Review Article

The Profile of Human Sperm Proteome; A Mini-review

Kambiz Gilany [1*], Niknam Lakpour [2], Mohtaram Vafakhah [1], Mohammad Reza Sadeghi [3]

1- Reproductive Biotechnology Research Center, Avicenna Research Institute, ACECR, Tehran, Iran
2- Nanobiotechnology Research Center, Avicenna Research Institute, ACECR, Tehran, Iran
3- Monoclonal Antibody Research Center, Avicenna Research Institute, ACECR, Tehran, Iran

* *Corresponding Author:*
Kambiz Gilany,
Department of Embryology, Reproductive Biotechnology Research Center, Avicenna Research Institute, Shahid Beheshti University, Tehran, Iran
E-mail:
k.gilany@avicenna.ac.ir

Received: Feb. 19, 2011
Accepted: Jun. 29, 2011

Abstract

New advances in mass spectrometry-based proteomics technology are having a major impact on our understanding of how human spermatozoa acquire their capacity for fertilization. A complete analysis of the proteins found in the human spermatozoa is essential for understanding the events leading up to, and including, fertilization and early embryo development. In this short review, we have collected the human sperm proteome from the literature and analyzed it by the Database for Annotation, Visualization and Integrated Discovery (DAVID) software. Bioinformatics analysis demonstrated that the collected 1,300 proteins were involved in various metabolic pathways including catabolic processes. Additionally, the majority of the collected human sperm proteome belonged to cytoplasm. Application of the multi-dimensional protein identification technology (MudPIT) for obtaining a better coverage of the hydrophobic and basic proteins of the human sperm proteome is recommended.

Keywords: DAVID bioinformatics, Gene ontology, Mass spectrometry-based proteomics, Proteome, Sperm.
To cite this article: Gilany K, Lakpour N, Vafakhah M, Sadeghi MR. The Profile of Human Sperm Proteome: a Mini-review. J Reprod Infertil. 2011;12(3):193-199.

Introduction

Spermatogenesis is a unique process in male gender to produce male haploid germ cells from diploid progenitor cells. Spermatogenesis includes two sequential division of meiosis that convert one diploid spermatogia cell to four haploid cells. Spermatogenesis is a specialized process of differentiation of haploid round spermatid cells to the highly specialized sperm cell, the spermatozoon. Sperm function is to deliver the paternal genome to the oocyte.

Identification of protein molecules involved in sperm function, fertilization and early embryo development increase our knowledge about sperm biology and it will be applied in reproductive medicine and treatment of some inborn genetic diseases to generate a healthier offspring. The importance and the easy accessibility of sperm cells have favored the study of its composition and mechanisms involved in its differentiation and function (1 - 5). The protein content of the sperm was one of the first cells to be studied. It was the pioneering work done by Friedrich Miescher in 1874 that led to the isolation and identification of protamine. Recently, use of mass spectrumetry-based (MS) proteomics technology has further contributed to the identification of the proteome that make up spermatozoa (6 - 12).

In the current short review we focused on the proteins of the spermatozoa identified by MS proteomics technology. As a methodological approach we considered for inclusion all the articles retrieved from PubMed search with the keywords "human", "sperm", "spermatozoa", "spermatozoon", combined with the key word "proteome", "proteomics" or "mass spectrometry". We analyze the collected human sperm proteome by the

Database for Annotation, Visualization and Integrated Discovery (DAVID) software. Using the DAVID software we particularly focused on the enriched biological themes, gene ontology (GO) terms, and discovered enriched functional-related gene groups (13).

Proteome definition: The proteome has been defined as the protein complement of the genome. However, the definition of proteome has changed since it was first defined by Wilkins et al. in 1995 (14). Today, the term 'proteome' has developed to be: "The proteome of an individual is defined by the sum and the time dynamics of all protein species occurring during the life-time of this individual". This definition of proteome includes the protein expression of the individual protein, the isoforms of a protein and post-translational modifications of a protein (15).

