

REVIEW

Open Access

Proteomics: a subcellular look at spermatozoa

Stefan S du Plessis¹, Anthony H Kashou², David J Benjamin², Satya P Yadav³, Ashok Agarwal^{2*}

Abstract

Background: Male-factor infertility presents a vexing problem for many reproductively active couples. Many studies have focused on abnormal sperm parameters. Recent advances in proteomic techniques, especially in mass spectrometry, have aided in the study of sperm and more specifically, sperm proteins. The aim of this study was to review the current literature on the various proteomic techniques, and their usefulness in diagnosing sperm dysfunction and potential applications in the clinical setting.

Methods: Review of PubMed database. Key words: spermatozoa, proteomics, protein, proteome, 2D-PAGE, mass spectrometry.

Results: Recently employed proteomic methods, such as two-dimensional polyacrylamide gel electrophoresis, mass spectrometry, and differential in gel electrophoresis, have identified numerous sperm-specific proteins. They also have provided a further understanding of protein function involved in sperm processes and for the differentiation between normal and abnormal states. In addition, studies on the sperm proteome have demonstrated the importance of post-translational modifications, and their ability to bring about physiological changes in sperm function. No longer do researchers believe that in order for them to elucidate the biochemical functions of genes, mere knowledge of the human genome sequence is sufficient. Moreover, a greater understanding of the physiological function of every protein in the tissue-specific proteome is essential in order to unravel the biological display of the human genome.

Conclusion: Recent advances in proteomic techniques have provided insight into sperm function and dysfunction. Several multidimensional separation techniques can be utilized to identify and characterize spermatozoa. Future developments in bioinformatics can further assist researchers in understanding the vast amount of data collected in proteomic studies. Moreover, such advances in proteomics may help to decipher metabolites which can act as biomarkers in the detection of sperm impairments and to potentially develop treatment for infertile couples. Further comprehensive studies on sperm-specific proteome, mechanisms of protein function and its proteolytic regulation, biomarkers and functional pathways, such as oxidative-stress induced mechanisms, will provide better insight into physiological functions of the spermatozoa. Large-scale proteomic studies using purified protein assays will eventually lead to the development of novel biomarkers that may allow for detection of disease states, genetic abnormalities, and risk factors for male infertility. Ultimately, these biomarkers will allow for a better diagnosis of sperm dysfunction and aid in drug development.

Background

As infertility affects nearly 15% of all couples of reproductive age [1,2], and family sizes continue to shrink, concerns about the reproductive potential of future generations are growing. Although infertility was believed to originate from female abnormalities, recent discoveries have revealed that as many as 50% of cases stem from male-factor defects, with no identifiable cause in 25% of infertile

men [3,4]. Male infertility presents an interesting, yet vexing, problem for men all over the world. Suboptimal sperm quality due to abnormal parameters—motility, morphology, concentration, DNA fragmentation, and genetic composition—has been linked to this issue. Nevertheless, our current understanding of the spermatozoa and its pathological and physiological effects are lacking and vaguely defined.

The study of cellular components on a molecular level offers much hope in deciphering the metabolic pathways essential for the diagnosis of male infertility. Advances

* Correspondence: agarwaa@ccf.org

²Center for Reproductive Medicine, Cleveland Clinic, Cleveland, Ohio, USA
Full list of author information is available at the end of the article

in spermatozoa research—including those in the field of proteomics—have allowed for enhanced characterization and identification of both the structural and functional proteins of spermatozoa. However, these findings remain limited. Because the spermatozoon is an extremely complex and highly accessible cell, it is remarkably suitable for proteomic analysis [5]. Moreover, because the sperm cell is responsible for transporting the paternal genome to the oocyte, examining its genetic composition may provide beneficial insight into ensuing disorders in offspring.

Recently employed proteomic techniques, such as 2D polyacrylamide gel electrophoresis (2D-PAGE), mass spectrometry (MS), and differential in gel electrophoresis (DIGE), have allowed for the identification of numerous sperm-specific proteins. These approaches have provided a greater understanding of protein function involved in sperm processes such as motility, capacitation, acrosome reaction, and fertilization. Studies of the sperm proteome have demonstrated how post-translational modifications, such as phosphorylation, glycosylation, proteolytic cleavages and mutations, bring about the physiological changes in spermatozoa function. Furthermore, proteomic analysis has allowed for the study of spermatozoa in different functional states—immature versus mature, uncapacitated versus capacitated, normal versus defective, and low sperm count versus high sperm count—all of which impact the male reproductive potential.

In this review, various proteomic techniques and their usefulness in diagnosing sperm dysfunction, as well as the possible application in the clinical setting will be examined.

Understanding the need for proteomic research

The human genome was first sequenced 10 years ago [6]. Since then, a vast amount of DNA sequences, including the diploid genome sequence of a human, have been made available in public databases [7]. Investigators are now beginning to understand that the promise of molecular medicine as a cure for genetic diseases was over emphasized, and that merely knowing the human genome sequences is not enough to elucidate the biochemical functions of genes. To unravel the biological display of the human genome, there must be greater understanding of the physiological function of every protein in the tissue-specific proteome. The genetic basis of human disease is much more complex than it was previously thought as a multitude of metabolic and regulatory pathways, post-translational modifications and complex protein-protein interactions play a major role [8]. Therefore organ- and tissue-specific proteomic studies will contribute directly to understand the fundamentals of spermatozoa and its physiological function in male fertility.

Further comprehensive studies on sperm-specific proteome, mechanisms of protein function and its proteolytic regulation, biomarkers and functional pathways, such as oxidative-stress induced mechanisms, will provide better insight into physiological functions of the spermatozoa. Large-scale proteomic studies using purified protein assays will eventually lead to the development of novel biomarkers that may allow for detection of disease states, genetic abnormalities, and risk factors for male infertility. Ultimately, these biomarkers will allow for a better diagnosis of sperm dysfunction and aid in drug development.

Proteomics—a brief overview

During the 1980s there were several new advances in molecular biology such as gene cloning, sequencing, and expression analysis. This transpired to the links between observed activity and protein function and its encoded gene to give way to protein chemistry. These novel approaches of identifying and characterizing proteins by means of both qualitative and quantitative analysis became known as proteomics. Scientists no longer only focused on the genomes of an organism, but rather turned their attention equally towards the structure of proteins and the functional interactions between the proteins.

As research soared in the mid-1990s, scientists attempted to discover a link between the genome expressing the proteins of a living cell and advances in new high-throughput techniques to aid in the process. This interest allowed for the development of successful sequencing analysis, reversible approaches that focused on phenotype instead of genotype, and gene sequencing databases. It was believed that if a link between the genotype and phenotype could be made, it would be plausible to examine all of an organism's proteins instead of concentrating on one to further understand their structure, function and their biochemical role in the body.

Proteomic approaches are generally now utilized for protein profiling, comparing the protein expression in specific organs and tissues, localizing and identifying post-translational modifications, and studying protein-protein interactions. New fractionation and labeling techniques have further enhanced protein identification of some of the least abundant proteins. It is these less prevalent proteins that have major impacts on biological systems and may be potentially valuable in clinical diagnosis and prescribing treatment.

