Journal of Proteomics 126 (2015) 296–302

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Beyond single modification: Reanalysis of the acetylproteome of human sperm reveals widespread multiple modifications

Tao Zhou *, Xiyang Xia, Juanjuan Liu, Gaigai Wang, Yueshuai Guo, Xuejiang Guo, Xinru Wang *, Jiahao Sha

State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 210029, PR China

ARTICLE INFO

Article history:
Received 9 March 2015
Received in revised form 17 May 2015
Accepted 19 May 2015
Available online 26 May 2015

Keywords:
Human sperm
Post-translational modification
Reanalysis
Tandem proteomics
Acetylation

ABSTRACT

Sperm is an ideal model for studying post-translational modifications since its transcriptional and translational activities are nearly silent. Thus, sperm functions are mainly regulated at the protein level, especially by means of post-translational modifications. Published proteomic datasets may contain valuable undiscovered information. In this study, we reanalyzed the raw data from previous acetylproteome study on human capacitated sperm to include two additional modifications: phosphorylation and ubiquitination. We successfully identified approximately 500 proteins with multiple types of modifications. Compared with recently developed serial enrichment strategy for multiple modifications, reanalysis of single modification enriched data provides a direct and efficient alternative approach. These results greatly expand our knowledge of protein modifications in human sperm.

1. Introduction

Mammalian sperm is a highly specialized haploid cell type. During spermiogenesis, most of the spermatid cytoplasm is discarded, and the sperm head contains nucleus with highly condensed chromosomes. The transcriptional and translational activities of sperm are nearly silent [1,2]. Thus, it is widely accepted that sperm functions are mainly regulated at the protein level, and post-translational modifications (PTMs) are particularly important.

Due to the complex composition of modifications and the low proportion of modified proteins in a cell, a general sample based proteomics strategy can struggle to elucidate global existence of PTMs effectively. Recently, various modified peptide enrichment approaches have been successfully combined with tandem proteomics for large-scale PTM analysis [3]. For human sperm, lectin enrichment and deglycosylation techniques have been used to identify N-linked glycosylated peptides [4]. The phosphoproteome of capacitated human sperm was characterized using immobilized metal affinity chromatography (IMAC) [5], and the lysine acetylproteome of capacitated human sperm was determined through immunopurification enrichment using an anti-acetyllysine antibody [6].

In most large-scale proteomics studies of PTMs, only one type of functional modification is usually considered, and proteins with multiple modifications are ignored. For example, only acetylated sites are searched in acetylated peptides enriched data. However, lysine acetylated proteins can also be modified by other functional modifications. For example, acyl-CoA synthetase 1 (ACSL1) contains numerous clusters of peptides with both phosphorylation and acetylation sites which may be important for enzymatic function or protein stability [7]. Combinations of modifications have been investigated for single proteins, and one recent study applied serial enrichment of different post-translational modifications (SEPTM) to study phosphorylation, ubiquitination and acetylation separately in human leukemia cells [8].

Increasing evidence supported the view that different modifications on the same or different proteins can interact with each other to regulate protein stability and activity, and the cross-talk between different PTM types falls into two main classes: cooperation and competition. For example, acetylation and phosphorylation of erythroid transcription factor GATA-1 are both required to regulate the degradation of active GATA-1 via the ubiquitin proteasome pathway (UPP) cooperatively [9]. However, acetylation of the cellular tumor antigen p53 may compete with ubiquitination which maintains protein stability and activity since the same lysine sites could be modified [10]. Thus, cross-talk between acetylation, phosphorylation and ubiquitination may be widespread and serve to regulate protein stability and activity through dual mechanisms. Accordingly, it is important to identify the composition of proteins with at least any two types of acetylation, phosphorylation and ubiquitination sites together in order to explore potential interactions between these modifications.

For the raw data from single modification enrichment, Matic et al. suggested to reanalyze the fragmentation spectra to uncover new types of modifications [11]. In the present study, we aimed to identify proteins with multiple modification types by reanalyzing raw data from the acetylproteome of human capacitated sperm [6]. First, we evaluated the results of reanalysis with three variable functional

* Corresponding authors.
E-mail addresses: zhoutao@njmu.edu.cn (T. Zhou), xrwang@njmu.edu.cn (X. Wang).

http://dx.doi.org/10.1016/j.jprot.2015.05.022
1874-3919/

modifications (including phosphorylation and ubiquitination in addition to acetylation). To test whether these additional modifications are co-enriched or randomly identified, we further compared the results of the reanalysis with the general proteome of human capacitated sperm without acetylation enrichment. Based on the combined results of the reanalysis and the general proteome, we annotated the functions of proteins with multiple modifications. Finally, we discussed the potential mechanisms underlying the cross-talk between the three types of PTMs and the relevance in regulation of sperm functions.