Techniques used in human sperm proteome mapping: There are several initial reports using MS proteomics technology to identify a limited number of proteins from the human sperm using two-dimensional gel electrophoresis (2-DE) coupled to MALDI-TOF-MS analysis (16 - 23). An extensive human sperm proteome analysis using 1D-SDS-PAGE combined with electrospray liquid chromatography tandem mass spectrometry (GeLC-MS/MS) approach identified 1,760 proteins (24). However, no protein list was published. The only far-reaching human sperm proteome analysis available to date is work done by Baker et al. (25). Using GeLC-MS/MS technique, they were able to map 1,056 unique proteins from the human spermatozoa. Literature review of the distribution of techniques used for mapping human spermatozoa showed that two studies had used GeLC-MS/MS. Additionally; 2-DE had been used in 8 studies to map human sperm proteome. To our best knowledge, no other techniques had been used for proteome profiling of human spermatozoa, including multidimensional protein identification technique (MudPIT) or combined fractional diagonal chromatography (COFRADIC) technique. A more extensive human sperm proteome could be obtained by combining different MS proteomics techniques. We have shown that different MS proteomics techniques are able to identify a unique set of proteins (26).

How many proteins are expressed in human sperm?: One of the big questions in the proteome analysis has been how big the human proteome size is? The near-complete sequencing of the human genome has yielded the total gene estimates that, at first glance, seem surprisingly low; of the order of 30000 open reading frames (27, 28). However, when a gene is expressed it is subjected to alternative splicing mechanisms and post-translational modifications. It is estimated each gene could produce between 5 to 6 mRNAs by an alternative splicing mechanism and each of these mRNA species is in turn translated into proteins that are processed in various ways, generating on the order of 8–10 different modified forms of each polypeptide chain. Thus, the human genome may potentially produce on the order of (30000×6×10) 1.8 million different protein species (29). Defining each and every one of these proteins is what global collaborations, such as the Human Proteome Organization (HUPO) [1] is set to undertake.

The question 'how many proteins, the most highly differentiated and unique cell type in the human body, the spermatozoa, contain?' is often posed in the literature (25, 30). Of course, it is quite difficult to predict the size of the human spermatozoa proteome from the existing proteomics data, knowing the current limitation of MS proteomics technology (26, 31 - 33). However, Baker et al (30) used the current proteomics data available from yeast proteome to predict the number of protein species of the human spermatozoa to be 2000-2500. As Baker et al. also point out, this is much lower than the identified proteome of bovine sperm (~ 4000) (34). However, Baker et al. argue that the high number of protein identified in the bovine sperm proteome is caused by false positive identification (30).

Collected human sperm proteome analyzed by DAVID: The collected human sperm proteome were functionally categorized based on Gene Ontology (GO) annotation terms using the Database for Annotation, Visualization and Integrated Discovery (DAVID) program package[2] (13, 35 - 37). For any gene or protein list, DAVID software tools are able to identify enriched biological themes, particularly GO terms, discover enriched functional-related gene groups, visualize genes or proteins on BioCarta and KEGG pathway maps, explore gene or protein names in batch, link gene-

1- http://www.hupo.org/
2- http://david.abcc.ncifcrf.gov/

disease associations, etc. Approximately 1,300 proteins of the human sperm cell, sum of 2-DE and GeLC-MS/MS techniques, were analyzed by DAVID software.

Biological function of human sperm proteome: Table 1 shows the ten most important catalogue outputs for biological function analysis by DAVID software DAVID software was only able to catalogue 793 of the submitted proteins. This means that biological functions of about 500 proteins out of the collected human sperm proteome are still unknown. As it is shown in the Table 1, the most important biologically functional proteins in the human sperm proteome belong to catabolic processes (16%), including proteins for the breakdown of carbon compounds with the liberation of energy used for sperm movement. DAVID categorized glucose catabolic processes and oxidative phosphorylation which is necessary for the homeostasis. In the table, we also find proteins belonging to spermatogenesis (3.6%) and spermiogenesis (0.9%).

Cellular component of human sperm proteome: Table 2 shows the top ten outputs of cellular localization of the collected human sperm proteome from DAVID software. The software was able to map 850 of the identified proteins. Around 450 of submitted proteins to DAVID were categorized as unknown localization.

Surprisingly, the most enriched groups from the collected human sperm proteome belong to cytoplasma (59%, 7.9E-48). It is well known that the human sperm lost most of its cytoplasm during spermiogensis process. A large number of proteins were categorized to be from mitochondria. Mitochondrial protein is not astonishing since the neck of human sperm is rich in mitochondria. Additionally, protein enriched parts belonging to the tail of human sperm were identified as cytoskeleton (12.6%, 1E-9) and flagellum (1.5%, 6.6E-8).