Although the total number of human protein products has been estimated to be approximately one million, and comparison of different proteomes is possible via analysis of post-translational modifications, there still remain several limitations and hurdles [9-11]. The proteomes of mammalian cells, tissues, and bodily fluids,

and their readily active concentrations and local environment are so intricate that even new techniques are unable to make fully accurate measurements. The large quantity of data that has been obtained thus far places further challenges on data processing and analysis. Nevertheless, while researchers now face additional limitations in fully sequencing proteins and peptides due to the complex nature and physiological processes of the proteome, extensive advancements in technology, methods of analysis and ways to overcome these matters hold for a promising future.

The basic process

In order to understand how human seminal proteins can be identified and characterized, the basic process of proteomic analysis will be described in brief. After the collection and liquefaction of a neat semen sample, it can be used as is or further processed (e.g. separating into mature and immature sperm populations by density gradient centrifugation). The recovered spermatozoa can be extracted and purified, after which the protein concentration can be determined in order to allow for equal protein-loading gradient on electrophoretic gels. Because proteins differ from each other in terms of their charge and mass, they can be separated using gel electrophoresis methods, such as one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (1D SDS-PAGE), two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (2D SDS-PAGE) or DIGE.

Various stains (e.g. Coomassie Brilliant Blue and Silver) are used to identify the positions of individual proteins within the gel, which appear as spots or smudges. These in turn can be excised and analyzed by MS. Peptides generated by in-gel digestion of the proteins in the 1D/2D SDS-PAGE gels and those digested by trypsin in DIGE samples can be extracted and subsequently purified by liquid chromatography. The extracted peptides can be directly analyzed by matrix assisted laser desorption ionization-time of flight (MALDI-TOF-TOF) or liquid chromatography - mass spectrometry/mass spectrometry (LC-MS/MS). These purified peptides can then be subjected to MS, or tandem MS (MS/MS), to identify the protein by peptide mass mapping as well as by sequencing the peptides by further fragmentation characteristic mass-to-charge ratio (m/z) of ions that is resembled in a specific mass spectrum. Finally, the data obtained can be analyzed, and the amino acid sequence can be submitted into a database to search for matching peptide sequences and thereby identifying the protein(s) that are most likely present.

Techniques

The initial purpose of proteomics was simply to analyze a proteome—the protein complement expressed by a

genome—using techniques such as two-dimensional gel electrophoresis (2D-E) and protein staining followed by in-gel digestion, peptide extraction and identification of peptide protein spots using MALDI-TOF-TOF or MS/MS [12-16]. The highly dynamic proteome depends on environmental conditions as the proteins' activity and their abundance vary at different physiological states and locations of the cell or tissue. Over time, the field expanded into profiling functional and structural proteomics, specifically aiming to identify and characterize a complete set of proteins present in a cell, organ, or organism [12]. In order to accommodate this higher level of focus, a wide range of techniques has been employed.

Profiling proteomics attempts to differentiate proteins expressed between two samples (cells or organisms) while exemplifying the variations in expression levels of two different states [17]. Distinguishing alternations in the proteome during normal or stressed conditions may help elucidate the processes of signaling pathways in certain disease states. On the other hand, functional proteomics is employed when searching for protein functions on post-translationally modified proteins, which are essential in understanding the functions and role of protein in a living organism [17]. Finally, structural proteomics focuses on the tertiary structure of proteins and their complexes with other small molecules and proteins [17]. Understanding how proteins interact with their corresponding binding substrates may help uncover the mechanistic pathways these molecules undergo and subsequently further improve our knowledge of the proteome and its relation to human life adversities.

Many researchers have begun to note the abundance of proteins and their functions in humans. The human genome harbors 31000 protein-encoding genes [18]. New advances in proteomics have contributed to the most recent knowledge and understanding of how spermatozoa function and acquire the ability to fertilize [19]. Post-translational modifications (PTM) are essential for the functionality of spermatozoa, both during the maturation of the cells in the epididymis and post-ejaculatory capacitation in the female reproductive tract (See Figure 1) [19]. The modifications that are made via glycosylation, methylation, or phosphorylation are capable of altering the functional properties of spermatozoa and seminal plasma proteins that consequently result in abnormal and difficult to detect effects. Therefore, understanding the function of each protein requires information not only on quantitative levels of gene expression at both the mRNA (transcriptome) and protein (proteome) levels but also on quantitative levels of proteolytic activity and eventually information on endogenous protein substrates and products along with their physiological significance.

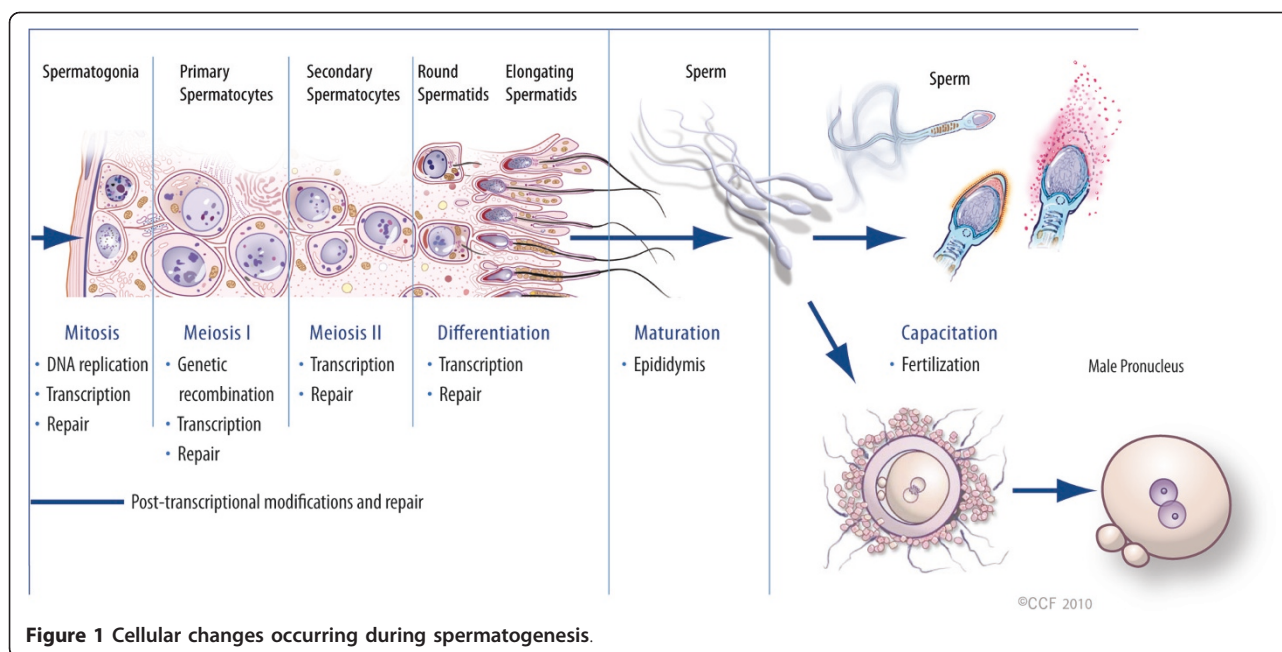


Figure 1 Cellular changes occurring during spermatogenesis.

