POSTPRINT VERSION CORRESPONDING TO: "*Characterization of Human Sperm Soler-Ventura A, Gay M, Jodar M, Vilanova M, Castillo J, Arauz-Garofalo G, Villarreal L, Ballescà JL, Vilaseca M, Oliva R. Protamine Proteoforms through a Combination of Top-Down and Bottom-Up Mass Spectrometry Approaches.J Proteome Res. 2020 Jan 3;19(1):221-237. doi: 10.1021/acs.jproteome.9b00499*."

Title: Characterization of human sperm protamine proteoforms through a combination of top-down and bottom-up mass spectrometry approaches

Authorship:

Ada Soler-Ventura[†‡#], Marina Gay[†ǁ#], Meritxell Jodar[†‡#], Mar Vilanova[†ǁ], Judit Castillo[†‡], Gianluca Arauz-Garofalo[†ǁ], Laura Villarreal[†ǁ], Josep Lluís Ballescà[†¥], Marta Vilaseca[†ǁ], Rafael Oliva[†‡§*]

[†] EUGIN-UB Research Excellence Program.

[‡] Molecular Biology of Reproduction and Development Research Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Fundació Clínic per a la Recerca Biomèdica (FCRB), Faculty of Medicine and Health Sciences, University of Barcelona, Barcelona 08036, Spain,

[§] Biochemistry and Molecular Genetics Service, Hospital Clínic, Barcelona 08036, Spain.

[ǁ] Mass Spectrometry and Proteomics Core Facility, Institute for Research in Biomedicine (IRB Barcelona), BIST (Barcelona Institute of Science and Technology), Barcelona 08028, Spain.

[¥ǁ] Clinic Institute of Gynecology, Obstetrics and Neonatology (ICGON), Hospital Clínic, Barcelona, Spain.

[#] These authors have contributed equally to this work

***Corresponding address:** Molecular Biology of Reproduction and Development Research Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Fundació Clínic per a la Recerca Biomèdica, Faculty of Medicine and Health Sciences,

University of Barcelona, Casanova 143, Barcelona 08036, Spain. Telephone: +34 93 402 18 77; Fax: +34 93 403 52 78; e-mail: roliva@ub.edu

Graphical Abstract /Table of Contents

Abstract: 200 words

Protamine 1 (P1) and protamine 2 family (P2) are the most abundant nuclear basic sperm-specific proteins, packing 85-95% of the paternal human genome. P1 is synthesized as a mature form, whereas P2 components (HP2, HP3 and HP4) arise from the proteolysis of the precursor (pre-P2). Due to the particular protamine physical-chemical properties, its identification by standardized bottom-up mass spectrometry (MS) strategies based on trypsin-digestion are not straightforward. Therefore, the aim of this study was to identify the sperm protamine proteoforms profile including P1 and P2 members and their post-translational modifications in normozoospermic individuals using two complementary strategies, a top-down MS approach and a proteinase K-digestion based bottom-up MS approach. For P1 and pre-P2, previously described and new truncated proteoforms were identified by the top-down MS approach. Likewise, intact naïve P1, pre-P2, and P2 mature forms and their phosphorylation pattern, as well as a +61 Da modification in different proteoforms were detected. Additionally, the bottom-up MS approach revealed other phosphorylated residues for pre-P2 and the P2 isoform 2, previously classified as a missing protein. Implementing these strategies in comparative studies of different infertile phenotypes would permit to determine alterations in the protamine proteoforms pattern, especially phosphorylation/dephosphorylation dynamics and elucidate its potential role in male fertility.

Keywords: 10 keywords

Top-down proteomics; bottom-up proteomics; electron transfer dissociation, protamines; sperm; post-translational modifications; phosphorylation; missing proteins; male infertility.

Introduction

During the final phase of spermatogenesis, called spermiogenesis, human sperm chromatin undergoes an intricate remodeling, in which histones are sequentially replaced by specific histone variants, transition proteins and, finally, by protamines (1). This process results in a unique chromatin structure with a 85-95 % of sperm DNA packed by protamines and a remaining 5-15 % of the paternal DNA attached to histones (1)–(5). Protamines are small and extremely basic sperm-specific proteins rich in positively-charged arginine residues (50-70 %). This particular amino acid composition promotes the formation of highly-condensed toroidal complexes alongside the negatively-charged sperm DNA. In addition, protamines harbor cysteine residues able to form disulfide bonds among intra- and inter-protamine molecules that stabilize the nucleoprotamine structure (1),(6). It has also been postulated that free thiol groups could bind to metals such as zinc, allowing the formation of salt bridges, which could further have a role in sperm chromatin dynamics (7). Alltogether, the protamine properties result in a highly compact chromatin structure that protects sperm DNA from nucleases and contributes in obtaining the required hydrodynamic shape for mature sperm functionality (1),(8)–(11).

In humans, protamines are encoded by a single copy of protamine 1 (*PRM1*) and protamine 2 (*PRM2*) genes clustered together with transition nuclear protein 2 (*TNP2)* gene in chromosome 16p13.3 (12),(13). Of note, although both protamine genes are present in all mammals, *PRM2* is only translated in some species, including human and mice (1),(14)–(16). Two types of protamines have been described in the human at the protein level, the protamine 1 (P1) and the protamine 2 family (P2). Interestingly, while P1 is directly synthesized as a mature protein of 51 amino acid residues, P2 derives from a precursor (pre-P2) of 102 amino acid residues (17). By proteolysis, pre-P2 gives rise to three P2 mature proteoforms (HP2, HP3, and HP4), comprising residues 46-102 (HP2), 49-102 (HP3), and 45-102 (HP4), differing among them only on the 1-4 amino

acid residues from the N-terminus end. HP2 is the most abundant proteoform followed by HP3, and the minority proteoform is HP4 (1),(15),(18)–(20). Other studies have also identified pre-P2 intermediate processed forms namely HPI1, HPI2, HPS1, and HPS2, which could give rise to mature HP2, HP3, and HP4 (21).

The normal ratio between the amounts of P1 and P2 members was established around 1 (0.8-1.2) through acid-urea polyacrylamide gel electrophoresis (PAGE)-based experiments. Alterations in this P1 / P2 ratio have been connected to male infertility, sperm DNA damage and/or impairments in fertilization and embryo development (reviewed elsewhere (22),(23)). Nevertheless, Nanassy and colleagues established in a fertile population a normal P1 / P2 ratio ranging 0.54 – 1.43 (24) and the meta-analysis performed by Ni and colleagues concluded that P1 / P2 ratio deregulations seemed to be associated only with subfertility (25). The presence of different protamine proteoforms including truncated and/or post-translational modifications (PTMs) could explain the controversial results about the associations between protamine content and fertility outcomes, since a deregulation of the proteoform pattern that could not be detected by standard methods might result in male infertility. Despite the fact that only a small part of sperm DNA remain packed by histones (1),(3), sperm histone variants and their corresponding PTMs have been widely studied and their role as epigenetic marks beyond fertilization has been postulated (4),(26)–(33). In contrast, the knowledge about protamine proteoforms and their associated potential regulatory activities is scarce.