As it is shown in Table 2 no transmembrane proteins were categorized from the collected human sperm proteome which are important types of proteins for the oocyte and sperm interaction. This probably is caused by MS proteomics techniques used for the proteome mapping of human sperm. It is a well-known fact that the hydrophobic proteins, such as transmembrane proteins, rarely appear in gel-based techniques (26). Using gel-free techniques, such as MudPIT, will improve the deeper coverage of human sperm proteome. However, MudPIT is not a straightforward technique and it needs some expertise (38, 39).

Table 1. Tabulated are the ten important biological functions with the greatest statistical significance for enrichment in the collected proteome data set of the human sperm (GOTERM: level ALL)

Biological functions	%	P-value
Catabolic processes	16	1.6E-24
Proteasomal ubiquitin-dependent protein catabolic processes	4.1	4.4E-24
Proteasomal protein catabolic processes	4.1	4.4E-24
Generation of precursor metabolites and energy	6.4	4.6E-20
Glucose catabolic processes	2.6	1E-16
Cell cycle processes	7.4	5.5E-12
Cell redox homeostasis	1.4	3.8E-5
Spermatogenesis	3.6	3.8E-5
Oxidative phosphorylation	1.6	3.5E-4
Spermiogenesis	0.9	1E-2

The percentage is calculated as: involved proteins divided by the total number of proteins multiplied by one-hundred. The enrichment P-value (compared to the theoretical human proteome) is calculated based on EASE Score, a modified Fisher's Exact Test and ranges from 0 to 1. Fisher's Exact P-value=0 represent perfect enrichment. Usually the P-value must be equal to or smaller than 0.05 to be considered strongly enriched in the annotation categories. The closer the value is to zero, the more enriched is the category

Table 2. Tabulated are the top ten important molecular functions with the greatest statistical significance for enrichment in the collected proteome data set of the human sperm (GOTERM: level ALL)

Cellular localization	%	P-value
Cytoplasm	59	7.9E-48
Mitochondrion	15.8	1.1E-32
Proteasome complex	3.6	4.5E-29
Mitochondrial part	9.7	1.1E-23
Intracellular	70	8.3E-22
Mitochondrial matrix	5.5	3.6E-21
Cytoskeleton	12.6	1.0E-9
Flagellum	1.5	6.6E-8
Eukaryotic translation elongation factor 1 complex	0.5	3.9E-5
Cilia	1.7	1.4E-3

Explanations for the percentage and p-values can be found in Table 1

Functional categorization of the collected human sperm proteome: The most statistically significant functional annotation by DAVID software were the actylated proteins (36.7, 2.2E-89) and phosphoprotein (47.4%, 1.8E-16) groups. This is to our knowledge that the most post-translated proteins identified so far were identified by using techniques such as 2-DE and GeLC-MS/MS (26, 40). However, the exact function of these large numbers of post-translational modifications is unknown (personal communication with Baker M, author of the largest human sperm proteome published to date (25)).

Metabolic pathway enriched in the collected human sperm proteome: One of the functions of DAVID software is to show the enriched KEGG pathways. The most significant metabolic pathway which were enriched in the collected human sperm proteome were proteasome (3%, 2E-22), fatty acid metabolism (1.6%, 3.3E-8), TCA cycle (1.4%, 5.8E-8), Glycolysis/Gluconegenesis (1.9%,7.7E-8) and pyruvate metabolism (1.4%,1.9E-6). Observing the enrichment of fatty acid and pyruvat metabolism is not surprising since sperm is under hypoxic condition.

Sperm: a silent cell?: One of the discussions in sperm cell biology is whether any protein synthesis takes place in the sperm cells or not? (30, 41). Martinez-Heredia et al (21) identified transcription factor proteins in the proteome mapping of human sperm using 2-DE technique, in the pI range 5-8. Additionally, in the analysis of the collected data on human sperm proteome by DAVID software we are able to localize protein in the eukaryotic translation elongation factor 1 complex (Table 2). However, a confirmation of these proteins by Western blotting technique is necessary in order to show that a protein synthesis actually takes place in sperm cells.