As the field of proteomics continues to grow, advances in technology and techniques have allowed for enhanced protein separation, fractionation, purification, detection, isolation, identification and characterization. Most of the recent proteomic studies have focused on improving the resolution of separation and the identification of every protein by gel electrophoresis and MS analysis [15]. In addition, peak and multidimensional separations have given us further insight into a protein sample while also elevating the sensitivity for sample analysis [20-23]. Nevertheless, separation of a complex mixture of protein and peptides still remains one of the most difficult challenges. If it is possible to quantitatively identify uniquely expressed proteins in a disease state (as a biomarker) then a link to possible disease states can be established and potential therapeutic treatments may be discovered. The following subsections will give an overview of some of the most commonly used proteomic techniques that have allowed for the extraordinary advances in the field, particularly in sperm cell biology (See Figure 2).

Gel electrophoresis

Gel electrophoresis is a common gel-based technique that has played a revolutionary role in the field of protein separation studies. This method separates deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and proteins from complex mixtures, allowing for the detection and approximate quantification of proteins. In general, electrophoresis utilizes an applied electromotive force to a cross-linked polymer gel matrix that can limit the motion of molecules based on their mass and charge. Since 1D gel electrophoresis methods have limited resolving power to separate complex mixtures of

proteins, particularly low-abundance proteins, multidimensional approaches have been currently employed to overcome these challenges.

Two-dimensional gel electrophoresis

2D-E is a widely accepted electrophoresis method that analyzes qualitative and quantitative characterization of proteins and changes in the proteome at high resolution on a large scale. It can thus serve as an initial screening method to acquire hypotheses and establish future explorative directions. The proteins on the gel may represent a complement of active proteins in a particular tissue or cell at any given time.

In addition, the resolving power and sensitivity as well as the low equipment cost are attractive to researchers. However, heterogeneities in different gels, the electric fields, pH gradients, and thermal fluctuations are only a few factors that may influence enzymatic activity, thereby making protein verification difficult. Therefore, much focus has been given to improving methods to overcome these variations and detect low-abundance protein modifications, which are the primary proteins of interest. This would give us a better understanding of protein structure and function as well as provide a plausible diagnostic tool.

2D-PAGE examines proteins based on their isoelectric point (pI) via isoelectric focusing separation in the first dimension followed by separation in the second dimension based on molecular mass using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE specifically separates proteins according to their electrophoretic mobility—a function of polypeptide

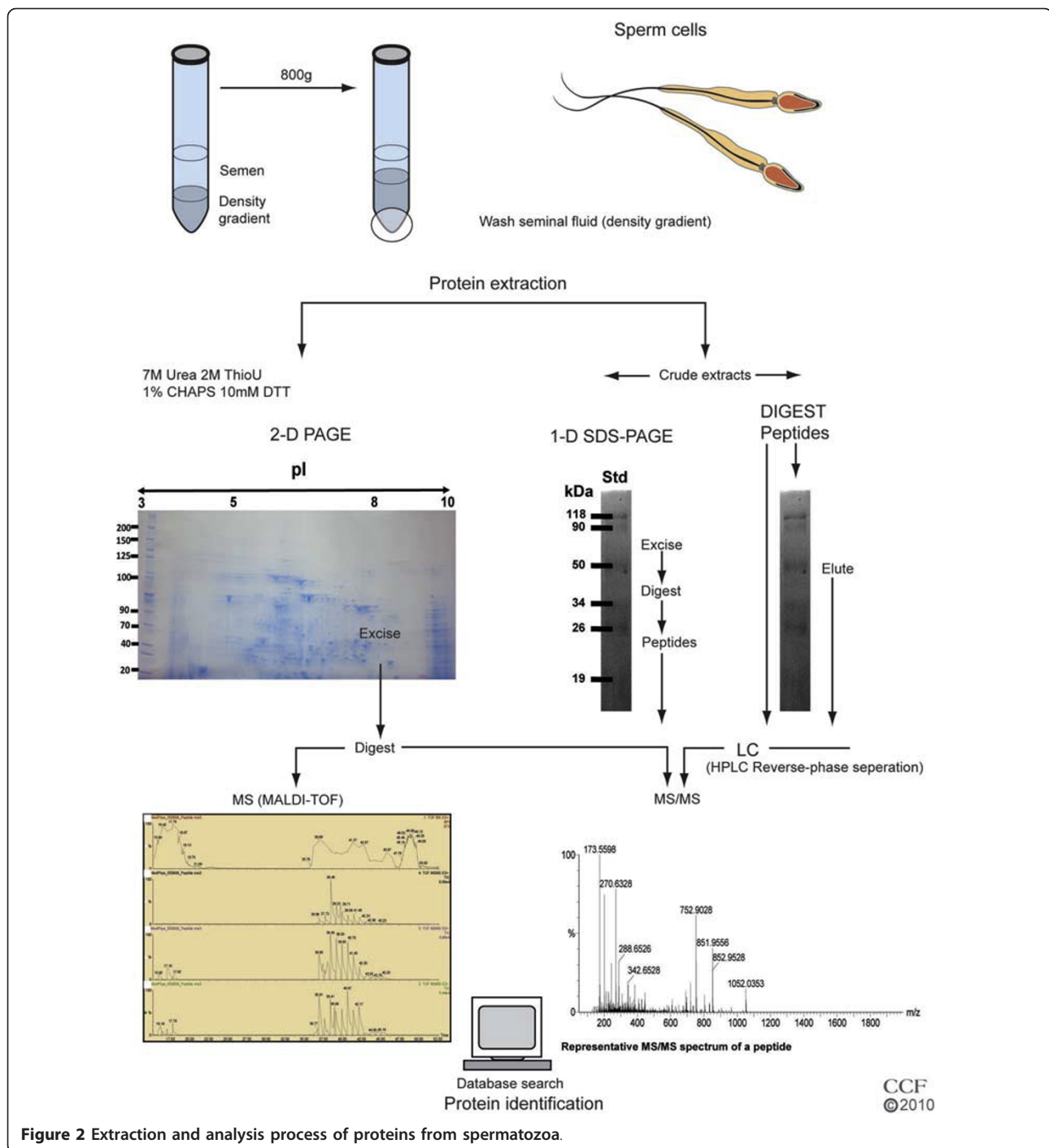


Figure 2 Extraction and analysis process of proteins from spermatozoa.

chain length or molecular mass. Polyacrylamide provides a support matrix through which proteins can migrate. Since 2D-PAGE is not sensitive enough to detect rare or low-abundance proteins, many proteins will not be resolved, and further splitting of the sample into different subcellular fractions may be necessary to reduce the complexity of protein mixtures prior to 2D-PAGE analysis.

