Sperm proteomics: potential impact on male infertility treatment

Ashok Agarwal\textsuperscript{a,} b, Ricardo Pimenta Bertolla\textsuperscript{a} and Luna Samanta\textsuperscript{c}

\textsuperscript{a}American Center for Reproductive Medicine, Department of Urology, Cleveland Clinic, Cleveland, OH, USA; \textsuperscript{b}Department of Surgery, Division of Urology, Human Reproduction Section, Federal University of São Paulo, São Paulo, Brazil; \textsuperscript{c}Department of Zoology, School of Life Sciences, Ravenshaw University, Cuttack, India

\section*{ABSTRACT}
Spermatozoa are unique cells that have highly compact DNA, motility (and hypermotility) patterns, a specific morphology, localized mitochondria and an apical acrosome. They are the end product of a dynamic process termed spermatogenesis. Sperm are therefore produced with specific proteins in order to effect different traits, such as the presence of cysteine-rich protamines in DNA, which effectively compacts DNA. Moreover, specific proteins are transferred during epididymal maturation and after ejaculation in order to render sperm capable of undergoing post-ejaculatory alterations, generally termed capacitation, which confers capacity to fertilize a mature oocyte. In addition, sperm exhibit several post-translational modifications, which are fundamental to their function, such as SUMOylation and ubiquitination. Discussed in this review is the current knowledge of the sperm proteome in terms of its composition and the function that these proteins determine, as well as their post-translational modifications and how these alter sperm functional integrity. Studies are emphasized that focus on shotgun proteomics – untargeted determination of the protein constituent of a cell in a given biological condition – and technologies currently applied toward that end are reviewed.

\section*{Introduction}
Infertility affects 15\% of couples of reproductive age, and in 50\% of these cases, there is a male component [1]. Various cellular mechanisms are related to? Cause? Male infertility, such as sperm functional alterations (i.e. DNA fragmentation, acrosome integrity, among others) related to failed fertilization [2], poor embryo development [3,4], and assisted reproduction outcome [5,6]. Recently, focus has been placed on investigating the molecular machineries that affect these mechanisms, with a special focus on the protein composition of ejaculated sperm [7–9].

Spermatozoa are the product of spermatogenesis – a complex process in which diploid spermatogonia undergo meiosis and extensive morphological modeling within the seminiferous tubules in the testes [10]. In humans, the process of spermatogenesis is responsible for generating between 20 and 240 million sperm per day and depends on tight regulation of cellular metabolism [10]. Moreover, a number of selective mechanisms within the testes and epididymides help ensure that high quality sperm are available for ejaculation [10].

Many studies have hinted at a specific proteome signature that allow for the differentiation of high quality from low quality sperm [9], as well as specific proteins that activate following ejaculation in order to orchestrate acquisition of specific motility patterns [11], exocytosis of the acrosomal content, [11,12] and fusion of the sperm membrane to the oolemma [13,14]. However, many proteins in sperm may be added after spermatogenesis via microvesicles within the epididymal lumen (epididymosomes) [15–18] or even after ejaculation through prostatic microvesicles (prostasomes) [19–21].

Moreover, different subsets of sperm seem to coexist within an ejaculate. D’amours et al., for example, demonstrated that dead bovine sperm (but not live sperm) are tagged, via transfer by epididymosomes, with a protein containing multiple fibronectin II domains, which are known to bind to collagen – Epididymal Sperm Binding Protein 1 (ELSPBP1) [15]. This protein has also been associated with increased sperm DNA fragmentation in humans [9]. Other bovine studies have localized Fas – a membrane receptor of the tumor necrosis factor family that initiates apoptosis – to at least three different sperm regions, indicating that these are indeed subsets with different activities [22].

This division of sperm into subsets poses important considerations in regards to study design because the whole proteome of an ejaculate may not reflect its true fertility potential. It has been demonstrated, for example, that sperm marked for specific processes, such as apoptosis, are indeed of lower reproductive potential than those unmarked [23]. Conversely, specific proteins, which may offer sperm resistance within the female reproductive tract may also mark a population of viable sperm [24].

The field of proteomics has greatly benefited from the development of mass spectrometry (MS)-based systems [25,26]. While initial studies relied on gel separation of proteins by their isoelectric point and their mass (two-dimensional (2D) gel electrophoresis) [27,28], the exponential rate at which mass spectrometric resolution and sensitivity, allied to the development of protein- and peptide-specific chromatographic separation techniques, have led to systems that can detect and quantify a greater number of proteins at a higher dynamic range [26,29–31]. The latter is usually achieved via digestion of proteins into their constitutive peptides (this may...
be preceded by protein fractionation) and identification and quantification of these peptides by liquid chromatography–tandem mass spectrometry (LC–MS/MS) in an approach termed shotgun proteomics [25,26,29].