Conclusion

Although, the sperm protein content was one of the first cells to be analyzed, there is still a limited number of identified human sperm protein compared to other samples, such as brain proteome (7792 proteins) or the human neuroblastoma cell line SH-SY5Y proteome (3707 proteins) (42, 43). A deeper coverage of the human sperm proteome can be obtained using gel-free techniques such as MudPIT or COFRADIC (44, 45). It is well-established today that gel-free techniques have a better performance for the identification of basic, acidic and hydrophobic proteins than gel-based techniques (39, 46, 47). Chu et al (48) were able to identify very basic proteins using the MudPIT technology from *C. elegans* sperm proteome which is impossible to identify by gel-based techniques. Additionally, it should be kept in mind that a proteome is much more complex than a genome. The absence of a particular protein from any MS proteomics list does not necessarily mean that it is not present in the spermatozoa of that species. An alternative explanation is that the proteomic coverage could have been incomplete,

the protein had been in too low abundance or the protein in question might have been missed by chance. Although, the human sperm proteome is small and less complex than other cells, the functions of many of the identified proteins of human sperm are still unknown at the present. Immunolocalization can be readily used to obtain some clues to their function through determining their location within the sperm and the expression pattern of the corresponding proteins. Also knockouts, knockdowns and conditional knockdowns should further contribute to the identification of their function. As the sperm proteome from different species becomes available, the comparison of conserved proteins and domains would also provide important clues towards the essential conserved functions and evolution of sperm proteins.

Acknowledgement

Authors declare no conflict of interest.

References

1. Baccetti B, Afzelius BA. The biology of the sperm cell. Monogr Dev Biol. 1976;(10):1-254.
2. Mezquita C. Chromatin composition, structure and function in spermatogenesis. Revis Biol Celular. 1985;5:V-XIV, 1-124.
3. Oliva R, Dixon GH. Vertebrate protamine genes and the histone-to-protamine replacement reaction. Prog Nucleic Acid Res Mol Biol. 1991;40:25-94.
4. Dadoune JP. Expression of mammalian spermatozoal nucleoproteins. Microsc Res Tech. 2003;61(1): 56-75.
5. Ainsworth C. Cell biology: the secret life of sperm. Nature. 2005;436(7052):770-1.
6. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422(6928):198-207.
7. Conner SJ, Lefièvre L, Kirkman-Brown J, Michelangeli F, Jimenez-Gonzalez C, Machado-Oliveira GS, et al. Understanding the physiology of prefertilisation events in the human spermatozoa--a necessary prerequisite to developing rational therapy. Soc Reprod Fertil Suppl. 2007;63:237-55.
8. Domon B, Aebersold R. Mass spectrometry and protein analysis. Science. 2006;312(5771):212-7.
9. Bailey JL, Tardif S, Dubé C, Beaulieu M, Reyes-Moreno C, Lefièvre L, et al. Use of phosphorproteomics to study tyrosine kinase activity in capacitating boar sperm. Kinase activity and capacitation. Theriogenology. 2005;63(2):599-614.
10. Bohring C, Krause W. The characterization of human spermatozoa membrane proteins--surface antigens and immunological infertility. Electrophoresis. 1999;20(4-5):971-6.
11. Miller MA. Sperm and oocyte isolation methods for biochemical and proteomic analysis. Methods Mol Biol. 2006;351:193-201.
12. Wang Y, Zhou ZM. [Update of the researches on sperm proteome]. Zhonghua Nan Ke Xue. 2007;13 (3):250-4. Chinese.
13. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.
14. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, et al. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis. 1995;16(7): 1090-4.
15. Jungblut PR, Holzhütter HG, Apweiler R, Schlüter H. The speciation of the proteome. Chem Cent J. 2008;2:16.
16. de Mateo S, Martínez-Heredia J, Estanyol JM, Domínguez-Fandos D, Vidal-Taboada JM, Ballescà JL, et al. Marked correlations in protein expression identified by proteomic analysis of human spermatozoa. Proteomics. 2007;7(23):4264-77.
17. Shetty J, Diekman AB, Jayes FC, Sherman NE, Naaby-Hansen S, Flickinger CJ, et al. Differential extraction and enrichment of human sperm surface proteins in a proteome: identification of immunocontraceptive candidates. Electrophoresis. 2001;22 (14):3053-66.
18. Pixton KL, Deeks ED, Flesch FM, Moseley FL, Björndahl L, Ashton PR, et al. Sperm proteome mapping of a patient who experienced failed fertilization at IVF reveals altered expression of at least 20 proteins compared with fertile donors: case report. Hum Reprod. 2004;19(6):1438-47.
19. Baker MA, Witherdin R, Hetherington L, Cunningham-Smith K, Aitken RJ. Identification of post-translational modifications that occur during sperm maturation using difference in two-dimensional gel electrophoresis. Proteomics. 2005; 5(4):1003-12.
20. Li LW, Fan LQ, Zhu WB, Nien HC, Sun BL, Luo KL, et al. Establishment of a high-resolution 2-D reference map of human spermatozoal proteins

from 12 fertile sperm-bank donors. Asian J Androl. 2007;9(3):321-9.