Proteins have varying charges, shapes and sizes and thus travel at different rates through the gel. To address this problem, proteins are usually denatured in a detergent, such as SDS, which coats the protein with a negative charge. In turn, the charge of the protein no longer becomes a contributing factor to its movement on the gel, and protein size becomes the main determinant. In other words, the larger the protein, the more restricted

its movement will be towards the positively charged cathode. The protein's molecular mass is then compared to a standard ladder, which is loaded in a separate lane. Upon staining, a visible spot indicates the presence of a protein or possibly a multitude of proteins. Proteins can be then identified through qualitative analysis.

Although 2D-E-based studies have generated great insight on the overall nature and complexity of the sperm proteome in terms of iso-electric point and molecular mass, precise protein identification and characterization remains a challenge with this approach. 2D-PAGE is a labor-intensive technique that is susceptible to gel matrix irregularities due to experimental conditions such as laboratory humidity, voltage fluctuations, and preparation slip-ups. Alternative methods can be used to circumvent these challenges: MS and two-dimensional difference in gel electrophoresis (2D-DIGE).

Mass spectrometry

MS is an analytical tool that identifies proteins or peptides by measuring the masses of molecules converted into ions via their m/z ratio. It is even capable of determining the elemental composition of a molecule. In addition, MS can be clinically useful in profiling biomarkers from tissues or bodily fluids as a method to diagnose different medical conditions.

Two of the most common MS techniques that have been applied to study sperm proteins are MALDI-TOF MS and LC-MS/MS. Both can account for systematic biases when characterizing protein spots from a previously-run-electrophoresis gel. The spectral data obtained from MS gives the protein's m/z ratio and its associated intensity, which aids in determining protein differential expression and modifications.

With MALDI-TOF analysis, it is possible to determine a protein's m/z ratio by excising protein bands from the gel and subsequently digesting them with trypsin. After the peptide masses or peptide fingerprints have been recorded, they can then be matched to sets of theoretically digested protein reference masses from a database to identify the protein (peptide mass mapping).

On the other hand, LC-MS/MS is a highly sensitive and specific analytical technique that combines the physical separation of peptides generated by trypsin digestion of a spot by HPLC (high performance liquid chromatography) followed by mass determination and subsequent sequence in MS/MS set up. The use of HPLC prior to MS has fueled tremendous growth in proteomics over the past few years. Its initial application for protein or peptide separation is based on the number of unique properties such as charge, hydrophobicity and the presence of a specific tag or amino acid(s) [17]. The subsequent coupling to a tandem mass spectrometer with an interface allows for rapid separation and

identification of the complex protein mixtures and in turn provides the primary peptide sequence of the original complex protein.

Both 2D separation-based MALDI-TOF and LC-MS/MS methods can be reproduced in laboratories and provide a 2D-map of the proteins in a complex mixture based on their iso-electric point and molecular mass. Since MALDI-TOF MS can be less sensitive and time consuming, advancements in LC-MS/MS analysis provide an enhanced method for protein identification throughput and a potential future diagnostic tool.

Nevertheless, MS is not perfect. "Noisy" data from external interference may give false positive peaks, making it difficult to detect protein changes, especially in low-copy and low-abundance proteins. Alternatively, tandem mass spectrometry (MS/MS) can be applied to further fragment protein mixtures, and more importantly, to improve the detection limits of compounds and the signal-to-noise ratio relative MS. However, the total ion current of some compounds may be decreased using MS/MS while also possibly affecting the efficiency, reproducibility and detection of the mass spectra.

Difference in gel electrophoresis

DIGE is a type of gel electrophoresis used to detect quantitative changes in protein abundance. This technique has played an essential role in the advancement of proteomics in that it takes the focus off the whole protein and places it onto its constituent peptides [19]. It is at the protein level where most PTMs occur that cause the significant variations in the three-dimensional protein structure and gel electrophoretic behavior. Thus, it is the peptides that provide the best basis of proteomic comparison [19]. DIGE can overcome 2D-E limitations that are due to inter-gel variation by including an internal standard in each gel and also provide faster throughput for subsequent analysis-since the proteins from the different sample types are run on the same gel, they can be directly compared. Therefore, this technique has the potential to be used one day in a clinical setting to compare samples of healthy and diseased states to formulate a prognosis for future direction and treatment.

Bioinformatics

One of the most daunting challenges currently facing proteomics is making sense of the vast data collected thus far. Bioinformatics, which employs statistical analysis and algorithms, can assist proteomic-based studies in analyzing data and identifying proteins of interest from MS or other technical data [24]. One system used by bioinformatics is Gene Ontology (GO), which can reveal meaningful patterns found in proteins [25]. A recent study using fly models employed GO, leading to the development of new ideas on gene regulation and chromatin organization [26]. Proteomics will expand on the

current knowledge of human spermatozoa as bioinformatics applications continue to be developed in the future.

Future direction of techniques

Due to the complexity of proteomic samples, 1D separation techniques are no longer sufficient. Because PTMs occur at the peptide level and they are the main factors that distinguish similar samples, additional multidimensional separation techniques must be employed to reduce a protein's complex nature. Advances in chromatographic columns based on restricted access materials and monolithic columns have been suggested to improve selectivity. In addition, stationary phases have been introduced to improve the enrichment of low-abundance proteins and their detection. Isotope-coded affinity tag (ICAT) and isobaric tag for relative and absolute quantification (iTRAQ) are recently developed proteomic methods used to identify and quantify low-abundance and low-concentration proteins. Nevertheless, advancements in MS must keep up with the fast-developing multidimensional separation techniques to achieve optimum resolution.

Furthermore, both consistent analysis of LC-MS/MS and sample throughput have been a few of the key bottlenecks for many proteomic studies [27]. LC-MS/MS data requires multiple stages of analysis, and protein separation often takes several hours to separate a single sample. A proposed approach to improve sample throughput is to develop faster scanning mass spectrometers. In order to overcome some of the most common challenges in the field, it is necessary to improve the effectiveness of sample preparation to reduce the protein's complexity and to enrich the low abundance proteins while decreasing the more abundant, less significant ones. Miniaturization and extensive data processing and analysis are also important.

Proteomics and mature spermatozoa

Until recently, most proteomic analysis of spermatozoa was largely conducted utilizing the 2D-PAGE approach. In terms of molecular mass and iso-electric point, 2D-E-based studies have been valuable in generating data on the overall nature and complexity of the sperm proteome and the nature of some of those proteins targeted by the human immune system. An advanced 2D-PAGE resolution using narrow pH ranges for iso-electrofocusing reported 3,872 different protein spots—16 proteins of significance were identified [28].

Early sperm proteome analyses identified 1397 protein spots—at least 98 surface-related—and were catalogued by means of 2D-PAGE analysis [29]. An *in vitro* experiment conducted with prolactin, one of the tyrosine kinase surface receptors, demonstrated improved fertilizing potential of mammalian spermatozoa after the

addition of the hormone [30]. Earlier studies have noted insulin to be a hormone of interest from its displayed interaction with both the plasma membrane and sperm acrosome [31]. Moreover, these reports further reinforced the essential role of surface receptors during fertilization, and any defect or abnormality may result in suboptimal reproductive potential or even infertility. Additional proteomic analysis at the basal level of the sperm's surface can provide additional knowledge and insight into the possible biological mechanistic processes involved in acrosome reaction, capacitation, and ultimately, fertilization.