It is known that during spermatogenesis, P1 is extensively phosphorylated by SRSF Protein Kinase 1 (SRPK1), while P2 is phosphorylated by Calcium / Calmodulin Dependent Protein Kinase IV (CAMK4) after its proteolytic cleavage (8),(27),(34)–(37). Noteworthy, CAMK4 knockout mice are infertile showing spermatogenesis defects due to the specific loss of P2 and TNP2, suggesting that protamine PTMs, such as P2 phosphorylation, could be crucial for male fertility (37). Once P1 and P2 mature proteoforms are tightly bound to the DNA, an extended dephosphorylation occurs before

spermatozoa enter the epididymis (27),(38). In fact, recently, Itoh and colleagues described in an infertile Hspa4l-/- null mice a decreased localization of phosphatase Ppp1cc2 in the chromatin fraction, which led to an abnormal increase of the *Prm2* phosphorylation state in epididymal sperm (39). The substitution of *Prm2* phosphorylatable residue S56 (S59 in humans) to a non-phosphorylatable residue A56 in this infertile model rescued the fertility state and mice were fertile as wild-type animals (39). These results suggest the importance of a regulated balance of phosphorylation/dephosphorylation dynamics in sperm protamines and their crucial role in fertility. However, few studies based on phosphoserine conversion followed by protein sequencing, electrospray mass spectrometry and mass spectrometry (MS) have provided evidence of remaining protamine phosphorylated residues in human mature sperm (40)–(42), which seems to aid the DNA proper binding (8),(27) and might have a role beyond oocyte fertilization during maternal histone replacement (7).

In order to perform a comprehensive proteoform characterization of protamines in human sperm, the use of high-throughput proteomic techniques seems to be warranted. However, due to the particular physical-chemical properties of protamines (extreme basicity, amino acid sequence rich in arginine and cysteine residues, small size and high-similarity among P2 mature proteoforms (1)), the application of standardized bottom-up MS strategies based on trypsin digestion of arginine and lysine residues is not straightforward. For that reason, technical modifications such as the use of other proteases should be optimized in order to overcome this limitation. Alternatively, a top-down MS strategy could direct identify protamine proteoforms thanks to their small size (43). It has been described that a combination of top-down MS and bottom-up MS strategies is optimal to describe PTM patterns in proteins with similar characteristics to protamines, as for example histones (44). Remarkably, recent studies have applied these MS variations to assess protamine PTMs in mouse and human sperm (42),(45),

but all these studies were performed in pooled samples, which do not provide information about inter-individual variability.

Therefore, the aim of this study is to set up the methodology to identify P1 and P2 family members' proteoforms and their associated PTMs in independent normozoospermic samples in order to establish the normal protamine proteoforms profile. Two complementary approaches have been used to accomplish our objective: 1) a top-down MS approach with novel proteomic workflows and data analysis pipelines, and 2) a bottom-up MS approach based on proteinase K-digestion of protamines, rather than trypsin digestion. The implementation of these complementary techniques in future comparative studies including different subtypes of infertile patients could resolve the impact of dysregulations in the protamine proteoforms profile and the importance of its balance on male fertility.

Experimental Procedures

Sample collection

Human semen samples (n=7) were obtained from patients undergoing routine semen analysis at the Assisted Reproduction Unit from the Clinic Institute of Gynecology, Obstetrics and Neonatology, at the Hospital Clínic of Barcelona (Spain). All patients gave signed informed consent in accordance with the Declaration of Helsinki. The ejaculates were collected by masturbation into sterile containers after 3-5 days of sexual abstinence. The evaluation of the seminal parameters was performed using the automatic semen analysis system CASA (Proiser, Paterna, Spain). After seminal analysis, all samples were classified as normozoospermic according to World of Health Organization guidelines (46). Three semen samples were pooled for the initial bottom-up MS experiments and four semen samples were used as individual samples for the top-down approach (2 samples), the bottom-up MS strategy (1 sample) and the RNA experiments (1 sample). Sperm from individual samples was purified through 50 %

Puresperm® density gradient separation (NidaCon International AB, Gothenburg, Sweden), following the manufacturer's instructions.

Human sperm protamine purification

Protamine enriched fraction from spermatozoa was purified as previously described in Soler-Ventura et al. 2018 (23) except that histones and other basic proteins weakly associated to DNA were removed from the sperm by 0.5 M HCl extraction prior to disulfide bond reduction. This was accomplished by adding 0.5 M HCl to the purified sperm cells, followed by vortexing and incubation for 10 min at 37 ºC and centrifuged at 2000 g for 20 min at 4 ºC (three times) prior to suspending the sperm pellet with 0.5 % Triton X-100, 20 mM (pH8) Trizma® base HCl, and 2 mM $MgCl_2$. Then, samples were processed as described in Soler-Ventura et al. 2018 (23). For the top-down MS approach, protamine-enriched fraction equivalent to 15 million spermatozoa was dried at room temperature and kept at -20 ºC until the MS/MS procedure. For the bottom-up approach, the protamine-enriched fraction was dried at room temperature and suspended with 20 µl of a solution containing 50 mM Tris-HCl pH 7.5, 5 mM $CaCl_2$. Subsequently, protamines were digested with proteinase K (Promega, Madison, Wisconsin, US) during 4 h at 37 ºC and the reaction was stopped with 5 mM PMSF. For both the top-down and the bottom-up MS approaches, successful protamine purification was monitored in an acid-urea PAGE (23).

Protamine quantification

Purified extracts of protamines were quantified as described in Soler-Ventura et al. 2018 (23). Briefly, an aliquot of protamines for each sample was separated in an acid-urea PAGE together with increasing amounts of a protamine standard obtained as described elsewhere (47). P1 and P2 approximate amounts were calculated from the protamine optical densities measurements and the quantification obtained with the protamine

standard through Quantity One 1-D Analysis Software (Bio-rad, Hercules, California, US).

Protamine identification by mass spectrometry

The overall procedure used to characterize protamine PTMs is schematized in figure 1. For the top-down MS approach, we employed a total of 2 biological replicates, from those, 6 experimental replicates and 10 technical replicates were analyzed and the data was integrated. After the LC, we conducted the MS analysis, in which the deconvoluted MS1 permitted assign PTMs to the intact naïve protamines according to the difference between the exact mass of the modification and the intact proteoform, and MS2 fragmentation allowed the localization of modified residues (Figure 2). Finally, we employed three different engines to characterize protamine proteoforms and their associated PTMs (Supporting Figure 1). For the bottom-up MS approach, an individual sample and the pooled samples were treated as biological replicates, and the data was integrated. After proteinase K digestion, protamine peptides were subjected to LC-MS/MS and the MS1&MS2 analysis conducted identified protamine peptides and their associated PTMs (Figure 1).

Mass Spectrometry analysis

The MS proteomics data have been deposited to the PRIDE Archive (XXXXXX) via the PRIDE partner repository with the data set identifier PXDxxxx and 10.6019/PXDxxxx.

Confirmation of P2 isoform 2 at the RNA level.

Purified spermatozoa from an ejaculate were prepared and stored at -80 ºC in RLT buffer (Qiagen, Hilden, Germany) and 1.5 % β-mercaptoethanol until further processing. Sperm long RNA fraction was extracted following the RNeasy Mini Kit protocol (Qiagen) with some modifications previously established in Goodrich et al. 2013 and Jodar et al. 2015 (48),(49). High-quality sperm RNA free of DNA and somatic RNA contamination (48),(49)

was employed for further analysis (Primers used for quality controls are shown in Supporting Table 1). To validate the presence of P2 isoform 2 at RNA level, 50 ng of sperm RNA was reverse transcribed using Maxima RT (Thermo Fisher Scientific) and random primers (Applied Biosystems), followed by a real-time PCR using PowerUp™ SYBR™ Green PCR (Applied Biosystems) targeting an specific region of the P2 isoform 2 transcript (Supporting Table 1 and Supporting Figure 2). PCR products were visualized in a 3 % agarose gel at 90 V. After that, 2 bands were obtained, one showing the P2 isoform 2 predicted size and another one corresponding to a product shorter in length (Supporting Figure 2). Both bands were excised and eluted using the QIAEX II Gel Extraction Kit (Invitrogen) according to the manufacturer's instructions. Whereas the new product was directly analyzed by direct sequencing (Supporting Figure 2), P2 isoform 2 required a semi-nested PCR using PowerUp™ SYBR™ Green real-time PCR (Thermo Fisher Scientific) with specific primers (Supporting Table 1) before direct sequencing (Supporting Figure 2).