21. Martínez-Heredia J, Estanyol JM, Ballescà JL, Oliva R. Proteomic identification of human sperm proteins. Proteomics. 2006;6(15):4356-69.
22. 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(2):436-8.
23. Liao TT, Xiang Z, Zhu WB, Fan LQ. Proteome analysis of round-headed and normal spermatozoa by 2-D fluorescence difference gel electrophoresis and mass spectrometry. Asian J Androl. 2009;11 (6):683-93.
24. 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.
25. Baker MA, Reeves G, Hetherington L, Müller 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(5): 524-32.
26. Gilany K, Van Elzen R, Mous K, Coen E, Van Dongen W, Vandamme S, et al. The proteome of the human neuroblastoma cell line SH-SY5Y: an enlarged proteome. Biochim Biophys Acta. 2008; 1784(7-8):983-5.
27. Harrison PM, Kumar A, Lang N, Snyder M, Gerstein M. A question of size: the eukaryotic proteome and the problems in defining it. Nucleic Acids Res. 2002;30(5):1083-90.
28. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409 (6822):860-921.
29. Jensen ON. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Curr Opin Chem Biol. 2004; 8(1):33-41.
30. Baker MA, Aitken RJ. Proteomic insights into spermatozoa: critiques, comments and concerns. Expert Rev Proteomics. 2009;6(6):691-705.
31. Fago A, Hundahl C, Dewilde S, Gilany K, Moens L, Weber RE. Allosteric regulation and temperature dependence of oxygen binding in human neuroglobin and cytoglobin. Molecular mechanisms and physiological significance. J Biol Chem. 2004;279(43):44417-26.
32. Maes MB, Lambeir AM, Gilany K, Senten K, Van der Veken P, Leiting B, et al. Kinetic investigation of human dipeptidyl peptidase II (DPPII)-mediated hydrolysis of dipeptide derivatives and its identification as quiescent cell proline dipeptidase (QPP)/dipeptidyl peptidase 7 (DPP7). Biochem J. 2005;386(Pt 2):315-24.
33. de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Fröhlich F, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature. 2008;455 (7217):1251-4.
34. Peddinti D, Nanduri B, Kaya A, Feugang JM, Burgess SC, Memili E. Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. BMC Syst Biol. 2008;2:19.
35. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003;4(5):P3.
36. Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome Biol. 2003;4 (10):R70.
37. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9.
38. Wu CC, MacCoss MJ, Howell KE, Yates JR 3rd. A method for the comprehensive proteomic analysis of membrane proteins. Nat Biotechnol. 2003; 21(5):532-8.
39. Lohaus C, Nolte A, Blüggel M, Scheer C, Klose J, Gobom J, et al. Multidimensional chromatography: a powerful tool for the analysis of membrane proteins in mouse brain. J Proteome Res. 2007;6 (1):105-13.
40. Schirle M, Heurtier MA, Kuster B. Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics. 2003;2(12): 1297-305.
41. Oliva R, de Mateo S, Estanyol JM. Sperm cell proteomics. Proteomics. 2009;9(4):1004-17.
42. Wang H, Qian WJ, Chin MH, Petyuk VA, Barry RC, Liu T, et al. Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. J Proteome Res. 2006;5(2):361-9.
43. Birkeland E, Nygaard G, Oveland E, Mjaavatten O, Ljones M, Doskeland SO, et al. Epac-induced Alterations in the Proteome of Human SH-SY5Y

Neuroblastoma Cells. J Proteomics Bioinform. 2009;2:244-54.

44. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR, et al. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat Biotechnol. 2003;21(5):566-9.
45. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19(3):242-7.
46. Chen EI, Hewel J, Felding-Habermann B, Yates JR 3rd. Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT). Mol Cell Proteomics. 2006;5(1):53-6.
47. Staes A, Van Damme P, Helsens K, Demol H, Vandekerckhove J, Gevaert K. Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics. 2008;8(7):1362-70.
48. Chu DS, Liu H, Nix P, Wu TF, Ralston EJ, Yates JR 3rd, et al. Sperm chromatin proteomics identifies evolutionarily conserved fertility factors. Nature. 2006;443(7107):101-5.