quantitation of membrane proteomes using multidimensional MS-based proteomic technologies. *Expert Rev Proteomics* **1**: 153–163.
- Bowler RP, Ellison MC, Reisdorph N (2006). Proteomics in pulmonary medicine. *Chest* **130**: 567–574.
- Chen WG, White FM (2004). Proteomic analysis of cellular signaling. *Expert Rev Proteomics* **1**: 343–354.
- Chevalier F, Martin O, Rofidal V, Devauchelle AD, Barteau S, Sommerer N *et al.* (2004). Proteomic investigation of natural variation between Arabidopsis ecotypes. *Proteomics* **4**: 1372–1381.
- Consortium TF (2003). The FlyBase database of the Drosophila genome projects and community literature. *Nucleic Acids Research* **31**: 172–175.
- Cristoni S, Bernardi LR (2004). Bioinformatics in mass spectrometry data analysis for proteomics studies. *Expert Rev Proteomics* **1**: 469–483.
- Domon B, Aebersold R (2006). Mass spectrometry and protein analysis. *Science* **312**: 212–217.
- Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL (2006). Genomic and functional evolution of the Drosophila melanogaster sperm proteome. *Nat Genet* **38**: 1440–1445.
- Dworzanski JP, Deshpande SV, Chen R, Jabbour RE, Snyder AP, Wick CH *et al.* (2006). Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification. *J Proteome Res* **5**: 76–87.
- Fazeli A, Affara NA, Hubank M, Holt WV (2004). Sperm-induced modification of the oviductal gene expression profile after natural insemination in mice. *Biol Reprod* **71**: 60–65.
- Fels LM, Buschmann T, Meuer J, Reymond MA, Lamer S, Rocken C *et al.* (2003). Proteome analysis for the identification of tumor-associated biomarkers in gastrointestinal cancer. *Dig Dis* **21**: 292–298.
- Fricker LD, Lim J, Pan H, Che FY (2006). Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. *Mass Spectrom Rev* **25**: 327–344.
- Gage MJ (2005). Reproductive biology: direct delivery of costly sex peptides. *Curr Biol* **15**: R124–R126.
- Gingras AC, Aebersold R, Raught B (2005). Advances in protein complex analysis using mass spectrometry. *J Physiol* **563**: 11–21.
- Gorg A, Weiss W, Dunn MJ (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**: 3665–3685.
- Guarente L, Kenyon C (2000). Genetic pathways that regulate ageing in model organisms. *Nature* **408**: 255–262.
- Guarino F, Specchia V, Zapparoli G, Messina A, Aiello R, Bozzetti MP *et al.* (2006). Expression and localization in spermatozoa of the mitochondrial porin isoform 2 in *Drosophila melanogaster*. *Biochem Biophys Res Commun* **346**: 665–670.
- Guerrera IC, Kleiner O (2005). Application of mass spectrometry in proteomics. *Biosci Rep* **25**: 71–93.
- Han MJ, Lee SY (2003). Proteome profiling and its use in metabolic and cellular engineering. *Proteomics* **3**: 2317–2324.

- Hauser R (2006). The environment and male fertility: recent research on emerging chemicals and semen quality. *Semin Reprod Med* **24**: 156–167.
- Heifetz Y, Wolfner MF (2004). Mating, seminal fluid components, and sperm cause changes in vesicle release in the *Drosophila* female reproductive tract. *Proc Natl Acad Sci USA* **101**: 6261–6266.
- Hellstrom WJ, Overstreet JW, Sikka SC, Denne J, Ahuja S, Hoover AM *et al.* (2006). Semen and sperm reference ranges for men 45 years of age and older. *J Androl* **27**: 421–428.
- Henzel WJ, Watanabe C, Stults JT (2003). Protein identification: the origins of peptide mass fingerprinting. *J Am Soc Mass Spectrom* **14**: 931–942.
- Hitchen PG, Dell A (2006). Bacterial glycoproteomics. *Microbiology* **152**: 1575–1580.
- Jones TM, Elgar MA (2004). The role of male age, sperm age and mating history on fecundity and fertilization success in the hide beetle. *Proc Biol Sci* **271**: 1311–1318.
- Lin D, Tabb DL, Yates JRr (2003). Large-scale protein identification using mass spectrometry. *Biochim Biophys Acta* **1646**: 1–10.