Results and Discussion

Top-down and bottom-up mass spectrometry approaches technical issues

The LC for the top-down MS strategy has been able to separate in different elution times P2 from P1 (Figure 2). Subsequent MS analysis has identified up to 5.263 counts, comprising 32 different proteoforms. Specifically, we identified intact naïve P1, pre-P2 and P2 mature forms; previously described truncated proteoforms for pre-P2 and two novel truncated proteoforms for both P1 and pre-P2; and different combinations of PTMs for P1, pre-P2, HP2 and HP3 (Table 1; Figure 3; Supporting Figure 3). Additionally, we have detected a total of 47 peptide spectrum matches (PSMs) comprising 35 different peptides for the bottom-up MS workflow (Table 2; Figure 4; Supporting Figure 1). Particularly, those peptides correspond to the intact naïve and mono- or di-phosphorylated pre-P2 and the naïve or monophosphorylated P2 isoform 2. It is interesting to note that P1 and P2 mature forms were not identified by bottom-up MS

approach most probably due to the more distinct physical-chemical properties of P1 and the high similarity among the P2 mature components.

Novel protamine proteoforms

Using the top-down MS approach we have fully identified canonical protamines, namely P1 (P04553), pre-P2 (P04554) and P2 mature forms HP2, HP3 and HP4 (Table 1; Figure 3). Moreover, we have identified pre-P2 truncated proteoforms, specifically spanning residues 22-102 (HPI2), 34-102 (HPS1) and 37-102 (HPS2), all of them previously described (50)–(52). Interestingly, we detected two novel truncated proteoforms, P1 truncated proteoform spanning residues 8-51 and pre-P2 truncated proteoform spanning residues 8-102 (Table 1; Figure 3; Supporting Figure 4). Although pre-P2 has been proposed as highly proteolytically processed in a regulated manner (53),(54), the mechanisms and the specific team-players implicated in this processing have not yet been described (21). In accordance with pre-P2 processing, the same post-translational cleavage could proteolyze the truncated form of P1. In both cases, the functions and roles of these truncated forms remain to be clarified.

With the bottom-up MS approach, we were able to identify peptides corresponding to pre-P2 and P2 isoform 2 (P04454-2), which has been classified as a missing protein and it was only validated at the RNA level according to the UniProtKB database (https://www.uniprot.org/). It is important to highlight that a missing protein should be identified by two or more unique peptides with > 9 amino acid residues and an FDR < 1 % to be considered experimentally validated at the protein level (55). In our study, we have identified with clearly spectra in both the pool of three samples and in the individual sample (Table 2; Figure 4) two unique peptides of P2 isoform 2 with 9 and 10 amino acid residues, respectively, after proteinase K digestion (Figure 4A). Although Vanderbrouck and colleagues reported in 2016 one unique peptide of P2 isoform 2, the strict criteria required to fully characterize a missing protein, it was not accomplished (56). Therefore,

the findings showed in the present study confirm for the first time the presence of this new P2 isoform 2 in human sperm at the protein level.

P2 isoform 2 arise from the splice site variant ENST00000435245.2 and results in a 140 amino acid length protein differing from pre-P2 from residue 91 (57),(58). The presence of P2 isoform 2 was also validated at the RNA level in an independent sperm sample. Surprisingly, two different products were amplified using a specific set of primers for P2 isoform 2, (Figure 5). Direct sequencing of each product revealed two splice variants of P2, the already known P2 isoform 2 and a novel splice variant that we named it P2 isoform 3 (Figure 5 and Supporting figure 2). However, P2 isoform 3 was not detected at the protein level using neither top-down nor bottom-up MS approaches, suggesting that either this product is not translated or the employed techniques have been unable to identify it at the protein level. Looking in detail the intron nucleotide sequence of *PRM2,* the splice sites corresponding to P2 isoform 2 and isoform 3 are solely conserved among different higher primate species, but not in other mammals such as *Mus Musculus* and *Rattus Norvegicus*, suggesting that there is no specific evolutionary conservation in mammals (Supporting Figure 6). However, the possible function of P2 isoform 2 and the putative translation of P2 isoform 3 are unrevealed yet, being necessary further studies to elucidate the potential role of these novel protamine isoforms.

Protamines contain multiple post-translational modifications

A phosphorylated pattern has been established for P1 through the top-down MS approach, MS1 identified, in decrease count number, naïve P1 and P1 mono-, di-, and tri-phosphorylated (Table 1; Figure 3A; Supporting figure 3). The automated validation of phosphorylated residues is summarized in supporting figure 5 and to ascertain and confirm modified residues, MS2 deconvoluted fragmentation spectra were manually validated. Following this pipeline, MS2 analysis was able to confirm P1 mono-phosphorylated at S11 or S13 and di-phosphorylated at S11-S22 or S11-S29 (Figure 3B) but MS2 could not detect tri-phosphorylated specific amino acid residues. Moreover, we

identified a phosphorylation in S11 in the truncated P1 proteoform (Figure 3B; Supporting Figure 4). Additionally, pre-P2 was detected in its native form as well as containing several PTMs. Specifically, MS2 deconvolution showed a phosphorylation pattern for pre-P2 including mono-phosphorylation at residue S8, di-phosphorylation at residues S8 and S10, tri-phosphorylation at residues S8, S10 and S59 and tetra-phosphorylation at residues S8, S10, S13 and S59 (Figure 3B). For P2 mature forms, S59 was detected phosphorylated for both HP2 and HP3, but no PTMs were identified in HP4 (Figure 3B). Total counts confirmed that naïve HP2 was the most identified mature P2 form followed by HP2 mono-phosphorylated, naïve HP3, HP3 mono-phosphorylated and finally naïve HP4. From the truncated P2 proteoforms, only HPS1 was detected harboring modified residues (Table 1; Figure 3). In particular, HPS1 proteoforms included mono-phosphorylation at S59 and the modification of N-terminus Q34 to pyroglutamic acid (Table 1; Figure 3B). The PTM of N-terminus glutamine residue modified to pyroglutamic acid is a reaction commonly found in plants and animals, including humans, and it has been proposed to be involved in stabilizing proteins and peptides structures (59). Complementary to these results, the bottom-up MS approach allowed the identification modifications in the sequence of pre-P2. Specifically, mono-phosphorylated peptides at S37 or Y44 residues corresponding to the pre-P2 were detected. In addition, P2 isoform 2 was found phosphorylated at T124 residue (Table 2; Figure 4A).