Furthermore, several proteomic studies (Refer to Table 1) have uncovered numerous housekeeping sperm proteins involved in fundamental processes such as oxidative phosphorylation and glycolysis [28,29,32-40]. Aitken *et al.* (1998) discussed the indispensable role of NADPH oxidases in sperm capacitation to help drive cAMP-induced tyrosine phosphorylation signal transduction cascades through possible suppression of tyrosine phosphatase activity [41]. Inhibiting protein kinase A (PKA) phosphorylation of A-kinase anchor proteins (AKAPs) revealed hyperactivation impairment, resulting in suboptimal reproductive function and subsequent infertility [38,41-43]. Human spermatozoa have been found to contain all the machinery necessary to sustain a cAMP-signaling pathway in the sperm tail and the components associated with the 26S proteasome [29,37,44]. The 26S proteasome is an ATP-dependent proteolytic complex essential in the sperm-egg interaction [29,37,44]. Indeed, past proteomic studies hold for a promising future of discovering new sperm proteins, which in turn will allow researchers and clinicians to better understand and diagnose sperm dysfunction (See Figure 3).

Comparative studies employing proteomic techniques have also helped to identify proteins of interest in infertile men in comparison to fertile men. Several comparative studies have recently been completed, most notably a 2007 study performed by Zhao *et al.* [45]. The study compared 8 men with asthenozoospermia with healthy controls. The authors identified 10 of 17 proteins with either increased or decreased frequency in the asthenozoospermic samples. They were able to link these differences to enzymes involved in sperm metabolism. Such comparative studies are expected to identify biomarkers, which will aid clinicians in better diagnosing male-factor infertility.

Proteomic approaches are now focused on characterizing the tyrosine phosphorylated proteins to better understand and clarify the mechanistic processes behind the capacitation-dependent alternations involved in sperm function. In the future, with improvements in proteomic methods, it will be possible to generate an

Table 1 Proteomic studies on mature human spermatozoa

Authors	Study	Proteomic Technique	Outcome
Johnston <i>et al.</i> 2005 [37]	First extensive analysis of human sperm proteome	LC-MS/MS	1760 proteins identified
Martinez-Heredia <i>et al.</i> 2006 [36] de Mateo <i>et al.</i> 2007 [40]	Goal to provide a reference 2D map of the mature human sperm proteome	MALDI-TOF	131 proteins identified
Ficarro <i>et al.</i> 2003 [38]	Phosphoproteome analysis of capacitated human sperm	LC-MS/MS	60 proteins identified
Domagala <i>et al.</i> 2007 [32]	Identification of sperm immunogenic antigens	LC-MS	35 proteins identified
Secciani <i>et al.</i> 2009 [33]	Protein profile of capacitated versus ejaculated human sperm	MS/MS	116 proteins identified
Shetty <i>et al.</i> 2001 [39]	Identification of immunocontraceptive candidates	MALDI-MS	8 proteins identified
Bohring <i>et al.</i> 2001 [34]	Isolation and identification of sperm membrane antigens	MALDI-MS	18 proteins identified
Baker <i>et al.</i> 2007 [29]	Identification of proteins in human sperm using LC-MS/MS analysis	LC-MS/MS	1056 proteins identified
Lefievre <i>et al.</i> 2007 [35]	Identify targets for S-nitrosylation in human sperm	LC-MS/MS	240 proteins identified
Li <i>et al.</i> 2007 [28]	Establish high-resolution 2-D map of human sperm proteins	MALDI-MS	16 proteins identified

accurate proteomic profile of normal human spermatozoa, which can aid in understanding and differentiating between the various functional stages of sperm cells (e.g., mature versus immature, capacitated versus uncapacitated, and normal versus defective parameters). With this information, it would be possible to verify whether impaired sperm function is related to PTMs involved in generating normal functional spermatozoa. Further understanding of various physiological and biochemical changes in post-testicular sperm maturation will unravel the molecular basis of defective sperm function characterized in subfertile men.

A study by Shetty *et al.* identified several novel sperm protein sequences from a 2D-E gel [39]. Many of those detected shared identical sequences, further stressing the importance of PTMs to sperm proteins. Another detailed MALDI-TOF analysis of the human sperm proteome consisting of over 100 mapped spots on 2D gels by Martinez-Heredia *et al.* identified 98 different proteins, and 23% of them had yet to be recorded on the human spermatozoa [36]. In addition, proteomics is a continuously evolving field and researchers are constantly exploring new avenues to further define and characterize the human proteome. The recent application of a highly sensitive nanoscale liquid chromatography technique combined with MS identification yielded 1056 gene products, 8% of which had not been previously characterized [29].