In the first part of this review, we will discuss the technologies that are currently used for sperm proteomic studies as well as the challenges in data analysis. In the second part, we will extensively review the data that have been generated regarding the human sperm proteome.

Review criteria
A literature search was performed using PubMed and Google Scholar electronic databases, using the following keywords: ‘human’, ‘spermatozoa’, ‘proteome’, ‘protein–protein interaction network’, ‘shotgun proteomics’, ‘proteomics’, ‘spermogenesis’, ‘epididymal maturation’, ‘post-translational modifications’, ‘infertility’, and ‘sperm function’. Only articles published in peer-reviewed journals were selected. The search targeted studies that were published in English-based journals within the past 10 years.

Methods for determining the sperm proteome
Proteomics is the identification and quantification of all the proteins in a given fluid or cell under a specific biological condition [32]. Initial untargeted studies used 2D-gel electrophoresis (2D-GE) in which proteins are separated initially by the isoelectric points – the pH at which their net charge is zero – followed by separation by mass in conventional laemmli gels and staining. Each protein thus forms a spot on a gel, and spot intensities are compared across gels in order to observe differential expression levels. Spots of interest may then be identified by MS (in the same principle that will be discussed further), and the relative quantification information is extracted from the gel (spot intensities) (for a review, see [27]).

Due to the variable nature of gel polymerization, these studies usually have high intra- and inter-assay variation [33,34]. To decrease this variability, difference gel electrophoresis (DIGE) can be used. With DIGE, proteins of one sample are tagged with a fluorescent stain whereas proteins of another sample are tagged with a different fluorescent stain. These samples are then pooled and run through all the steps together. Fluorescent gel analysis then quantifies the fluorescent signals from each sample to provide a more accurate differential analysis [27,35].

More recently, MS-based proteomics pipelines have been developed, increasing dynamic range, quantification, and sensitivity with direct integration to chromatographic separation [26,29,31,36]. In these studies, proteins are digested into their constituent peptides, which are extensively analyzed and quantified. Proteins are therefore identified and quantified based on in silico construction of their identified peptides (i.e. a bottom-up approach). This is achieved using MS equipment in an approach termed ‘shotgun proteomics’ [26,29,37].

Mass spectrometers analyze the masses of ionized molecules in a gas phase. The spectrometers contain an ionization source and one or more mass analyzers [31]. Ionization is achieved through the use of ‘soft’ ionization – usually matrix-assisted laser desorption/ionization [38,39] and electrospray ionization (ESI) [40]. Because ESI ionizes molecules that are in a liquid phase, they are utilized for direct integration with liquid chromatography equipment in a setup (LC–MS) that allows quantitative information to be derived from the constructed chromatograms [36].

Following ionization, the mass analyzers select or analyze the mass of the ionized molecules, utilizing different principles. In general, quadrupole (Q) and ion trap mass analyzers filter and accumulate the ions of interest, whereas time-of-flight and orbitrap mass analyzers detect mass values with high resolution [31]. Current shotgun proteomic platforms usually couple an ultraperformance liquid chromatographer (UPLC) to a hybrid mass spectrometer (MS/MS) in a setup termed LC–MS/MS [7,9,41–43].

In most MS-based shotgun proteomics studies, an initial scan is generated in the high-resolution mass analyzer, and the most important parent scan ions are determined. In subsequent scans, each parent ion selected in the first mass analyzer is fragmented in a collision chamber, and the fragment ions are identified in the high-resolution mass analyzer [31]. This approach is considered a data-dependent approach because an initial scan is necessary.

More recently, data-independent acquisition techniques have been generated, such as MS² [42,44] and sequential window acquisition of all theoretical fragment ion spectra-MS [45–47]. In data-independent acquisition, there is no precursor ion selection; all ions are analyzed in their unfragmented and fragmented forms. These techniques improve quantitative analysis of the identified proteins as well as reproducibility of the data analyses [48].

Another important point is that coverage of the proteome depends on the resolution of the chromatographic separation, because coelution of peptides decreases the odds that LC–MS/MS will be able to identify them, and this effect is more important than low peptide quantity [49]. This ultimately means that decreasing the sample complexity will allow for a greater number of peptides (and thus of proteins) to be determined. Fractionation of proteins before the LC–MS/MS experiment has increasingly been used toward that end. Fractionation may be performed using a number of methods: running 1D gels to preseparate proteins based on their mass and then slicing the generated protein lanes into mass fractions; in-solution isoelectric focusing, and prefractionation utilizing affinity columns [50–52].