Our results are supported by previous publications describing retained phosphorylated residues in mature sperm through phosphoserine conversion followed by protein sequencing, electrospray mass spectrometry or by similarity from other species. For instance, P1 has been described by others as phosphorylated in S9, S11 and S29 residues and pre-P2 in S8, S10, S37, Y47, S50, S59 and S73 (41),(60)–(62). A phosphorylation pattern for P1, including mono- di- and tri- phosphorylation and methylations and acetylations in HP3 were described through a top-down MS approach in human sperm but they were unable to specifically localize the modified residues (42). In mice, Brunner *et al.* 2016 described different combinations of protamine PTMs using

a combination of top-down and bottom-up MS approaches. Specifically, they identified phosphorylated, methylated and acetylated residues for both protamines (45).

The conventional analysis of the P1 / P2 ratio based on their quantity derived from urea-PAGE experiments is not suitable to identify alterations in the protamine proteoform profile and proteomic strategies are adequate to detect them. Recently, it has been described the crucial role of P2 dephosphorylation in an infertile mouse model to rescue the fertility state (39), pointing out the existence of a tightly regulated balance of phosphorylation/dephosphorylation state, which could compromise fertility. Impairment in this dynamic network could be the underlying cause of different types of infertile patients, being necessary to be addressed in future protamine proteoforms comparative studies to elucidate its role. Likewise, our results show that naïve proteoforms are more abundant than proteoforms with phosphorylated residues, supporting the hypothesis that a high dephosphorylation wave occurs during sperm maturation (1),(8),(27). Remaining phosphorylated residues in mature human sperm could be simple spermatogenesis remnants find randomly across some protamine molecules or in contrast could act as specific epigenetic marks during early embryo development. Since the attachment of phosphorylated protamines to the DNA is not stronger as for naïve protamines, due to the reduction of positive charge, the phosphorylated protamines could mark the preferential paternal protamine replacement by maternal histones during early embryo development (7). However, the function of these retained phosphorylated protamines beyond fertilization, if any, remains to be elucidated.

Surprisingly, through the top-down MS approach a +61 Da modification has been widely identified solely or in combination with other PTMs across P1, HP2, HP3 and HPS1 (Table 1; Figure 3A; Supporting Figure 3). In particular, MS2 fragmentation showed the localization of this modification in residues C30 and/or C51 of P1 as mono-modified or in conjunction with mono-, di- and tri-phosphorylation, and in residue C58 of HP2, HP3 and HPS1 (Figure 3B). In HPS1, this modification was detected solely or in combination with Q34 modification to pyroglutamic acid (Figure 3B). The incorporation of 61 Da to the

molecular weight of protamines might correspond to the presence of Zn^{2+} ion attached to protamine molecules. It has been demonstrated that low levels of Zn^{2+} in seminal plasma negatively correlates with sperm quality and fertility (63),(64). Moreover, it has also been proposed that Zn^{2+} could also aid to stabilize sperm chromatin (7). The attachment of Zn^{2+} to P2 has been previously demonstrated (65), and a stoichiometric binding of one P2 molecule to one Zn^{2+} ion has been proposed (66). In addition, it has been described as a zinc-finger motif in P2 (Cys2-His2) that could allow P2 proteins to act as zinc-finger proteins (67). In the present study, we have also been able to detect the +61 Da modification in P1, although the attachment of Zn^{2+} to P1 has not been previously described. Further studies are required in order to clarify if this modification in both P1 and P2 corresponds to Zn^{2+} ion and the potentially related function.

Conclusions

In this study, we have demonstrated that the combination of top-down and bottom-up MS approaches is suitable to study the protein sequence and the associated proteoforms of sperm-specific protamines, giving complementary information. These approaches have permitted, for the first time, to establish the normal phosphorylation pattern for human protamines, identifying known and novel phosphorylated residues for P1, pre-P2, HP2, and HP3. Remarkably, our top-down MS results suggest that the normal protamine proteoform profile comprises naïve P1 as the most widely identified proteoform, followed descending by mono- and di-phosphorylated P1, naïve mature HP2, mono-phosphorylated HP2, naïve HP3, mono-phosphorylated HP3 and finally naïve HP4. Moreover, in both P1 and pre-P2 previously detected and novel truncated proteoforms have been identified, possibly because of post-translational proteolytic cleavage, some of them containing modified residues. Interestingly, in addition to phosphorylation, both protamines harbor a +61 Da modification that might correspond to Zn^{2+} ion. Furthermore, the proteinase K-based bottom-up MS approach unequivocally identified the P2 isoform 2, considered until now as a missing protein. Since the utility of conventional P1 / P2

ratio based on their quantity is limited and cannot identify alterations in the protamine proteoform profile, our results have established the normal profile of protamine proteoforms in human individual sperm samples. The study of proteoform dysregulations in different types of patients, particularly phosphorylation/dephosphorylation dynamics open a window to determine their potential role on male fertility, along with its potential function as epigenetic mark beyond fertilization.

Acknowledgements

This work was supported by grants from EUGIN-UB Research Excellence Program (EU -REP 2014), the Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad; Fondos FEDER 'Una manera de hacer Europa', PI 13 /00699, and PI 16 /00346), from Fundación Salud 2000 (SERONO 13 –015) to R.O. M.J. is supported by a Postdoctoral Fellowship from the Government of Catalonia (Generalitat de Catalunya, pla estratègic de recerca i innovació en salut, PERIS 2016 -2020, SLT SLT002 /16 /00337). J.C. is supported by the Sara Borrell Postdoctoral Fellowship from the Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad, Acción Estratégica en Salud, CD17/00109).

Bibliography

(1) Oliva, R. Protamines and Male Infertility. *Hum. Reprod. Update* **2006**, *12* (4), 417–435.

(2) Kimmins, S.; Sassone-Corsi, P. Chromatin Remodelling and Epigenetic Features of Germ Cells. *Nature* **2005**, *434* (7033), 583–589.

(3) Balhorn, R. The Protamine Family of Sperm Nuclear Proteins. *Genome Biol.* **2007**, *8* (9), 227.

(4) Bao, J.; Bedford, M. T. Epigenetic Regulation of the Histone-to-Protamine Transition during Spermiogenesis. *Reproduction* **2016**, *151* (5), R55–R70.

(5) Barral, S.; Morozumi, Y.; Tanaka, H.; Montellier, E.; Govin, J.; de Dieuleveult, M.; Charbonnier, G.; Couté, Y.; Puthier, D.; Buchou, T.; et al. Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. *Mol. Cell* **2017**, *66* (1), 89–101.e8.

(6) Balhorn, R.; Corzett, M.; Mazrimas, J. A. Formation of Intraprotamine Disulfides in Vitro. *Arch. Biochem. Biophys.* **1992**, *296* (2), 384–393.

(7) Björndahl, L.; Kvist, U. Human Sperm Chromatin Stabilization: A Proposed Model Including Zinc Bridges. *Mol. Hum. Reprod.* **2010**, *16* (1), 23–29.

(8) Oliva, R.; Dixon, G. H. Vertebrate Protamine Genes and the Histone-to-Protamine Replacement Reaction. *Prog. Nucleic Acid Res. Mol. Biol.* **1991**, *40* (C), 25–94.

(9) Oliva, R.; Castillo, J. Proteomics and the Genetics of Sperm Chromatin Condensation. *Asian J. Androl.* **2011**, *13* (1), 24–30.

(10) Oliva, R.; Bazett-Jones, D.; Mezquita, C.; Dixon, G. H. Factors Affecting Nucleosome Disassembly by Protamines in Vitro. Histone Hyperacetylation and Chromatin Structure, Time Dependence, and the Size of the Sperm Nuclear Proteins. *J. Biol. Chem.* **1987**, *262* (35), 17016–17025.