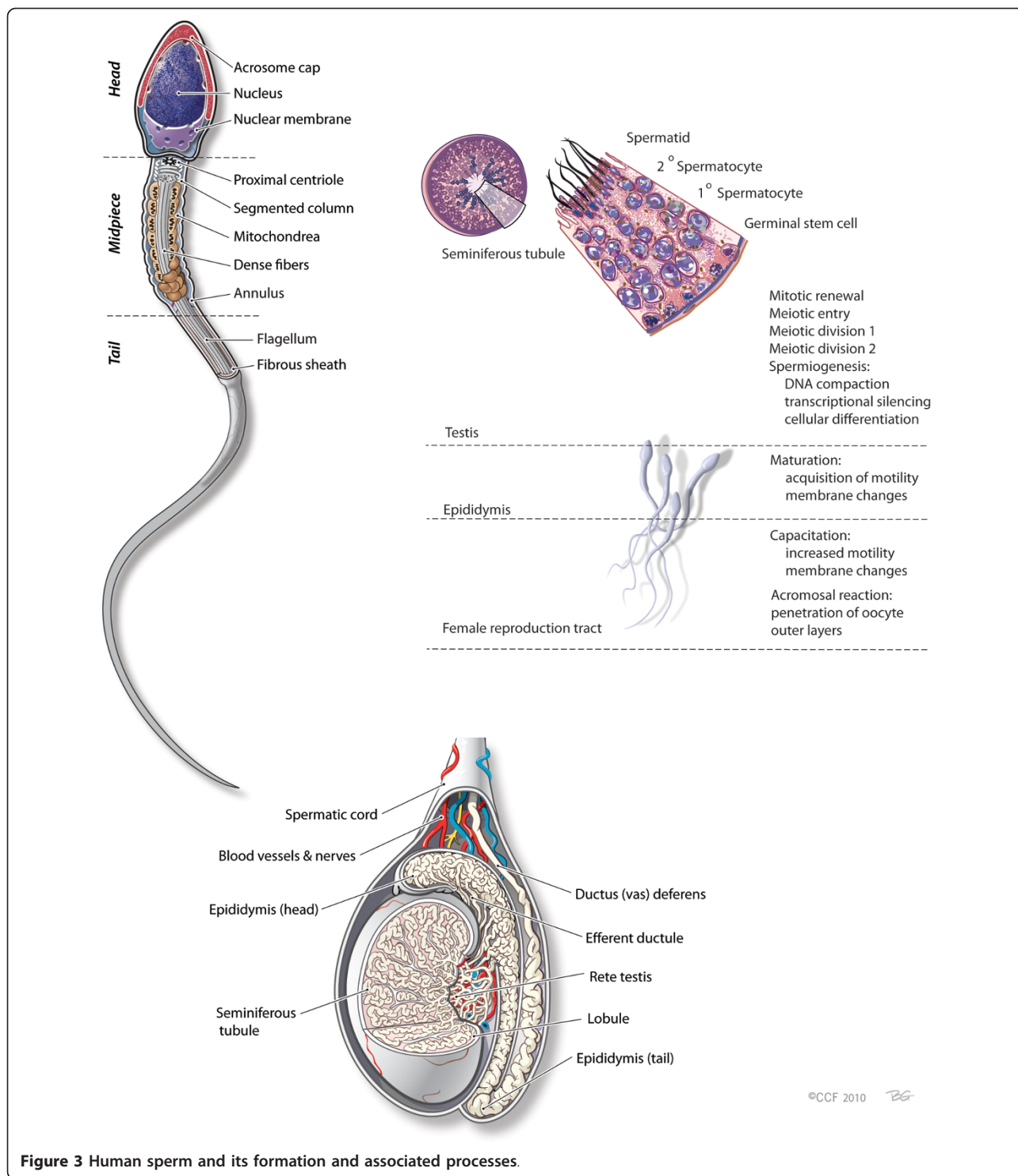
Clinical applications of proteomics

Although proteomics has only been used to study human spermatozoa in the laboratory setting, proteomic-based

studies may one day be used in the clinical setting to identify sperm dysfunction. Sperm chromatin, which consists of DNA and proteins, is essential for proper sperm function and embryonic development, as several studies have linked defects in sperm chromatin to potential reproductive errors [46-48]. Proteomics offers the potential for expanding on clinical diagnostic testing of sperm for infertility, which is currently limited to count, motility and morphology.

In order to apply proteomics to a clinical setting, several areas of study must be expanded upon. The first milestone that must be reached in order to better understand human spermatozoa is to determine the entire human sperm proteome. While a study by Baker and colleagues identified 1056 proteins and a more recent study by Li and colleagues identified 3872 spots, very little is known about the sperm proteome [28,29]. Prior studies have already been conducted on seminal plasma. Therefore, once more studies are completed, a better understanding of human ejaculate will be gained [49]. Future proteomics studies will allow for a better understanding of PTMs such as phosphorylation and sperm capacitation [38,50]. These PTMs are vital for consequent stages of sperm formation, maturation and activation [51]. Currently, there is a very limited understanding of these modifications, but this area of study is continuing to expand in scope with the pending publication of new studies.

A second milestone is the publication of comprehensive comparative studies between normal spermatozoa and abnormal spermatozoa. Such studies will highlight the potential of proteomics to identify biomarkers that can, in turn, help clinicians pinpoint certain peptides or



metabolites that may be linked to male infertility. Furthermore, proteomics can be employed in comparative studies between human sperm samples and model organisms such as mice, fish, worms, flies, and sea urchins, all of which have already contributed to greater understanding of human spermatozoa [5,52].

Protamines present a starting point for proteomic studies to identify proteins with altered expression in infertile patients. For example, one recent study by de Mateo and colleagues found an altered ratio of two protamine proteins-protamine 1 and protamine 2, also known as P1 and P2, respectively [40]. In addition, proteomic

studies using mouse models have identified the chaperone protein HSPA2 as an important assessment factor for infertility in humans [5]. One study using asthenozoospermic patients identified HSPA2 as one of 17 proteins with abnormal levels in comparison to samples from normal donors [53]. Since HSPA2 is believed to be vital to several sperm processes such as plasma membrane remodeling and cytoplasmic extrusion, HSPA2 resembles a biomarker that can aid in clinical diagnosis of male infertility (e.g., less or absent sperm-specific proteins may correspond to incompletely formed or defective sperm chromatin) [54].

Studies such as these can produce vast amounts of data, but with the help of bioinformatics applications, researchers can discover meaningful patterns that may be applied in a clinical setting [55]. Furthermore, similar proteomic studies may identify additional biomarkers that are indicative of male-factor infertility signaling pathways.

Factors affecting proteomic studies

Several factors affect proteomics, all of which must be taken into account when putting proteomics into practice in the clinical setting. The 2D-PAGE technique is laborious and indefinite in identifying and characterizing proteins [5]. In a study by Li *et al.*, only 16 of 3872 spots were identified using this technique, while in similar experiment conducted by Martinez-Heredia *et al.*, only 98 of 1000 spots were identified [28,53].

Variations in study findings can be attributed to numerous factors, many of which are uncontrollable in current proteomic advances. Pre-analytical variables may affect the testing for biomarkers during sample collection, pipetting, and dilution [55]. Human biological variation with age, race, nutrition, life style, stress, environment or global location of the sample population may also affect experimental results [56].

Nevertheless, biomarker discovery remains a very challenging task due to the complexity of the sample as well as the wide dynamic range of protein concentrations [24]. As the field of proteomics emerges to face several challenges, bioinformatics can provide statistical support in analyzing the vast amounts of data resulting from proteomics testing. Although proteomics is an emerging field facing many challenges, bioinformatics can provide statistical support in analyzing the vast amounts of data resulting from proteomics testing. However, it is expected that the future proteomics studies may lead to the development of novel diagnostic tools and therapeutic drugs to improve sperm dysfunction by elucidating potential causes of sperm impairment and providing insights to its underlying mechanistic pathway [50].

Conclusion

As researchers and scientists note the sudden decline in male reproductive potential worldwide, many have focused their attention on male-factor defects. Several reports have linked suboptimal sperm quality to this growing concern [57,58]. Although much remains unknown about human spermatozoa and its pathological and physiological effects on reproductive function, recent advances in proteomic techniques have provided insight into sperm function and dysfunction. Several multidimensional separation methods have been utilized to identify and characterize spermatozoa [28,29,37]. Future developments in bioinformatics can further assist researchers in understanding the vast amount of data collected in proteomic studies. In addition, such advances in proteomics may help to decipher metabolites directly involved in normal growth, development, and reproduction, which can stand as biomarkers in the detection of disease or sperm impairments. Multinuclear nuclear magnetic resonance-based metabolomic approaches along with advanced proteomic techniques may provide essential information of the cellular metabolome in order to further understand cellular function. With the identification of novel biomarkers through proteomic studies, clinical tests and treatments for sperm dysfunction might be developed to potentially help infertile couples.

Acknowledgements

The research for this article was supported by the Center for Reproductive Medicine, Cleveland Clinic.