Finally, in order to decrease inter-assay variability, many studies have utilized differential mass tags. These mass tags, such as iTRAQ, iCAT, tandem mass tags [53], and dimethyl labels [54] allow for the differential marking of samples. These samples are then pooled prior to the LC–MS/MS experiment. Software processing allows for relative quantification between different samples in a same run, thus increasing reproducibility.

In studies focusing on sperm proteins, some other important points should be considered. If sperm are collected by ejaculation, it is important to remove any contaminating round cells [9]. Moreover, some authors have suggested that different subpopulations of sperm coexist in the human ejaculate [22,55,56]. For example, if this is the case, researchers may want to select only viable sperm using density gradient or
swim-up techniques. Sperm selection based on functional traits could also generate important information on sperm physiology (such as selecting sperm with fragmented DNA or sperm bound to the zona pellucida of mature oocytes). This information is probably diluted in whole ejaculates, which enhances the importance of planning sperm prefractionation even before protein extraction.

Because sperm interaction with the surrounding media is fundamental in sperm physiology, approaches targeting specific sperm regions are also important. Isolating membrane proteins, for example, may be useful in determining sperm–egg interacting proteins [57] or proteins transferred during epididymal maturation [15,58]. Similarly, isolating mitochondrial proteins could generate information regarding oxidative phosphorylation energy-producing pathways in mature sperm under diverse conditions [59].

There are many methods for determining the constituent sperm proteome as well as differential proteomes under different biological conditions. It is important, however, to consider isolating specific subsets of sperm, or specific regions within these subsets, in order to focus on the intended organisms. Furthermore, technological considerations should be made such as whether a data-independent approach is feasible and if a label-free approach will suffice or if pooling samples with different labels for comparison within a same run will generate better information. Finally, studies may focus on differential proteome networks, as these will overexpress specific functions when acting conjointly. Or, a functional component may be added to these studies by verifying regulatory post-translational modifications in specific proteins or in a similar shotgun approach.

**Differences in form and function: differential proteome signatures and post-translational modifications**

Technological advances in MS-based proteomics and in peptide separation technologies have greatly increased coverage of the proteome [29,31]. As a result, studies have identified and quantified thousands of peptides and proteins within biological samples [60,61]. This has led to the understanding that proteins interact with one another to enrich specific functions and thus to produce a given phenotype [62].

Proteins thus form protein–protein interaction (PPI) networks and participate in specific pathways to produce specific functions, such as fertilization (Figure 1). These PPI networks are scale-free in that few proteins may lead to multiple interactions (hub proteins) while many proteins may lead to few interactions (Figure 2) [63]. This is important to understand because proteins of more centrality (or with more interactions), when altered or absent, will affect the function of many different proteins and pathways. These are usually lethal gene products. More often than not, however, mild phenotypic differences will be observed due to altered expression of proteins with few interactions [64].

Another important point is that proteins, even when present, may have altered function due to post-translational modifications (PTMs) [65,66]. There are many different types of PTMs, such as SUMOylation, phosphorylation, glycosylation, and ubiquitination [67]. In many cases, the proteome map itself may not be altered but differences in PTMs may cause a deviation in function. Sperm capacitation, for example, depends on an orchestrated cascade of protein phosphorylation which, when finely regulated, leads to alteration in motility patterns and eventually in acrosome exocytosis [68]. In addition, ubiquitination of sperm mitochondria prevents the paternal mitochondrial DNA from being transmitted to the offspring [69]. Finally, SUMOylation of specific proteins in sperm such as topoisomerase-IIα has been associated with DNA damage, thus demonstrating an effect of PTM on sperm function [55].

In order to convincingly establish protein function, as a function of their interaction networks, there is a necessity to design studies that provides as complete a picture of the proteome as possible. Moreover, enriching PTMs for their characterization in shotgun studies may lend a functional component to the understanding of the sperm proteome. Downstream data analysis therefore must be able to deal
with a large number of inter-correlated variables to filter out unimportant data.

### Data analysis in proteomics

Shotgun proteomic studies can identify and quantify hundreds to thousands of proteins at a time [7,9,41–43]. Results matrices are therefore quite complex, containing thousands of variables – many of which are intercorrelated. In addition, the variables are scaled differently from those used in conventional sperm analysis studies such as sperm concentration, motility, morphology, DNA integrity, and acrosome integrity as other information (age, hormonal profile, etc.). Therefore, the statistical analysis must consider the intercorrelation observed between the quantified proteins as well as filter unimportant variables to remove them from constructed models [41]. Furthermore, because proteins interact with one another, the data analysis must also include functional enrichment analysis [70,71].