(11) Oliva, R.; Mezquita, C. Marked Differences in the Ability of Distinct Protamines To Disassemble Nucleosomal Core Particles in Vitro. *Biochemistry* **1986**, *25* (21), 6508–6511.

(12) Domenjoud, L.; Kremling, H.; Burfeind, P.; Maier, W. M.; Engel, W. On the Expression of Protamine Genes in the Testis of Man and Other Mammals. *Andrologia* **1991**, *23* (5), 333–337.

(13) Nelson, J. E.; Krawetz, S. A. Mapping the Clonally Unstable Recombinogenic PRM1→PRM2→TNP2 Region of Human 16p 13.2. *Mitochondrial DNA* **1995**, *5*

(3), 163–168.

(14) McKay, D. J.; Renaux, B. S.; Dixon, G. H. The Amino Acid Sequence of Human Sperm Protamine P1. *Biosci. Rep.* **1985**, *5* (5), 383–391.

(15) McKay, D. J.; Renaux, B. S.; Dixon, G. H. Human Sperm Protamines: Amino-acid Sequences of Two Forms of Protamine P2. *Eur. J. Biochem.* **1986**, *156* (1), 5–8.

(16) De Yebra, L.; Ballesca, J. L.; Vanrell, J. A.; Bassas, L.; Oliva, R. Complete Selective Absence of Protamine P2 in Humans. *J. Biol. Chem.* **1993**, *268* (14), 10553–10557.

(17) Yelick, P.; Balhorn, R.; Johnson, P.; Corzett, M.; Mazrimas, J.; Kleene, K.; Hecht, N. Mouse Protamine 2 Is Synthesized as a Precursor Whereas Mouse Protamine 1 Is Not. *Mol. Cell. Biol.* **1987**, *7* (6), 2173.

(18) Jodar, M.; Oliva, R. Protamine Alterations in Human Spermatozoa. *Adv. Exp. Med. Biol.* **2014**, *791*, 83–102.

(19) Arkhis, A.; Martinage, A.; Sautiere, P.; Chevaillier, P. Molecular Structure of Human Protamine P4 (HP4), a Minor Basic Protein of Human Sperm Nuclei. *Eur. J. Biochem.* **1991**, *200* (2), 387–392.

(20) Balhorn, R.; Corzett, M.; Mazrimas, J.; Stanker, L. H.; Wyrobek, A. High-Performance Liquid Chromatographic Separation and Partial Characterization of Human Protamines 1, 2, and 3. *Biotechnol. Appl. Biochem.* **1987**, *9* (1), 82–88.

(21) de Mateo, S.; Ramos, L.; de Boer, P.; Meistrich, M.; Oliva, R. Protamine 2 Precursors and Processing. *Protein Pept. Lett.* **2011**, *18* (8), 778–785.

(22) Barrachina, F.; Soler-Ventura, A.; Oliva, R.; Jodar, M. Sperm Nucleoproteins (Histones and Protamines). In *A Clinician's Guide to Sperm DNA and Chromatin Damage*; Springer International Publishing: Cham, 2018; pp 31–51.

(23) Soler-Ventura, A.; Castillo, J.; de la Iglesia, A.; Jodar, M.; Barrachina, F.; Ballesca, J. L.; Oliva, R. Mammalian Sperm Protamine Extraction and Analysis: A Step-By-Step Detailed Protocol and Brief Review of Protamine Alterations. *Protein Pept. Lett.* **2018**, *25* (5), 424–433.

(24) Nanassy, L.; Liu, L.; Griffin, J.; Carrell, D. T. The Clinical Utility of the Protamine 1/protamine 2 Ratio in Sperm. *Protein Pept. Lett.* **2011**, *18* (8), 772–777.

(25) Ni, K.; Spiess, A.-N.; Schuppe, H.-C.; Steger, K. The Impact of Sperm Protamine Deficiency and Sperm DNA Damage on Human Male Fertility: A Systematic Review and Meta-Analysis. *Andrology* **2016**, *4* (5), 789–799.

(26) Siklenka, K.; Erkek, S.; Godmann, M.; Lambrot, R.; McGraw, S.; Lafleur, C.; Cohen, T.; Xia, J.; Suderman, M.; Hallett, M.; et al. Disruption of Histone Methylation in Developing Sperm Impairs Offspring Health Transgenerationally. *Science* **2015**, *350* (6261).

(27) Carrell, D. T.; Emery, B. R.; Hammoud, S. Altered Protamine Expression and Diminished Spermatogenesis: What Is the Link? *Hum. Reprod. Update* **2007**, *13* (3), 313–327.

(28) Hammoud, S. S.; Nix, D. A.; Zhang, H.; Purwar, J.; Carrell, D. T.; Cairns, B. R. Distinctive Chromatin in Human Sperm Packages Genes for Embryo Development. *Nature* **2009**, *460* (7254), 473–478.

(29) Castillo, J.; Amaral, A.; Azpiazu, R.; Vavouri, T.; Estanyol, J. M.; Ballesca, J. L.; Oliva, R. Genomic and Proteomic Dissection and Characterization of the Human Sperm Chromatin. *Mol. Hum. Reprod.* **2014**, *20* (11), 1041–1053.

(30) Arpanahi, A.; Brinkworth, M.; Iles, D.; Krawetz, S. A.; Paradowska, A.; Platts, A. E.; Saida, M.; Steger, K.; Tedder, P.; Miller, D. Endonuclease-Sensitive Regions of Human Spermatozoal Chromatin Are Highly Enriched in Promoter and CTCF

Binding Sequences. *Genome Res.* **2009**, *19* (8), 1338–1349.

(31) Royo, H.; Stadler, M. B.; Peters, A. H. F. M. Alternative Computational Analysis Shows No Evidence for Nucleosome Enrichment at Repetitive Sequences in Mammalian Spermatozoa. *Dev. Cell* **2016**, *37* (1), 98–104.

(32) Samans, B.; Yang, Y.; Krebs, S.; Sarode, G. V.; Blum, H.; Reichenbach, M.; Wolf, E.; Steger, K.; Dansranjavin, T.; Schagdarsurengin, U. Uniformity of Nucleosome Preservation Pattern in Mammalian Sperm and Its Connection to Repetitive DNA Elements. *Dev. Cell* **2014**, *30* (1), 23–35.

(33) Krejčí, J.; Stixová, L.; Pagáčová, E.; Legartová, S.; Kozubek, S.; Lochmanová, G.; Zdráhal, Z.; Sehnalová, P.; Dabravolski, S.; Hejátko, J.; et al. Post-Translational Modifications of Histones in Human Sperm. *J. Cell. Biochem.* **2015**, *116* (10), 2195–2209.

(34) Papoutsopoulou, S.; Nikolakaki, E.; Chalepakis, G.; Kruft, V.; Chevaillier, P.; Giannakouros, T. SR Protein-Specific Kinase 1 Is Highly Expressed in Testis and Phosphorylates Protamine 1. *Nucleic Acids Res.* **1999**, *27* (14), 2972–2980.

(35) Green, G. R.; Balhorn, R.; Poccia, D. L.; Hecht, N. B. Synthesis and Processing of Mammalian Protamines and Transition Proteins. *Mol. Reprod. Dev.* **1994**, *37* (3), 255–263.

(36) Pirhonen, A.; Linnala-Kankkunen, A.; Mënpää, P. H. P2 Protamines Are Phosphorylated in Vitro by Protein Kinase C, Whereas P1 Protamines Prefer cAMP-Dependent Protein Kinase. A Comparative Study of Five Mammalian Species. *Eur. J. Biochem.* **1994**, *223* (1), 165–169.

