Original article

Human sperm proteins identified by 2-dimensional electrophoresis and mass spectrometry and their relevance to a transcriptomic analysis

Karolina Nowicka-Bauer[a], Małgorzata Ozgo[b], Adam Lepczynski[b], Marzena Kamieniczna[a], Agnieszka Malcher[a], Wiesław Skrzypczak[b], Maciej Kurpisz[a,*]

[a] *Institute of Human Genetics, Polish Academy of Sciences, Strzeszynska 32, 60-479 Poznan, Poland*
[b] *Department of Physiology, Cytobiology and Proteomics, West Pomeranian University of Technology, 26 Doktora Judyma Street, 71-466 Szczecin, Poland*

ARTICLE INFO

ABSTRACT

The aim of this study was to identify and analyse human sperm proteins from normozoospermic men using 2-dimensional electrophoresis (2-DE) and mass spectrometry (MS). We identified 73 different sperm proteins, including two less characterized human sperm proteins, Annexin A7 (ANXA7) and c14orf105. Bioinformatic analysis of detected sperm proteins revealed new carbohydrate and lipid metabolic pathways, which supply energy to motile sperm. A comparison of our data with available mRNA microarray data from the human testis allows for validation of identified sperm proteins and aids in the recognition of their physiological pathways.

1. Introduction

The destiny of spermatozoa is to fertilize oocytes, and, to achieve this goal, male gametes must be equipped with fully functional machinery, including specific proteins. Compared to somatic cells, spermatozoa are highly sophisticated cells with specific components. To address infertility problems, it is important to understand the protein composition of sperm cells, and this can be achieved by proteomics.

In the last decade, due to the development of novel proteomic techniques, studies on the human sperm proteome have intensified. The various proteomic approaches used to identify sperm proteins have been shown to have variable efficiencies. Although two-dimensional electrophoresis (2-DE) has led to the identification of many sperm-specific proteins and their isoforms, traditional sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) is still used for protein resolution [1]. In some cases, separation of sperm proteins by electrophoresis was omitted and direct in-solution digestion of proteins was performed [2]. In addition, the type of mass spectrometry (MS), such as matrix assisted laser desorption/ionisation-time of flight (MALDI-TOF) or liquid chromatography-tandem MS (LC–MS/MS), used may affect the identification of sperm proteins. Although more sperm proteins have been identified by LC–MS/MS, many sperm proteins have also been identified by MALDI-TOF MS [3]. Nevertheless, it should be noted that much of our knowledge of the human sperm proteome is a result of studies on male infertility and/or sperm dysfunction. The revision made by Amaral et al. [4] included data from proteomic studies

on human sperm from normozoospermic individuals, from men with different sperm pathologies [5,6], from infertile men with idiopathic background [7], and from a functional proteomics study [8]. For a list of the proteins identified in human sperm pathologies, refer to the review by Nowicka-Bauer & Kurpisz [9], but note that several other proteomic studies of human sperm have been published [10], such as studies of differentially expressed sperm proteins in asthenozoospermia [11], in infertile men with unilateral varicocele [12], in normozoospermic men with idiopathic infertility [13], and in the sperm of tobacco smokers [14]. Further, the proteomic profiles of mature and immature sperms were compared [15], and the proteomic profiles of spermatozoa that were isolated by the swim-up method or by density gradient centrifugation were comparatively studied [16].

In this study, we aimed to create a human sperm proteome map from normozoospermic men and to compare our results with data from other human sperm proteome studies that used the same proteomic approaches of 2-DE/MS and microarray data.

2. Materials and methods

2.1. Sample preparation

Human semen samples were obtained from ten normozoospermic individuals who volunteered to participate in this study and signed informed consents. Semen was collected by masturbation into sterile disposable containers after 3–5 days of sexual abstinence and allowed

to liquefy for 30–60 min. Then, routine seminal analyses were performed according to the World Health Organization guidelines [17].

To generate a general protein profile of the sperm samples collected, sperm cells were isolated from semen using a single layer of 50% Percoll [5]. Total sperm pellets were solubilized in a lysis buffer containing urea, thiourea, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), octyl β-D-glucopyranoside, phenylmethylsulfonyl fluoride (PMSF) [7], and protease inhibitors (O'complete Protease Inhibitors Cocktail, Hoffmann-La Roche, Basel, Switzerland). Total sperm protein concentrations were measured using the RC-DC Protein Assay (Bio-Rad Laboratories, Hercules, CA, USA) with a bovine serum albumin (BSA) standard curve. Equal amounts of each sperm protein sample were pooled so that each sample contributed equally with respect to the protein content.

2.2. Sperm protein separation and identification

Ten sperm protein samples (400 μg) were applied to pH 5–8, 11 cm long nonlinear (NL) ReadyStrip™ IPG Strips (Bio-Rad Laboratories) using passive (12 h) and subsequently active (12 h, 50 V) rehydration processes. Numerous biological replicates were used to establish reproducible sperm protein profiles. Isoelectrofocusing (IEF) was performed using a Protean® i12™ IEF Cell (Bio-Rad Laboratories) as follows: (i) 250 V for 125 V-hours (Vh), (ii) 500 V for 250 Vh, (iii) 1000 V for 500 Vh, (iv) linear increase to 5000 V for 1.5 h, and (v) 5000 V for 35,000 Vh. After IEF, the IPG strips were reduced with dithiothreitol (DTT) in an equilibration buffer (6 M urea, 0.5 M Tris/HCl pH 6.8, 2% w/v SDS, 30% w/v glycerol, and 1% w/v DTT) and then alkylated with iodoacetamide (2.5% w/v). The second dimension of electrophoresis was performed in a 12% SDS polyacrylamide gel with a total of 920 Vh using a Protean Plus™ Dodeca Cell™ electrophoretic chamber (Bio-Rad Laboratories). After 2-DE separation, the gels were stained with colloidal Coomassie Brilliant Blue G-250 (CCB G-250) according to Pink et al. [18].