In regards to statistical analysis, a univariate approach is often used. However, in these cases, the analysis must include adjustments for multiple comparisons, such as permutation-based false discovery rate determination for p-value adjustment [72]. A multivariate statistical design may deal with these data more appropriately. Principal component analysis, an unsupervised data multivariate analysis, extracts components (highly correlated variables that explain patterns in the data) of independent variables. This is useful if general patterns are observed and if eventual sample outliers are present as confounding individuals in the data [72]. Partial least squares discriminant analysis, a supervised multivariate analysis test, extracts components from both the independent and the dependent variables in order to generate a model that best explains why two groups of patients or samples differ from one another [41]. Logistic regression models may also be constructed in order to predict group participation based on protein quantification. These models will produce odds ratios for individual protein quantities – variations of these quantities will increase the odds of that individual being of one group (e.g. infertile) or another [41].

Proteins can be annotated in different ways using online databases. Gene ontology terms, for example, are annotated into three general categories: biological process, molecular function, and cellular component [73]. However, proteins may also be annotated with pathway terms (Kyoto encyclopedia of genes and genomes – KEGG [74], Reactome [75], PantherDB [70]), domains (InterPro [76]), interactions (IntAct [77], BioGRID [78], HPRD [79], STRING [80], among others. Because a single protein is usually annotated with many terms from each database, statistical tests have been designed to observe if the frequency of these terms in the PPI network or in one of its clusters is higher than the frequency of these terms in the whole database or in reference samples [81]. This approach, termed functional enrichment analysis, will offer a functional inference as to what role that protein (and its counterparts in the PPI network) plays in determining a phenotype [71,81–83].

In this workflow, a PPI network is generated by searching for interaction information from the databases cited above. Thenceforth, functional enrichment analysis of the differentially expressed proteins generates information regarding the functions enriched by these clusters and possible mechanisms that explain the biological state [9,41–43].

The proteomic study of spermatozoa has taken a big leap ever since the historical discovery of protamines in salmon sperm by Miescher [84]. Interest in using proteomic studies to identify potential diagnostic markers has grown in recent years. In one study, 2045 proteins were identified from a group of patients with idiopathic infertility – 21 proteins were differentially expressed (>1.2-fold) in men whose sperm resulted in a clinical pregnancy via ART versus those who did not [85]. With the absence of physiologically active transcription and translation, spermatozoa could be regarded as outstanding candidates for proteomic study. This is further supported by the fact that they can be purified at high concentrations and can be pharmacologically manipulated into various functional states such as capacitated, non-capacitated, and acrosome reacted [86,87]. Moreover, understanding the differences in the proteomic profile of developing sperm populations – caput versus cauda epididymis and ejaculated verses capacitated, for example–will help unravel the post-

---

Figure 2. Scale-free protein-protein interaction sub-network of reviewed human sperm proteins (Uniprot [83]) highlighting hub proteins (proteins with many interactions) in red.
translational modifications relevant to sperm maturation [88,89]. In addition, comparative proteomic analysis will not only pave the way for understanding any ‘hidden’ sperm factors responsible for infertility but also provide a molecular biomarker for diagnosis and prognosis of male factor infertility – thus helping to develop an appropriate therapeutic approach to treat male infertility [90].

**Origin and characterization of sperm proteins – spermatogenesis, epididymal maturation, and post-ejaculation**

Spermatogenesis is a highly coordinated process that marks the differentiation of diploid spermatogonia through haploid round spermatid cells into the highly specialized, elongated, terminally differentiated spermatozoon that is packaged to deliver the paternal genome to an oocyte. Spermatozoa are left with a limited amount of cytosolic organelles and a complete loss of transcriptional ability during post-testicular maturation, while during epididymal transit, spermatozoa undergo a series of post-translational modifications and acquire new proteins, making it competent to fuse with the oocyte [91]. Most of our knowledge regarding human sperm is based on proteomic analysis of ejaculated human spermatozoa. However, data on the characterization of proteins responsible for spermatogenesis and epididymal maturation are as of yet still unavailable.

According to a systematic compilation by Amaral et al., approximately 6198 different sperm proteins have been identified – about 30% are known to be expressed in the testsis [92]. Gene ontology (GO) analysis revealed that these proteins were involved in various functional pathways, such as metabolism, apoptosis, cell cycle, meiosis, and membrane trafficking, among others. Apart from the expected abundance of cytoskeletal, mitochondrial, flagellar, and membrane proteins, GO analysis has also provided reports on the presence of a large proportion of proteins involved in transcription, protein synthesis, and turnover [93,94].

Human sperm proteins of testicular and epididymal origin were characterized by Li et al. and his group [95,96]. In their report, 112 of 319 identified human sperm-located proteins (35%) were exclusively of testicular-origin, 152 (48%) were exclusively of epididymal origin and 55 (17%) were common to both organs [96]. The report suggested that 47% of the identified proteins were intrinsic sperm proteins expressed at the spermatid stage while 23% were extrinsic sperm proteins that were acquired during epididymal transit and had an epididymal origin [95].