(37) Wu, J. Y.; Ribar, T. J.; Cummings, D. E.; Burton, K. A.; McKnight, G. S.; Means, A. R. Spermiogenesis and Exchange of Basic Nuclear Proteins Are Impaired in Male Germ Cells Lacking Camk4. *Nat. Genet.* **2000**, *25* (4), 448–452.

(38) Gusse, M.; Sautiere, P.; Belaiche, D.; Martinage, A.; Roux, C.; Dadoune, J. P.; Chevaillier, P. Purification and Characterization of Nuclear Basic Proteins of Human Sperm. *Biochim. Biophys. Acta* **1986**, *884* (1), 124–134.

(39) Itoh, K.; Kondoh, G.; Miyachi, H.; Sugai, M.; Kaneko, Y.; Kitano, S.; Watanabe, H.; Maeda, R.; Imura, A.; Liu, Y.; et al. Dephosphorylation of Protamine 2 at Serine 56 Is Crucial for Murine Sperm Maturation in Vivo. *Sci. Signal.* **2019**, *12* (574), eaao7232.

(40) Pirhonen, a; Linnala-Kankkunen, A.; Mäenpää, P. H. Identification of Phosphoseryl Residues in Protamines from Mature Mammalian Spermatozoa. *Biol. Reprod.* **1994**, *50* (5), 981–986.

(41) Chirat, F.; Arkhis, A.; Martinage, A.; Jaquinod, M.; Chevaillier, P.; Sautière, P. Phosphorylation of Human Sperm Protamines HP1 and HP2: Identification of Phosphorylation Sites. *Biochim. Biophys. Acta* **1993**, *1203* (1), 109–114.

(42) Castillo, J.; Estanyol, J. M.; Ballescà, J.; Oliva, R. Human Sperm Chromatin Epigenetic Potential: Genomics, Proteomics, and Male Infertility. *Asian J. Androl.* **2015**, *17* (4), 601.

(43) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; et al. How Many Human Proteoforms Are There? *Nat. Chem. Biol.* **2018**, *14* (3), 206–214.

(44) Bonet-Costa, C.; Vilaseca, M.; Diema, C.; Vujatovic, O.; Vaquero, A.; Omeñaca, N.; Castejón, L.; Bernués, J.; Giralt, E.; Azorín, F. Combined Bottom-up and Top-down Mass Spectrometry Analyses of the Pattern of Post-Translational Modifications of Drosophila Melanogaster Linker Histone H1. *J. Proteomics* **2012**, *75* (13), 4124–4138.

(45) Brunner, A. M.; Nanni, P.; Mansuy, I. M. Epigenetic Marking of Sperm by Post-

Translational Modification of Histones and Protamines. *Epigenetics and Chromatin* **2014**, *7* (1), 1–12.

(46) Cooper TG. WHO. Examination and Processing of Human Semen. Fifth Edition. *Who* **2010**.

(47) Mengual, L.; Ballescà, J. L.; Ascaso, C.; Oliva, R. Marked Differences in Protamine Content and P1 / P2 Ratios. *J. Androl.* **2003**, *24* (3), 438–447.

(48) Goodrich, R. J.; Anton, E.; Krawetz, S. A. Isolating mRNA and Small Noncoding RNAs from Human Sperm. *Methods Mol. Biol.* **2013**, *927*, 385–396.

(49) Jodar, M.; Sendler, E.; Moskovtsev, S. I.; Librach, C. L.; Goodrich, R.; Swanson, S.; Hauser, R.; Diamond, M. P.; Krawetz, S. A. Absence of Sperm RNA Elements Correlates with Idiopathic Male Infertility. *Sci. Transl. Med.* **2015**, *7* (295), 295re6.

(50) Sautière, P.; Martinage, A.; Bélaiche, D.; Arkhis, A.; Chevaillier, P. Comparison of the Amino Acid Sequences of Human Protamines HP2 and HP3 and of Intermediate Basic Nuclear Proteins HPS1 and HPS2. Structural Evidence That HPS1 and HPS2 Are pro-Protamines. *J. Biol. Chem.* **1988**, *263* (23), 11059–11062.

(51) Alimi, E.; Martinage, A.; Arkhis, A.; Belaiche, D.; Sautiere, P.; Chevaillier, P. Amino Acid Sequence of the Human Intermediate Basic Protein 2 (HPI2) from Sperm Nuclei. Structural Relationship with Protamine P2. *Eur. J. Biochem.* **1993**, *214* (2), 445–450.

(52) Martinage, A.; Arkhis, A.; Alimi, E.; Sautière, P.; Chevaillier, P. Molecular Characterization of Nuclear Basic Protein HPI1, a Putative Precursor of Human Sperm Protamines HP2 and HP3. *Eur. J. Biochem.* **1990**, *191* (2), 449–451.

(53) Carré-Eusèbe, D.; Lederer, F.; Lê, K. H. D.; Elsevier, S. M. Processing of the

Precursor of Protamine P2 in Mouse. Peptide Mapping and *N* -Terminal Sequence Analysis of Intermediates. *Biochem. J.* **1991**, *277* (1), 39–45.

(54) Chauvière, M.; Martinage, A.; Debarle, M.; Sautière, P.; Chevaillier, P. Molecular Characterization of Six Intermediate Proteins in the Processing of Mouse Protamine P2 Precursor. *Eur. J. Biochem.* **1992**, *204* (2), 759–765.

(55) Deutsch, E. W.; Overall, C. M.; Van Eyk, J. E.; Baker, M. S.; Paik, Y.-K.; Weintraub, S. T.; Lane, L.; Martens, L.; Vandenbrouck, Y.; Kusebauch, U.; et al. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. *J. Proteome Res.* **2016**, *15* (11), 3961–3970.

(56) Vandenbrouck, Y.; Lane, L.; Carapito, C.; Duek, P.; Rondel, K.; Bruley, C.; Macron, C.; Gonzalez de Peredo, A.; Couté, Y.; Chaoui, K.; et al. Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. *J. Proteome Res.* **2016**, *15* (11), 3998–4019.

(57) Zerbino, D. R.; Achuthan, P.; Akanni, W.; Amode, M. R.; Barrell, D.; Bhai, J.; Billis, K.; Cummins, C.; Gall, A.; Girón, C. G.; et al. Ensembl 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D754–D761.

(58) UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47* (D1), D506–D515.

(59) Schilling, S.; Wasternack, C.; Demuth, H.-U. Glutaminyl Cyclases from Animals and Plants: A Case of Functionally Convergent Protein Evolution. *Biol. Chem.* **2008**, *389* (8), 983–991.

(60) Pirhonen, a; Linnala-Kankkunen, a; Mäenpää, P. H. Identification of Phosphoseryl Residues in Protamines from Mature Mammalian Spermatozoa. *Biol. Reprod.* **1994**, *50* (5), 981–986.

(61) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.;

Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res.* **2015**, *43* (D1), D512–D520.

(62) Castillo, J.; Estanyol, J. M.; Ballescà, J. L.; Oliva, R. Human Sperm Chromatin Epigenetic Potential: Genomics, Proteomics, and Male Infertility. *Asian J. Androl.* **2015**, *17* (4), 601.

(63) Fallah, A.; Mohammad-Hasani, A.; Colagar, A. H. Zinc Is an Essential Element for Male Fertility: A Review of Zn Roles in Men's Health, Germination, Sperm Quality, and Fertilization. *J. Reprod. Infertil.* **2018**, *19* (2), 69–81.