- Liska AJ, Shevchenko A (2003). Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications. *Proteomics* **3**: 19–28.
- Lubarsky B, Krasnow MA (2003). Tube morphogenesis: making and shaping biological tubes. *Cell* **112**: 19–28.
- Lung O, Tram U, Finnerty CM, Eipper-Mains MA, Kalb JM, Wolfner MF (2002). The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. *Genetics* **160**: 211–224.
- Marvin LF, Roberts MA, Fay LB (2003). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. *Clin Chim Acta* **337**: 11–21.
- Nagele E, Vollmer M, Horth P, Vad C (2004). 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. *Expert Rev Proteomics* **1**: 37–46.
- Navas A, Albar JP (2004). Application of proteomics in phylogenetic and evolutionary studies. *Proteomics* **4**: 299–302.
- Orchard S, Taylor C, Hermjakob H, Zhu W, Julian R, Apweiler R (2004). Current status of proteomic standards development. *Expert Rev Proteomics* **1**: 179–183.
- O'Farrell PH (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**: 4007–4021.
- Pappin DJ, Hojrup P, Bleasby AJ (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **3**: 327–332.
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**: 3551–3567.
- Pilch B, Mann M (2006). Large-scale and high-confidence proteomic analysis of human seminal plasma. *Genome Biol* **7**: R40.
- Ragoussis J, Elvidge GP, Kaur K, Colella S (2006). Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. *PLoS Genet* **2**: e100.
- Robertson SA (2005). Seminal plasma and male factor signalling in the female reproductive tract. *Cell Tissue Res* **322**: 43–52.
- Sardiello M, Tripoli G, Oliva M, Santolamazza F, Moschetti R, Barsanti P *et al.* (2003). A comparative study of the porin genes encoding VDAC, a voltage-dependent anion channel protein, in *Anopheles gambiae* and *Drosophila melanogaster*. *Gene* **317**: 111–115.
- Schmidt PS, Paaby AB, Heschel MS (2005). Genetic variance for diapause expression and associated life histories in *Drosophila melanogaster*. *Evolution Int J Org Evolution* **59**: 2616–2625.
- Setchell BP, Sanchez-Partida LG, Chairussyuhur A (1993). Epididymal constituents and related substances in the storage of spermatozoa: a review. *Reprod Fertil Dev* **5**: 601–612.
- Spencer CC, Howell CE, Wright AR, Promislow DE (2003). Testing an 'aging gene' in long-lived *drosophila* strains: increased longevity depends on sex and genetic background. *Aging Cell* **2**: 123–130.
- Takats Z, Cotte-Rodriguez I, Talaty N, Chen H, Cooks RG (2005a). Direct, trace level detection of explosives on ambient surfaces by desorption electrospray ionization mass spectrometry. *Chem Commun (Camb)* **15**: 1950–1952.
- Takats Z, Wiseman JM, Cooks RG (2005b). Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology. *J Mass Spectrom* **40**: 1261–1275.
- Takats Z, Wiseman JM, Gologan B, Cooks RG (2004). Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* **306**: 471–473.
- Thiellement H, Bahrman N, Damerval C, Plomion C, Rossignol M, Santoni V *et al.* (1999). Proteomics for genetic and physiological studies in plants. *Electrophoresis* **20**: 2013–2026.
- Thiellement H, Zivy M, Plomion C (2002). Combining proteomic and genetic studies in plants. *J Chromatogr B Analyt Technol Biomed Life Sci* **782**: 137–149.
- Troedsson MH, Desvougues A, Alghamdi AS, Dahms B, Dow CA, Hayna J *et al.* (2005). Components in seminal plasma regulating sperm transport and elimination. *Anim Reprod Sci* **89**: 171–186.
- Walker MJ, Rylett CM, Keen JN, Audsley N, Sajid M, Shirras AD *et al.* (2006). Proteomic identification of *Drosophila melanogaster* male accessory gland proteins, including a pro-cathepsin and a soluble gamma-glutamyl transpeptidase. *Proteome Sci* **4**: 9.
- Whitelegge JP (2004). Mass spectrometry for high throughput quantitative proteomics in plant research: lessons from thylakoid membranes. *Plant Physiol Biochem* **42**: 919–927.
- Yates JRr (2004). Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct* **33**: 297–316.
- Zimmer JS, Monroe ME, Qian WJ, Smith RD (2006). Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom Rev* **25**: 450–482.