It was further validated that these proteins had a specific pattern of localization in sperm based on their origin. The epididymal proteins were more specifically concentrated at the principal piece while less so in the acrosomal area in comparison to testicular proteins [96]. HSPA2, a novel, testis-specific member of the HSP70 family, is expressed in the sperm plasma membrane along with another member HSP5 [97]. Underexpression of HSPA2 in oligoteratozoospermic men [98] suggests that human HSPA2 may act as a marker for sperm maturity [99]. It is postulated to play a secondary role in the remodeling of the sperm plasma membrane during spermatogenesis [100] to facilitate the formation of the ZP-binding sites. Another study also found this major sperm component to be underexpressed in oxidatively stressed spermatozoa [101]. Recent evidence states that HSPs, including 60, 70, and 90, migrate to the sperm surface and undergo phosphorylation during capacitation [102] since tyrosine-phosphorylated sperm heads were able to bind to the zona pellucida [103]. Wang et al identified about 4675 proteins in human spermatozoa of which only 227 proteins had a testicular origin [104].

**Sperm proteomics and male infertility**

In the field of male infertility, where conventional semen analysis sometimes fails to report the underlying cause of male factor infertility, recent advances in proteomic research have provided important biomarkers for diagnosis and prognosis of male infertility. One of the initial reports in the study of sperm defects through the use of 2D proteomics reported the proteomic mapping of a patient who experienced a failure in in vitro fertilization, where 20 different proteins were identified as compared to controls [105]. Subsequently, a differential expression profile of 17 proteins was observed in asthenozoospermia [106], and further reports demonstrated that clusterin and semenogelin were increased in these patients [101,107]. In an independent study, oligoasthenozoospermic (OAT) patients exhibited significantly higher incidence of heterozygosity for cytosine-adenine-guanine repeats and a higher mtDNA content with a lower percentage of sperm expressing polymerase γ and transcription factor A, mitochondrial in the spermatozoa of OAT group [108].

In globozoospermic patients, 35 proteins were differentially expressed, of which 9 proteins were upregulated and 26 proteins were downregulated in round-headed spermatozoa compared with normal spermatozoa [109]. These differentially expressed proteins were suggested to be involved in a variety of cellular processes and structures, including spermatogenesis, cell skeleton, metabolism, and motility of spermatozoa.

A focused study on proteomic signatures in infertile patients whose semen parameters were normal but with failed in vitro fertilization revealed 24 proteins with altered expression in the normozoospermic infertile group. These proteins were reported to be involved in energy production, structure and movement, and cell signaling and regulation [90]. In another study on patients with idiopathic infertility, a comparison of sperm proteome in men whose sperm resulted in a clinical pregnancy via assisted reproductive technology (ART) from those who did not, revealed 21 proteins to be differentially expressed (>1.2-fold) [85].

Abnormal DNA fragmentation and/or motility is associated with four main protein groups, viz. (i) sperm nuclear proteins such as the sperm protein associated with nucleus in the X chromosome (SPANX) isoforms and several types of histones [110]; (ii) mitochondria-related functions and oxidative stress proteins including mitochondrial ferritin,
mitochondrial single-stranded DNA binding protein, and several isoforms of Peroxiredoxins; (iii) proteins related to sperm motility such as microtubule-based flagellum and spindle microtubule, and (iv) proteins related to the ubiquitin–proteasome pathway.

In an attempt to detect the differential protein profile in men with and without varicocele using 2D gel electrophoresis; 15 consistent differences in protein expression were identified in patients with varicocele when compared with controls [111]. Heat shock proteins, mitochondrial proteins, and cytoskeleton proteins are the major proteins affected by varicocele. Particularly, a significant upregulation of heat shock proteins (HSP70 and HSP90) was observed in subjects with varicocele which was further validated by Western blot and immunocytochemistry [112].

In a recent study, 369 differentially expressed proteins (DEP) were detected in infertile patients with unilateral varicocele in comparison to fertile donors [111]. Of the total reported DEPs, 120 proteins were unique to the fertile group, and 38 were unique to the unilateral varicocele group 114 overexpressed and 97 underexpressed in the unilateral varicocele group. The major functional pathways of the 359 DEP involve metabolism, disease, immune system, gene expression, signal transduction, and apoptosis.

In another study comparing sperm of men with bilateral varicocele with fertile controls, 73 DEP were reported, the majority of which associated with metabolic processes, stress response, oxidoreductase activity, enzyme regulation, and immune system processes [112]. Comparison of the sperm proteomes of men with unilateral versus bilateral varicocele demonstrated 64 proteins in the bilateral group and 31 proteins in the unilateral group [7]. Core functions of the top protein interaction networks were post-translational modification, protein folding, free-radical scavenging, cell death, and survival.