Author details

¹Division of Medical Physiology, Stellenbosch University, Tygerberg, South Africa. ²Center for Reproductive Medicine, Cleveland Clinic, Cleveland, Ohio, USA. ³Molecular Biotechnology Core Laboratory, The Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, Ohio, USA.

Authors' contributions

SSDP conceived the study, contributed to the design of the study, information collection, drafting of the manuscript and critical review of the final paper. AK helped with the design of the study, information collection and drafting of the manuscript. DB helped with information collection and drafting of the manuscript. AA and SY have been involved in the critical revision of the manuscript. All authors read and approved the final manuscript.

Authors' information

Stefan S du Plessis, PhD, is the head of the Division of Medical Physiology and associate-professor in the Faculty of health Sciences at Stellenbosch University, Tygerberg, South Africa. Anthony H Kashou and David J Benjamin are summer interns at the Center for Reproductive Medicine, Cleveland Clinic, Cleveland, Ohio, USA. Satya Yadav is the director of the Molecular Biotechnology Core Laboratory of The Lerner Research Institute at the Cleveland Clinic Foundation, Cleveland, Ohio, USA. Ashok Agarwal is the Director of the Center for Reproductive Medicine, Glickman Urological & Kidney Institute and Ob/Gyn & Women's Health Institute, Cleveland Clinic, Cleveland, Ohio, USA

Competing interests

The authors declare that they have no competing interests.

Received: 5 November 2010 Accepted: 22 March 2011
Published: 22 March 2011

References

1. Stephen EH, Chandra A: **Updated projections of infertility in the United States: 1995-2025.** *Fertil Steril* 1998, **70**:30-34.
2. Thonneau P, Marchand S, Tallec A, Ferial ML, Ducot B, Lansac J, Lopes P, Tabaste JM, Spira A: **Incidence and main causes of infertility in a resident population (1 850 000) of three French regions (1988-1989).** *Hum Reprod* 1991, **6**:811-816.
3. Sharlip ID, Jarow JP, Belker AM, Lipshultz LI, Sigman M, Thomas AJ, Schlegel PN, Howards SS, Nehra A, Damewood MD, Overstreet JW, Sadovsky R: **Best practice policies for male infertility.** *Fertil Steril* 2002, **77**:873-882.
4. Turek PJ: **Practical approaches to the diagnosis and management of male infertility.** *Nat Clin Pract Urol* 2005, **2**:226-238.
5. Wu TF, Chu DS: **Sperm Chromatin.** *Mol Cell Proteomics* 2008, **7**:1876-1886.
6. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatide P, Ballew RM, Hussen DH, Russo Wortman J, Zhang Q, Kodira CD, Zheng XH, Chen L, Marian S, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, et al: **The sequence of the human genome.** *Science (Washington, DC)* 2001, **291**:1304-1351.
7. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Wing Chun Pang A, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers Y, Frazier ME, Scherer SW, Strausberg RL, et al: **The diploid genome sequence of an individual human.** *PLoS Biol* 2007, **5**: e254.
8. Mann M, Jensen O: **Proteomic analysis of post-translational modifications.** *Nat Biotechnol* 2003, **21**:255-261.
9. Wilkins MR, Gasteiger E, Gooley AA, Herbert BR, Molloy MP, Binz PA, Ou K, Sanchez JC, Bairoch A, Williams KL, Hochstrasser DF: **High-throughput mass spectrometric discovery of protein post-translational modifications.** *J Mol Biol* 1999, **289**:645-657.
10. Godovac-Zimmermann J, Brown LR: **Perspectives for mass spectrometry and functional proteomics.** *Mass Spectrom Rev* 2001, **20**:1-57.
11. Rabilloud T: **Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains.** *Proteomics* 2002, **2**:3-10.
12. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphrey-Smith I, Hochstrasser DF, Williams KL: **Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it.** *Biotechnol Genet Eng Rev* 1996, **13**:19-50.
13. O'Farrell PH: **High resolution two-dimensional electrophoresis of proteins.** *J Biol Chem* 1975, **250**:4007-4021.
14. Wassinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphrey-Smith I: **Progress with gene-product mapping of the mollicutes: Mycoplasma genitalium.** *Electrophoresis* 1995, **16**:1090-1094.
15. Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphrey-Smith I, Williams KL, Hochstrasser DF: **From proteins to proteomics: large scale protein identification by two-dimensional electrophoresis and amino acid analysis.** *Biotechnology (N Y)* 1996, **14**:61-65.
16. Bradshaw RA, Burlingame AL: **From proteins to proteomics.** *IUBMB Life* 2005, **57**:267-272.
17. Mitulovic G, Mechtler K: **HPLC techniques for proteomics analysis—a short overview of latest developments.** *Brief Funct Genomic Proteomic* 2006, **5**:249-260.
18. Baltimore D: **Our genome unveiled.** *Nature* 2001, **409**:814-816.
19. Aitken RJ, Baker MA: **The role of proteomics in understanding sperm cell biology.** *Int J Androl* 2008, **31**:295-302.
20. Moore RE, Klickler L, Schumann D, Lee TD: **A microscale electrospray interface incorporating a monolithic, poly (styrene-divinylbenzene) support for on-line liquid chromatography/tandem mass spectrometry analysis of peptides and proteins.** *Anal Chem* 1998, **70**:4879-4884.
21. Davis MT, Lee TD: **Rapid protein identification using a microscale electrospray LC/MS system on an ion trap mass spectrometer.** *J Am Soc Mass Spectrom* 1998, **9**:194-201.
22. Washburn MP, Wolters D, Yates JR: **Large-scale analysis of the yeast proteome by multidimensional protein identification technology.** *Nat Biotechnol* 2001, **19**:242-247.
23. Wolters DA, Washburn MP, Yates JR: **An automated multidimensional protein identification technology for shotgun proteomics.** *Anal Chem* 2001, **73**:5683-5690.
24. Baldwin MA: **Protein identification by mass spectrometry: issues to be considered.** *Mol Cell Proteomics* 2004, **3**:1-9.
25. Lan N, Montelione GT, Gerstein M: **Ontologies for proteomics; towards a systematic definition of structure and function that scales to the genome level.** *Curr Opin Chem Biol* 2003, **7**:44-54.
26. Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL: **Genomic and functional evolution of the Drosophila melanogaster sperm proteome.** *Nat Genet* 2006, **38**:1440-1445.
27. Baldwin MA: **Protein identification by mass spectrometry.** *Mol Cell Proteomics* 2004, **3**:1-9.
28. Li LW, Fan LQ, Zhu WB, Nie HC, Sun BL, Luo KL, Liao TT, Tang L, Lu GX: **Establishment of a high resolution 2-D reference map of human spermatozoal proteins from 12 fertile sperm-bank donors.** *Asian J Androl* 2007, **9**:321-329.
29. Baker MA, Reeves G, Hetherington L, Muller J, Baur I, Aitken RJ: **Identification of gene products present in Triton X-100 soluble and insoluble fractions of human spermatozoa lysates using LC-MS/MS analysis.** *Proteomics Clin Appl* 2007, **1**:524-532.
30. Fukuda A, Mori C, Hashimoto H, Noda Y, Mori T, Hoshino K: **Effects of prolactin during preincubation of mouse spermatozoa on fertilizing capacity in vitro.** *J In Vitro Fert Embryo Transf* 1989, **6**:92-97.
31. Silverstroni L, Modesti A, Sartori C: **Insulin-sperm interaction: effects on plasma membrane and binding to acrosome.** *Arch Androl* 1992, **28**:201-211.
32. Domagala A, Pulido S, Kurpisz M, Herr JC: **Application of proteomic methods for identification of sperm immunogenic antigens.** *Mol Hum Reprod* 2007, **13**:437-444.
33. Secciani F, Bianchi L, Ermini L, Cianti R, Armini A, La Sala GB, Focarelli R, Bini L, Rosati F: **Protein profile of capacitated versus ejaculated human sperm.** *J Proteome Res* 2009, **8**:3377-3389.
34. Bohring C, Krause E, Habermann B, Krause W: **Isolation and identification of sperm membrane antigens recognized by antisperm antibodies, and their possible role in immunological infertility disease.** *Mol Hum Reprod* 2001, **7**:113-118.
35. Lefevre L, Chen Y, Connor SJ, Scott JL, Publicover SJ, Ford WC, Barratt CL: **Human spermatozoa contain multiple targets for protein S-nitrosylation: an alternative mechanism of the modulation of sperm function by nitric oxide?** *Proteomics* 2007, **7**:3066-3084.
36. Martinez-Heredia J, Estanyol JM, Balleca JL, Oliva R: **Proteomic identification of human sperm proteins.** *Proteomics* 2006, **6**:4356-4369.
37. Johnston DS, Wooters J, Kopf GS, Qiu Y, Roberts KP: **Analysis of the human sperm proteome.** *Ann N Y Acad Sci* 2005, **1061**:190-202.
38. Ficarro S, Chertihin O, Westbrook VA, White F, Jayes F, Kalab P, Marto JA, Shabanowitz J, Herr JC, Hunt DF, Visconti PE: **Phosphoproteome analysis of capacitated human sperm. Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing protein/p97 during capacitation.** *J Biol Chem* 2003, **278**:11579-11589.
39. Shetty J, Diekmann AB, Jayes FCL, Sherman NE, Naaby-Hansen S, Flickinger CJ, Herr JC: **Differential extraction and enrichment of human sperm surface proteins in a proteome: identification of immunocontraceptive candidates.** *Electrophoresis* 2001, **22**:3053-3066.
40. de Mateo S, Martinez-Heredia J, Estanyol JM, Dominguez-Fandos D, Vidal-Taboada JM, Balleca JL, Oliva R: **Marked correlations in protein expression identified by proteomic analysis of human spermatozoa.** *Proteomics* 2007, **7**:4264-4277.
41. Aitken RJ, Harkiss D, Knox W, Paterson M, Irvine DS: **A novel signal transduction cascade in capacitating human spermatozoa characterised by a redox-regulated, cAMP-mediated induction of tyrosine phosphorylation.** *J Cell Sci* 1998, **111**:645-656.
42. Baker MA, Hetherington L, Aitken RJ: **Identification of SRC as a key PKA-stimulated tyrosine kinase involved in the capacitation-associated hyperactivation of murine spermatozoa.** *J Cell Sci* 2006, **119**:3182-3192.
43. Miki K, Willis WD, Brown PR, Goulding EH, Fulcher KD, Eddy EM: **Targeted disruption of the Akap4 gene causes defects in sperm flagellum and motility.** *Dev Biol* 2002, **248**:331-342.