2. Materials and methods

2.1. Ethics statement and proteomic analysis

The study of human sperm was approved by the Ethics Committee of Nanjing Medical University. Two datasets were analyzed in the present study: the reanalysis dataset and the general dataset. The raw files of the reanalysis dataset were obtained from our previously published lysine acetylproteome of human capacitated sperm (based on an acetylpeptide enrichment strategy) [6]. Experimental details of the acetylproteome including sample preparation and liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis can be found in the previous article [6]. For comparison and evaluation of the reanalysis dataset, we generated a novel general proteome from human capacitated sperm (without modification enrichment). The sample and procedure were similar to the previous study except affinity enrichment of lysine-acetylated peptides was not applied. In brief, semen samples were obtained from healthy adult male volunteers with normal semen quality assessed using the criteria recommended by the 5th World Health Organization (WHO) manual [12]. A Percoll gradient (GE Healthcare, Waukesha, WI, USA) was diluted to 60% with 1× Biggers–Whitten–Whittingham (BWW) medium (114.00 mM NaCl, 4.78 mM KCl, 1.71 mM $CaCl_2$, 1.19 mM $MgSO_4$, 1.19 mM KH_2PO_4, 21.58 mM sodium lactate, 5.56 mM glucose, 10.00 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) and 25.07 mM $NaHCO_3$, pH 7.6). The liquefied semen was gently stratified on the top of the 60% Percoll gradient and centrifuged at 350 ×*g* for 10 min to obtain purified sperm which was washed in 1× BWW medium twice, and then incubated to capacitation in human tubal fluid (HTF; In Vitro Care, Frederick, MD, USA) containing 10 mg/ml BSA for 2.5 h in a humidified CO_2 incubator at 37 °C under 5% CO_2 /95% air (v/v) as described previously [13].

For proteomic analysis, in vitro capacitated sperm samples were dissolved in 8 M urea, 75 mM NaCl, 50 mM Tris, pH 8.2 and 1% (v/v) protease inhibitor cocktail. Sperm proteins were extracted using an ultrasonic method and digested using trypsin (Sequencing Grade Modified Trypsin, Promega, WI, USA). The peptide mixture was separated using a 1 mm ID × 10 cm ion exchange column packed with Poros 10 S (DIONEX, Sunnyvale, CA) on an UltiMate 3000 HPLC system at a flow rate of 50 μl/min. Peptide analysis was performed using an LTQ Orbitrap (ThermoFinnigan, San Jose, CA, USA) coupled directly to the LC column. An MS survey scan was obtained in the mass-to-charge ratio (m/z) range 350–1800, and MS/MS spectra were acquired from the survey scan for the 20 most intense ions (as determined by Xcalibur mass spectrometer software in real time). Dynamic mass exclusion windows of 60 s were used, and siloxane (m/z = 445.120025) was used as a lock mass.

2.2. Database searches

Raw files were processed using the MaxQuant software package (version 1.3.0.5) integrated with the Andromeda search engine [14]. The reanalysis and the general datasets were searched separately with the same parameters. The database search was performed against UniProt (Universal Protein Resource) human protein sequences (updated May 2012) [15], combined with the standard MaxQuant contaminants database. The false discovery rate (FDR) of the identification was estimated by searching against the databases with the reversed amino acid sequences. The site, peptide and protein FDR were all set to 0.01. Enzyme specificity was set to be fully cleaved by trypsin (Trypsin/P), the maximum missed cleavage sites permitted was two. The minimum peptide length required was six. The mass tolerance for precursor ions was set to 20 ppm at the first search as applied in MaxQuant for initial mass recalibration. For the main search, the mass tolerance for precursor ions was set to 6 ppm and for fragment ions was set to 0.5 Da. MaxQuant includes three default modifications which are generally applied to improve peptide identification due to deliberate chemical protection or protein features. Thus, Carbamidomethyl (C) was set as a fixed modification. Variable modifications were Oxidation (M) and N-term Acetylation (Protein N-term). The maximum number of modifications per peptide was five. At least one unique peptide was required for protein identification.

The above parameters were the same as those used in the previous study on the acetylproteome, however, only acetylation on lysine (K) was considered previously. The present study searched for two additional functional modifications: ubiquitination on lysine and phosphorylation on serine (S), threonine (T) or tyrosine (Y). From the initial results of the reanalyzed acetylproteome, we found that although the identification of acetylation was robust, many of the additional PTM sites were ambiguous. Thus, to obtain reliable identification results, we applied strict criteria for identifying multiple modifications which further need a maximal posterior error probability (PEP) of 0.05 for a peptide spectrum match, and a minimal localization probability of 0.75 for site localization. Annotated MS/MS spectra for identified sites from both reanalysis and general datasets are shown in Supporting data 1. Detailed information including modified peptide, localization probability, PEP, match score, mass error, corresponding protein identity and gene name for each identified site are listed in Supporting data 2.

2.3. Statistical and bioinformatics analysis

For statistical comparison of different ratios of modified peptides among the identified spectra, the Fisher's exact test was performed and a P value less than 0.05 was considered statistically significant. To visually compare the sequence features between acetylation, phosphorylation and ubiquitination, we created sequence logos using the aligned modified window sequences (six amino acids from the modified site) using the WebLogo web server (version 3.4) [16]. The color scheme of the amino acids was based on their chemical features. Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were used to determine the functional class and regulatory network of a group of proteins [17]. To investigate potential functions of proteins with multiple modifications, we performed enrichment analysis based on GO and KEGG knowledgebase using WebGestalt web server [18]. The human genome was set as the background and an adjusted P value by Benjamini & Hochberg (BH) method less than 0.05 was controlled for significant enrichment. The structure graph of protein domain and sites was plotted using Domain Graph (DOG; Version 1.0) [19]. Protein–protein interaction (PPI) information was obtained from the Biological General Repository for Interaction Datasets (BioGRID) database (version 3.2.119) [20]. Cytoscape (Version: 3.01) was applied to generate the visual network of PPIs between proteins with multiple PTMs [21].

3. Results and discussion

3.1. Reanalysis of the acetylproteome with additional modifications

Generally, only designated modifications are considered in database searches for a specific experiment. For example, only acetylation of lysine residues is considered for samples of enriched acetylpeptides. Searching for additional modifications greatly increases the search time and reduces the specificity of identification. In this study, we