In another study, the sperm proteome from groups of men with high levels of reactive oxygen species (ROS+) were compared with that of men with low or physiological levels of ROS (ROS−) [101,113]. In the first case [113], a 2D-DIGE with differential fluorescence tagging and in gel digestion of the proteins followed by LC–MS detection. A total of 31 spots were differentially expressed with 6 significantly decreased and 25 increased in the ROS− sample compared with the ROS+ sample. The result showed overabundance of four antioxidant proteins in ROS− sperm that may exert essential cytoprotective effects against the build-up of ROS levels: lactotransferrin isoform 2, lactotransferrin isoform 1 precursor, peroxiredoxin-1, and Mn-SOD mitochondrial isoform A precursor [113]. In the second study, the samples were subjected to LC–MS/MS analysis through in-solution digestion of proteins for peptide characterization [101]. The findings revealed a total of 74 proteins in spermatozoa, out of which 15 proteins with more than twofold difference were overexpressed (ROS+ group) when compared to the ROS− group. The overexpressed proteins comprised the histone cluster 1 H2ba (HIST1H2BA); mitochondrial malate dehydrogenase precursor (MDH2); heat shock protein 90 kDa beta, member 1 (HSP90B1); heat shock 70 kDa protein 5 (HSPA5); glutamine synthetase (GLUL); transglutaminase 4 (prostate) (TGM4); glutathione peroxidase 4 isoform A precursor (GPX4); sperm acrosomal membrane protein 4 (SPACA4); olfactomedin 4 precursor (OLFM4); and chromosome 20 open reading frame 3 (C20orf3), whereas the underexpressed proteins included semenogelin II precursor (SEMG2); peroxiredoxin 6 (PRDX6); clathrin heavy chain 1 (CLTC); eukaryotic translation elongation factor 2 (EEF2); and enolase 1 (ENO1). Furthermore, there was significant change in localization profile of the differentially expressed proteins. While the overexpressed proteins were predominantly found in the cellular compartments of intracellular, organelle, macromolecular complex region and mitochondria; the underexpressed proteins were dominated in the cytoplasm extracellular, plasma membrane, protein complex (e.g. enolase) and the vesicular region (e.g. clathrin heavy chain). Remarkably, the endosome, lipid particle, membrane-bound organelles and the microtubules showed an exclusive abundance of overexpressed proteins in ROS+ group; while the proteinaceous extracellular matrix was restricted only to the underexpressed proteins. In a recent study, comparisons were made between protein profiles of spermatozoa from fertile donors and infertile men with varying levels of ROS. Identification of six DEPs (Calmegin, Tripeptidyl peptidase II, Dynine intermediate chain 2, axonemal, heat shock 70 kDa protein 4 L, early endosome antigen 1, and plasma serine protease inhibitor) that were present in all the three ROS groups with varying expression levels suggests that they may serve as potential candidates of oxidative stress markers [114].

Many of the proteins in its precursor or preprotein form have been identified in studies relating to male fertility anomalies, and it may be speculated that the accumulation of these protein is an indication of generalized defects at the post translational level. Assumptions are also made regarding the down regulation of some of the downstream functions related to these proteins [101,106,115]. Such differential alteration pattern may be due to PTM which could be used as an effective diagnostic strategy valuable for understanding male factor infertility.

Ubiquitination and acetylation were reported to be the principal PTM found in non-obstructive azoospermia (NOA). Impairment in production of different enzymes belonging to UPS, namely UBE2B protein (an ubiquitin-conjugating enzyme) may cause defects in spermatozoa maturation and result in NOA [116,117]. Besides, mutations in genes encoding for testis-specific ubiquitin proteases, namely USP26 [118] and USP9Y [119], as observed in azoospermic patients, elucidates the role of protein ubiquitination in azoospermia. On the other hand, hyperacetylation of histone H4 is reported in obstructive azoospermic spermatids [120]. USP might play some regulatory role in oligozoospermic patients [121]. Differential expression of ubiquitin-conjugating enzyme UBE2B [117,122], and protease USP26 [123] was noticed in
Potential involvement of altered protein phosphorylation events, such as phosphorylation of the gamma-tubulin complex associated protein 2 (GCP2) in the pathogenesis of asthenozoospermia (AZS) [124] suggests the role of this PTM in sperm motility. Parte et al. [125] further established this by identifying 66 differentially expressed phosphoproteins in AZS sperm, which comprised predominantly of the HSPs, cytoskeletal proteins, proteins associated with the fibrous sheath, and those associated with energy. Furthermore, a correlation between deficiency of tyrosine phosphorylation of tail proteins, especially those related to hyperactivated motility and AZS was also reported [126]. Nevertheless, decreased membrane fluidity is hypothesized to be the cause of hypophosphorylation of proteins during capacitation in these groups [127]. In two separate proteomic studies, several proteins having S-nitrosylation sites were identified to be differentially expressed in AZS sperm [106,128]. Similarly, testis-specific alpha-tubulin isoforms TUBA3 C and TUBA8, in their acetylated state, was decreased, and TUBA4A was increased in AZS as compared with normal spermatozoa [129]. A variation of ubiquitin protease USP26 was also reported to be directly associated with human sperm motility [123]. Moreover, in some AZS patients, a severe form of structural anomaly is found in the sperm known as dysplasia of the fibrous sheath (DFS). A study demonstrated that sperm with DFS were accompanied by increased ubiquitination of the sperm surface as well as the sperm mitochondrial proteins [130]. In AZS patients only, percentage of SUMO1-positive spermatozoa was found to be inversely correlated with total and progressive motility [56]. In addition, an association of AZS with anomalies in seminal glycome has also been documented [131,132].