(64) Colagar, A. H.; Marzony, E. T.; Chaichi, M. J. Zinc Levels in Seminal Plasma Are Associated with Sperm Quality in Fertile and Infertile Men. *Nutr. Res.* **2009**, *29* (2), 82–88.

(65) Bal, W.; Dyba, M.; Szewczuk, Z.; Jezowska-Bojczuk, M.; Lukszo, J.; Ramakrishna, G.; Kasprzak, K. S. Differential Zinc and DNA Binding by Partial Peptides of Human Protamine HP2. *Mol. Cell. Biochem.* **2001**, *222* (1–2), 97–106.

(66) Bench, G.; Corzett, M. H.; Kramer, C. E.; Grant, P. G.; Balhorn, R. Zinc Is Sufficiently Abundant within Mammalian Sperm Nuclei to Bind Stoichiometrically with Protamine 2. *Mol. Reprod. Dev.* **2000**, *56* (4), 512–519.

(67) Bianchi, F.; Rousseaux-Prevost, R.; Sautiere, P.; Rousseaux, J. P2 Protamines from Human Sperm Are Zinc -Finger Proteins with One Cys2His2 Motif. *Biochem. Biophys. Res. Commun.* **1992**, *182* (2), 540–547.

Tables and Figures legend

Table 1. Protamine proteoforms identified through the top-down MS approach.

Table 2. Protamine peptides identified with the bottom-up MS approach.

Figure 1. Overall strategy used for the identification of protamines and associated post-translational modifications (PTMs) in human sperm. After removing histones from semen samples, chromatin was decondensed and disulfide bonds reduced. Afterwards, chromatin was precipitated and purified protamines were obtained. For the top-down MS approach, disulfide bonds in intact protamines were re-reduced, protected with iodoacetamide and subjected to LC-MS after sample clean-up. MS1 and MS2 were deconvoluted, leading to the proteoform identification. For the bottom-up approach, purified protamines were digested with proteinase K and protamine peptides were subjected to LC-MS, which allowed the identification of protamine peptides and their associated PTMs. See supporting figure 1 for more details.

Figure 2. Overall top-down MS workflow. A) Chromatogram resolving P2 and P1 in different elution times. B) Full MS1 spectra of P2 and P1. C) Deconvolution of P2 with its mature forms HP3 with one phosphorylation, HP2 with one phosphorylation and naïve HP4, and P1 with the phosphorylation pattern. D) Deconvolution of MS2 for P1.

Figure 3. P1 and P2 proteoforms identified by the top-down MS approach. A) Graphical representation of the different protamine proteoforms identified with the ProSight engine, in which native proteoforms are located at point 0 and modifications are represented as a mass increase in the difference between the modification and the intact proteoform (Δm). B) Representation of P1 and P2 proteoforms identified and their associated PTMs with high confidence. Phosphorylated residues, +61 Da modified residues and pyroglutamic acid modified residues are indicated in green, light purple and orange, respectively. Scissors indicate the proteolytic site for mature forms (HP2 in yellow, HP3 in orange, and HP4 in blue).

Figure 4. Protamine peptides identified by the bottom-up MS approach. A) No peptides have been identified in the P1 amino acid sequence (upper box). The lower box shows the Identified peptides located into the amino acid sequence of pre-P2 and P2 isoform 2 including their associated PTMs represented in green squares. In light orange, the difference between the peptide sequence between P2 and P2 isoform 2 is highlighted. Scissors indicate the proteolytic site for mature forms (HP2 in yellow, HP3 in orange, and HP4 in blue). B) P2 isoform 2 spectra corresponding to the two unique peptides identified.

Figure 5. P2 isoform 2 (ENST00000435245.2) validated at the RNA level by direct sequencing. A) Graphical representation of the two exons comprising *PRM2* gene with the two different transcripts derived from splice variants. The red letters indicate the splice sites. B) P2 isoforms 2 and 3 separated by a 3 % agarose gel were excised and while direct sequencing directly sequenced P2 isoform 3, a semi-nested PCR followed by direct sequencing was necessary to obtain P2 isoform 2 sequence. The yellow box and brown letters indicate the final part of exon 1 common in all transcripts and the P2 isoform 3 and isoform 2 splice sites are shown in the green and grey boxes, respectively.

Supporting information

Supporting Table 1. Specific primers used for the validation of the P2 isoform 2 at the RNA level.

Supporting Figure 1. Workflow applied for the top-down MS and the bottom-up MS approaches. A) Top-down liquid chromatography, MS&MS/MS parameters and software used. B) Bottom-up liquid chromatography, MS&MS/MS parameters, database searches and software used.

Supporting Figure 2. Overall Strategy employed to identify PRM2 isoform 2 at the RNA level. A) Schematic representation of canonical PRM2 (in orange), PRM2 isoform 2 (in grey) and PRM2 isoform 3 (in green) genes and the primer sets used. B) Graphical

representation of the methodology performed to validate PRM2 isoform 2, in which PCR product was run in an agarose gel and the two products were excised. For PRM2 isoform 2 a semi-nested PCR was necessary prior to direct sequencing.

Supporting Figure 3. Fragmentation of the identified proteoforms.

Supporting Figure 4. Spectra corresponding to the truncated forms of P1 and P2. A) On the top, MS/MS spectrum of the naïve truncated form of P1 spanning residues 8-51 and the lower spectrum shows the truncated P1 form with S11 residue phosphorylated. B) MS/MS spectrum of the truncated pre-P2 spanning residues 8-102

Supporting Figure 5. Specific modified residues in P1 employing two different engines. A) Automated localization of phosphorylated sites using Prosight software. B) Automated localization of phosphorylated residues using TopPic software.

Supporting Figure 6. "CLUSTAL O" (http://www.ebi.ac.uk/Tools/msa/clustalo/) alignment corresponding to the intron sequences of the *PRM2* gene among different primate species. The *Homo Sapiens* sequence is shown in green and the splice site resulting in the splice variant of the P2 isoform 2 in bold.

* fully conserved residue, : conservation of strongly similar properties, . conservation of weakly similar properties.