44. Yi YJ, Manandhar G, Oko RJ, Breed WG, Sutovsky P: **Mechanism of sperm-zona pellucida penetration during mammalian fertilization: 26S proteasome as a candidate egg coat lysin.** *Soc Reprod Fertil Suppl* 2007, **63**:385-408.
45. Zhao C, Huo R, Wang FQ, Lin M, Zhou ZM, Sha JH: **Identification of several proteins involved in regulation of sperm motility by proteomic analysis.** *Fertil Steril* 2007, **87**:436-438.
46. Boe-Hansen GB, Fedder J, Ersbøll AK, Christensen P: **The sperm chromatin structure assay as a diagnostic tool in the human fertility clinic.** *Hum Reprod* 2006, **21**:1576-1582.
47. Bungum M, Humaidan P, Axmon A, Spano M, Bungum L, Erenpreis J, Giwercman A: **Sperm DNA integrity assessment in prediction of assisted reproduction technology outcome.** *Hum Reprod* 2007, **22**:174-179.
48. Cebesoy FB, Aydos K, Unlu C: **Effect of sperm chromatin damage on fertilization ratio and embryo quality post-ICSI.** *Arch Androl* 2006, **52**:397-402.
49. Pilch B, Mann M: **Large-scale and high-confidence proteomic analysis of human seminal plasma.** *Genome Biol* 2006, **7**:R40.
50. Barratt CLR: **The human sperm proteome: the potential for new biomarkers of male fertility and a transformation in our understanding of the spermatozoon as a machine: commentary on the article 'Identification of proteomic differences in asthenozoospermic sperm samples' by Martinez et al.** *Hum Reprod* 2008, **23**:1240-1241.
51. Dadoune JP, Siffroi JP, Alfonsi MF: **Transcription in haploid male germ cells.** *Int Rev Cytol* 2004, **237**:1-56.
52. Guo X, Shen J, Xia Z, Zhang R, Zhang P, Zhao C, Xing J, Chen L, Chen W, Lin M, Huo R, Su B, Zhou Z, Sha J: **Proteomic analysis of proteins involved in spermiogenesis in mouse.** *J Proteome Res* 2010, **9**:1246-1256.
53. Martínez-Heredia J, de Mateo S, Vidal-Taboada JM, Ballesca JL, Oliva R: **Identification of proteomic differences in asthenozoospermic sperm samples.** *Hum Reprod* 2008, **23**:783-791.
54. Huszar G, Ozkavukcu S, Jakab A, Celik-Ozenci C, Sati GL, Cayli S: **Hyaluronic acid binding ability of human sperm reflects cellular maturity and fertilizing potential: selection of sperm for intracytoplasmic sperm injection.** *Curr Opin Obstet Gynecol* 2006, **18**:260-267.
55. Colantonio DA, Chan DW: **The clinical application of proteomics.** *Clin Chim Acta* 2005, **357**:151-158.
56. Fraser CG: **Biological variation: from principles to practice.** *Biological Variation: From Principles to Practice*. 1 edition. Washington, DC: AACCC Press; 2001, 151.
57. Povey AC, Stocks SJ: **Epidemiology and trends in male subfertility.** *Hum Fertil (Camb)* 2010, **13**:182-188.
58. Merzenich H, Zeeb H, Blettner M: **Decreasing sperm quality: a global problem?** *BMC Public Health* 2010, **10**:24.

doi:10.1186/1477-7827-9-36

Cite this article as: du Plessis et al.: Proteomics: a subcellular look at spermatozoa. *Reproductive Biology and Endocrinology* 2011 **9**:36.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