Teratozoospermia patients (TZS) are documented to present differential tyrosine phosphorylation of six sperm proteins in comparison to normozoospermic men [133]. Furthermore, spermatozoa from TZS males exhibited a compromised ability to undergo tyrosine phosphorylation following capacitation [134]. Significant negative correlation with the phosphorylated levels of protein Stat3 was also observed in TZS [134]. Similarly, ejaculated sperm of the TZS group demonstrated increased binding of anti-ubiquitin antibodies to the sperm surface, reflecting the occurrence of enhanced sperm surface ubiquitination in nonsystematic cases of TZS [135].

A comparative study on assessment of tyrosine phosphorylation status in spermatozoa of varicocele patients revealed a reduced level of phosphorylation [136], which is attributed to an alteration in plasma membrane dynamics resulting in sperm dysfunction in these patients. Furthermore, enzymes linked to protein glycosylation namely acid beta-glucuronidase, alpha-mannosidase, alpha-glucosidase, alpha-galactosidase, beta-galactosidase, and beta-N-acetylglucosaminidase showed overexpression and hyperactivity in seminal plasma of infertile men with varicocele [137]. As these extracellular glycosidases are implicated to cause the modification of sperm plasma membrane glycoproteins during sperm maturation process, it could be postulated that such overexpression of acid glycosidase in the seminal plasma would mask the carbohydrate-containing molecules present in ejaculated spermatoza of varicocele group, altering their fertilization potential. One of the major pathways of PTM and protein folding found to be severely compromised in varicocele individuals is the UPS [7]. In addition, assessment of the ubiquitination median, a marker for functionality of the UPS was also lower in varicocele group, where it was reported to show a negative correlation with sperm morphology [138].

Expert commentary

In the post-genomic era, studies have mainly focused on the identification of novel protein biomarkers in complex biological systems. A biomarker is a distinctive biological or biologically derived indicator of a process, event, or condition, and is considered ideal if it serves the purpose of screening, diagnosis, and monitoring disease activity. In addition, they may hold responsible for targeted therapy or would assess therapeutic response [139]. With regard to male factor infertility, the main objective of a biomarker is to evaluate in an accurate and minimally invasive manner, the male potential to father a child. Furthermore, use of semen for proteomic analysis is complicated because it contains sperm as well as seminal plasma and the most important challenge in sperm proteomics is the fact that spermatozoa undergoes distinct physiological changes after ejaculation. The advantages of using proteome over routine semen analysis is that it confirms the qualitative and quantitative measure of a protein within the spermatozoa. Nevertheless, identification of the ideal biomarker for clinical diagnosis in such a complex scenario to distinguish the pathophysiological state is one of the main challenges. Furthermore, in a clinical proteomics study one sample per individual is used with the assumption that this sample, is representative of a disease. However, it is more likely representative of the time of day it was drawn, the nutritional status and life style of the patient, and their genetic background rather than the pathology under investigation. Therefore, longitudinal sampling where samples are collected over time to determine how the proteome changes over time within an individual is necessary for discovery of putative biomarker. Personalized reference ranges are where medicine will have a tremendous impact; discovery-based proteomics using longitudinal sampling will be critical to elucidate predictive biomarkers.