After manual excision from gels, the protein spots were identified using Microflex™ MALDI-TOF MS (Bruker Daltonics, Billerica, MA, USA) as previously described by Ożgo et al. [19]. Briefly, the protein spots were washed once with a 25 mM NH_4HCO_3 in 5% v/v acetonitrile (ACN) solution and washed twice with a solution 25 mM NH_4HCO_3 in 50% v/v ACN solution to remove bound CBB G-250. Decolorized gel pieces were dehydrated using 100% ACN and subsequently vacuum-dried.

Then, the dried protein spots were digested overnight at 37 °C in a trypsin solution (12.5 μg trypsin/mL in 40 mM NH_4HCO_3). The resulting peptides were extracted from the gel pieces with 100% ACN and subsequently loaded onto a MALDI-MSP AnchorChip 600/96 plate (Bruker Daltonics, Bremen, Germany). The peptide mixture was then immediately combined with an equal volume of matrix solution containing 2.5 mg/mL α-cyano-4-hydroxycinnamic acid (CHCA), 0.1% v/v trifluoroacetic acid (TFA), and 50% v/v ACN using the dry droplet technique. External calibration was performed against Peptide Mass Standard II with a mass range of 700–3200 Da (Bruker Daltonics). Mass spectra were acquired in the positive-ion reflector mode using a Microflex MALDI-TOF MS (Bruker Daltonics). The peptide mass fingerprinting (PMF) data were compared with data deposited in mammalian databases (SWISS-PROT: http://us.expasy.org/uniprot/and NCBI: http://www.ncbi.nlm.nih.gov/) using the MASCOT search engine (http://www.matrixscience.com/). The parameters used for the database search included monoisotopic mass, 150 ppm mass accuracy, trypsin as the enzyme with one missed cleavage allowed, carbamidomethylation of cysteine as a fixed modification, and methionine oxidation as a variable modification. The results were further validated by the MASCOT score value and protein sequence coverage.

The first step of our proteomic strategy was the separation of proteins using 2-DE. This technique is popular for separation of complex protein mixtures, however, it may not be effective for analysis of proteins characterized by low pH values (pH < 3) or high pH values (pH > 10). Additionally, 2-DE is not suitable for the separation of very small molecular weight proteins (Mw < 10 kDa) or membrane proteins because they have a tendency to precipitate during IEF. However, an advantage of this technique is that it can be used to study posttranslational modifications that change the pI and/or Mw of proteins.

MALDI-TOF MS also has some limitations. Protein spots with low protein concentrations may not be identified and mass determination of large molecules may be difficult because it may be impossible to obtain spectra for peptides > 600 Da. In our experiments, protein identification was based on PMF. The disadvantage of this technique is that it requires a database containing the protein, which has already been characterized in other species. Then, if a large enough number of accurately measured peptide masses (in the form of peaks) in our sample matched a database sample, the protein was unambiguously identified based on this 'fingerprint' and without sequence information. To achieve a high probability of protein identification, the number of matched peptides, which were measured as sequence coverage, was at least 16% for proteins of high Mw and 20% for proteins of low Mw. If the percentage of coverage was lower, the identification was found to be unreliable. When a protein was identified as a homologue of a protein from another species, this was confirmed by repeating the identification When confirmation was not obtained by repetition, validation of the identification was obtained using a different proteomic technique, such as western immunoblotting.

2.3. Bioinformatic analyses

All of the identified proteins were categorized according to subcellular localization based on data from Uniprot Knowledgebase (UniProtKB; http://www.uniprot.org). Gene Ontology (GO) analyses for biological functions and KEGG pathway analyses were performed using STRING bioinformatic software (http://string-db.org/). If a protein was identified as a homologue of a protein from other species, the protein sequences were compared using Basic Local Alignment Search Tool (BLAST; https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.4. Western immunoblotting

The same sperm protein extracts were analysed by western blot to validate the presence of Annexin VII (ANXA7). Sperm protein samples from 4 fertile men were separated by SDS-PAGE, transferred to polyvinylidene difluoride (PVDF) membranes (Merck Millipore, Billerica, MA, USA), blocked with 5% w/v BSA for 1 h, and incubated at 4° C overnight with anti-Annexin VII antibody [EPR17090] (ab197586, Abcam Cambridge, UK). The membranes were washed with 0.1% v/v Tween-20 in Tris-buffered saline (TBS) and then incubated with the goat anti-rabbit IgG H&L (ab97051, Abcam) horseradish peroxidase-conjugated secondary antibody. The immune-reactive complexes were detected using the Clarity™ Western ECL Substrate (Bio-Rad Laboratories). Anti-GAPDH antibody (ab9485, Abcam) was used as the sample loading control [20]. The images were visualized using a ChemiDoc™ MP System (Bio-Rad Laboratories).

2.5. Immunofluorescence assay

After semen liquefaction, portions of the semen samples were washed with PBS and spread onto glass slides. The sperm samples were fixed with 4% paraformaldehyde (PFA), and fixed spermatozoa were washed with PBS and incubated with 0.5% Triton X-100. Then, they were washed again with PBS and incubated in 1% w/v BSA for 1 h at RT. After washing with PBS, the sperm samples were incubated overnight with anti-Annexin VII antibody (ab197586, Abcam). The samples were then incubated with secondary antibody conjugated with Alexa Fluor® 488 dye (Goat anti-rabbit IgG H&L, ab150077, Abcam) for 30 min at RT. The samples were counterstained with DAPI and