As vide supra in the past two decades, studies on proteomic profiling of spermatozoa result in the generation of an increasing amount of data. The information generated mainly comprises of complex protein compositions and a large number of isolated proteins whose implications in the biological state is yet to be discovered. This limits the use of proteomic approaches to elucidate the role of biological markers [140]. Furthermore, an additional confounding factor is the inherent complexity of the disease, therefore, the pathogenesis cannot be singled out by a protein, but might indeed be a result of multiple interacting proteins.
across multiple pathways. Thus, employing a panel of biomarkers for identification of infertility is more likely to be successful. With relation to oxidative stress in the spermatozoa, proteins such as MDH2, TGM4, GPX4, GLUL, HSP90B1, and HSPA5 are suggested as possible biomarkers due to their increased expression in samples with high levels of ROS compared with those without oxidative stress [141]. Data from a number of independent laboratories suggests that an underrepresentation of a key molecular chaperone, HSPA2 is responsible for impaired sperm–egg recognition [142] which is reported to be modulated by oxidative stress (lipid aldehydes) [143]. While considering the use of multiple interacting-markers to provide specific valid signatures, the problems inherent to validation increases exponentially [144]. The question is how to test the validity. Is the change in expression of a candidate protein directly associated with the disease or is it due to the presence of a PTM(s) or might it only be an artefact resulting from technical variability? This fact is further compounded by the identification of proteins possessing PTM within the normal and abnormal spermatozoa, specific to different pathological states, such as oxidative stress. Nevertheless, the current trends to identify PTMs are only predictive [145]. Therefore, creation of new pipelines as well as analytical methods [146] that allow for analysis of specific PTMs on sperm function will impact the understanding of protein function in healthy and infertile spermatozoa. A study that attested to the inherent technical variability present in proteomic profiling, as published by members of the Human Proteome Organization (HUPO) showed that when an equimolar protein sample containing 20 human proteins was given to 27 different laboratories, only 7 out of 27 labs were able to correctly identify all the 20 proteins in the sample. The discrepancies in the results were attributed to the databases used for analysis, as they contain erroneous confidence levels, thereby missing several of the 20 proteins present in the sample [147].

Another aspect of understanding the sperm proteome in different pathophysiological states is to identify the abnormality in structure and function of the protein so as to target them for development of reversible male contraception. This necessitates the identification of sperm proteins on different domains, which are acquired by sperm mostly as they traverse the epididymis. The sperm proteome is dynamic due to the fact that proteins are transported into the spermatozoa from the epididymal, seminal, or prostatic fluid. Epididymal proteins are preferred targets for contraception development over testicular proteins, as the blockage of sperm maturational events will not obstruct spermatogenesis and testicular endocrine function, as well as allow for reversibility [148,149]. Both pre- and postfertilization events are possibly affected by targeting the sperm head and the flagellar proteins. Evaluation of sperm proteins thus allows for a straightforward assessment of testicular homeostasis, and is an increasingly important area of study in male infertility.

Five-year view

Among the main challenges that will need to be dealt with in studying the infertile male, sperm proteomics studies will greatly benefit from increased sensitivity, specificity and reproducibility of MS-based proteomics. Studying post-translational modifications will further allow for determination of understanding how the dynamic equilibrium of fertile sperm, co-inhabiting with altered sperm, may lead to positive or negative effects to fertility. Furthermore, the clinical impact of proteomics to sperm biology and functional state may (i) allow for determination of the biological systems leading to the observed cellular integrity and (ii) allow for pathway specific intervention. In a recent report, Intasqui et al. demonstrated a number of proteins associated with sperm DNA fragmentation, some of which are transferred specifically during epididymal transit through epididymosomes [9]. It is interesting that multiplatform studies may indeed enhance our current knowledge of sperm biology, as well as of male fertility.

Acknowledgement

Financial support for this study was provided by the American Center for Reproductive Medicine, Cleveland Clinic.

Financial & competing interests disclosure
The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

ORCID
Luna Samanta http://orcid.org/0000-0002-2969-0071
References

Papers of special note have been highlighted as:
• of interest
•• of considerable interest


* A key finding that establishes sperm proteome changes during epididymal transit.


46. One of the first papers to propose the targeted data extraction of tandem mass spectra by data-independent acquisition.


50. One of the first papers to propose the targeted data extraction of tandem mass spectra by data-independent acquisition.

51. One of the first papers to propose the targeted data extraction of tandem mass spectra by data-independent acquisition.

52. One of the first papers to propose the targeted data extraction of tandem mass spectra by data-independent acquisition.


54. One of the first papers to propose the targeted data extraction of tandem mass spectra by data-independent acquisition.

55. One of the first papers to propose the targeted data extraction of tandem mass spectra by data-independent acquisition.


67. This paper explains the importance of post-translational protein modifications of spermatozoa during epididymal maturation.


This paper reveals the importance of differential protein phosphorylation in asthenozoospermia.

- A crucial study on reproducibility of mass spectrometry-based proteomic data.
- A toolkit for simulation of protein post-translational modification to further our understanding on protein modification and its consequences.
- A proteomic study of sperm proteome expression in a pathological condition such as asthenozoospermia.