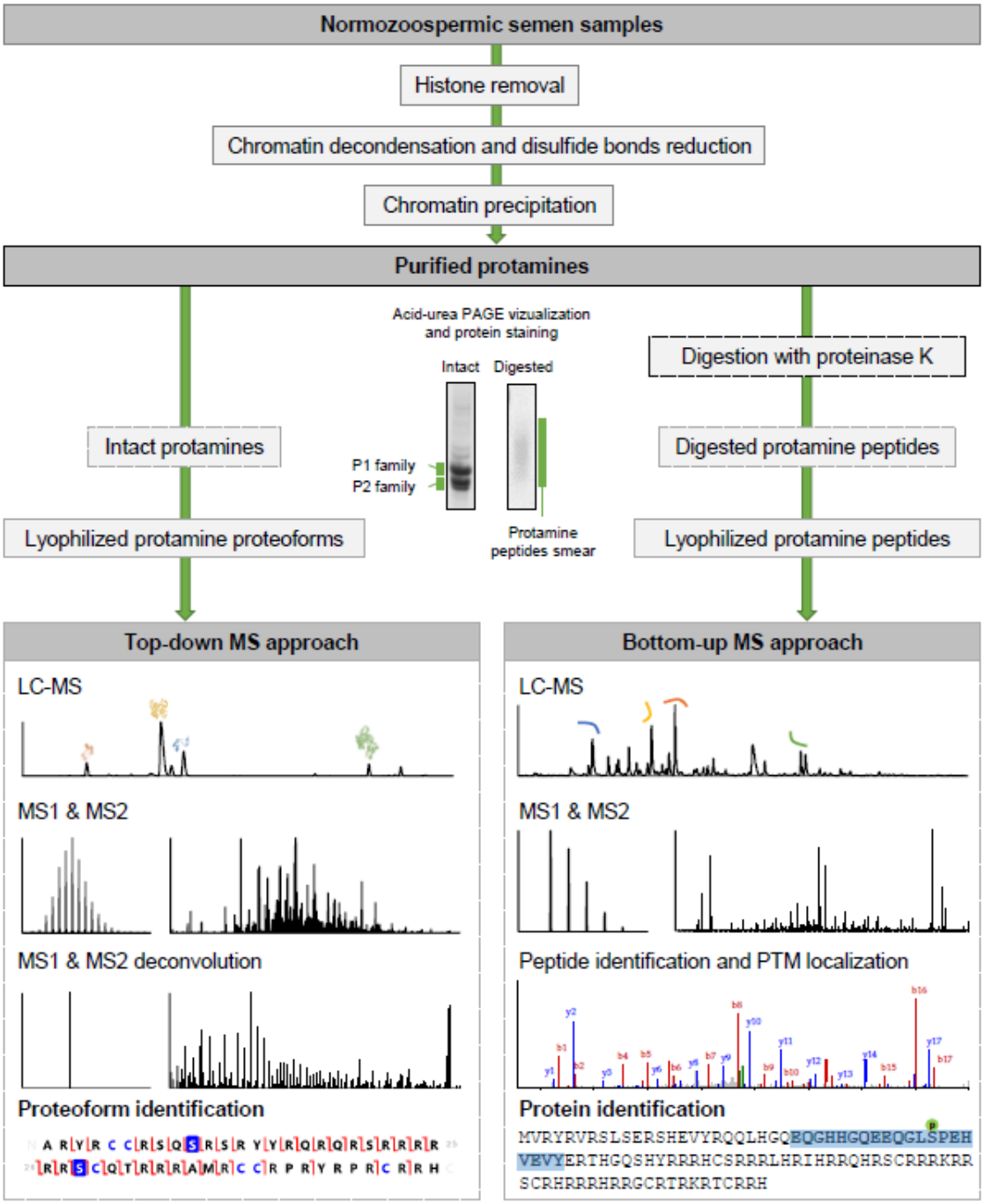


Figure 1. Overall strategy used for the identification of protamines and associated post-translational modifications (PTMs) in human sperm. After removing histones from semen samples, chromatin was decondensed and disulfide bonds reduced. Afterwards, chromatin was precipitated and purified protamines were obtained. For the top-down approach, disulfide bonds in intact protamines were re-reduced, protected with iodoacetamide and subjected to liquid chromatography followed by mass spectrometry (LC-MS) after sample clean-up. MS1 and MS2 were deconvoluted, leading to the proteoform identification. For the bottom-up approach, purified protamines were digested with proteinase K and protamine peptides were subjected to LC-MS, which allowed the identification of protamine peptides and their associated PTMs. See supporting figure 1 for more details.

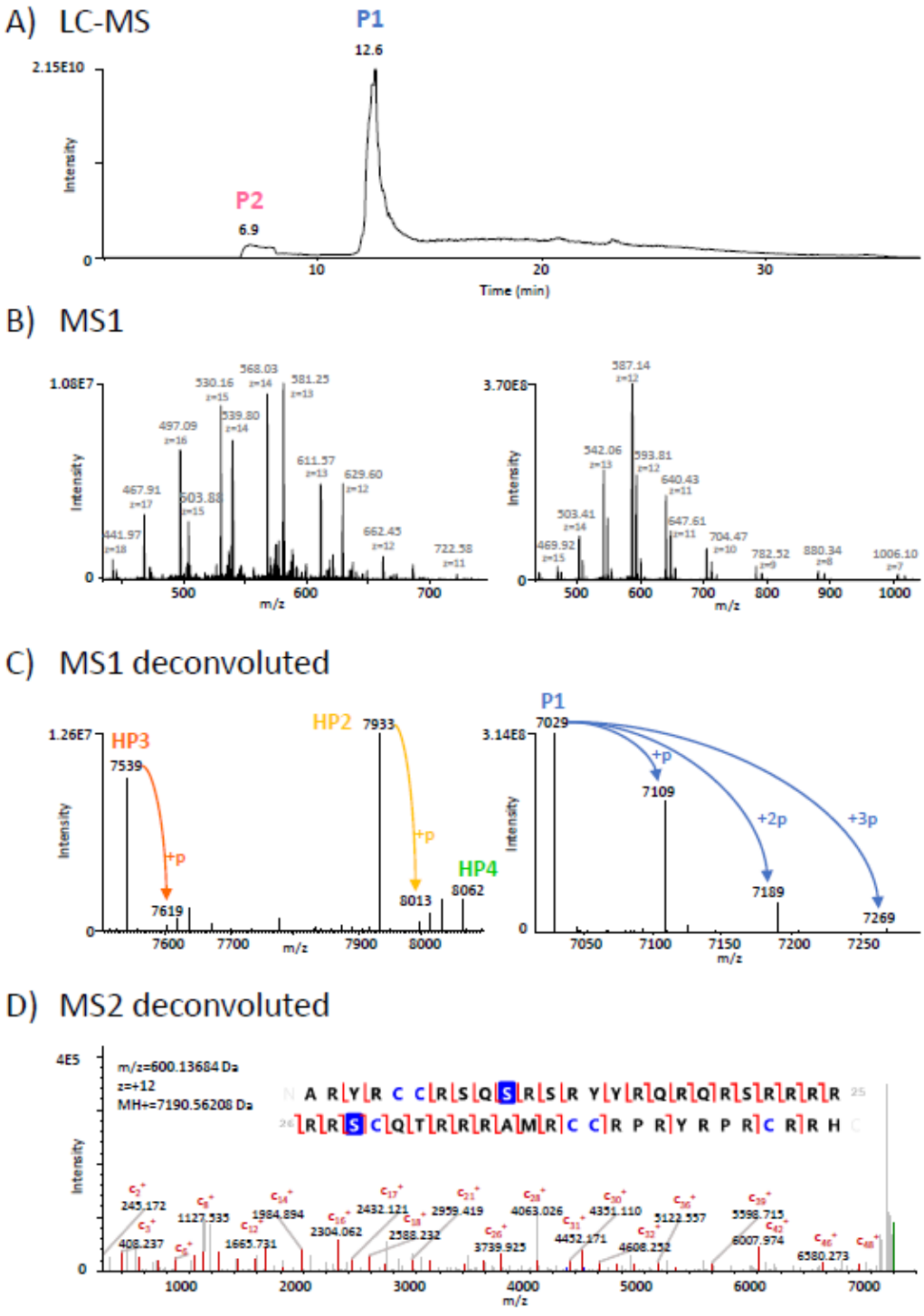


Figure 2. Overall top-down mass spectrometry workflow. A) Chromatogram resolving protamine 2 (P2) and protamine 1 (P1) in different elution times. B) Full MS1 spectra of P2 and P1. C) Deconvolution of P2 with its mature forms HP3 with one phosphorylation, HP2 with one phosphorylation and naïve HP4, and P1 with the phosphorylation pattern. D) Deconvolution of MS2 for P1.

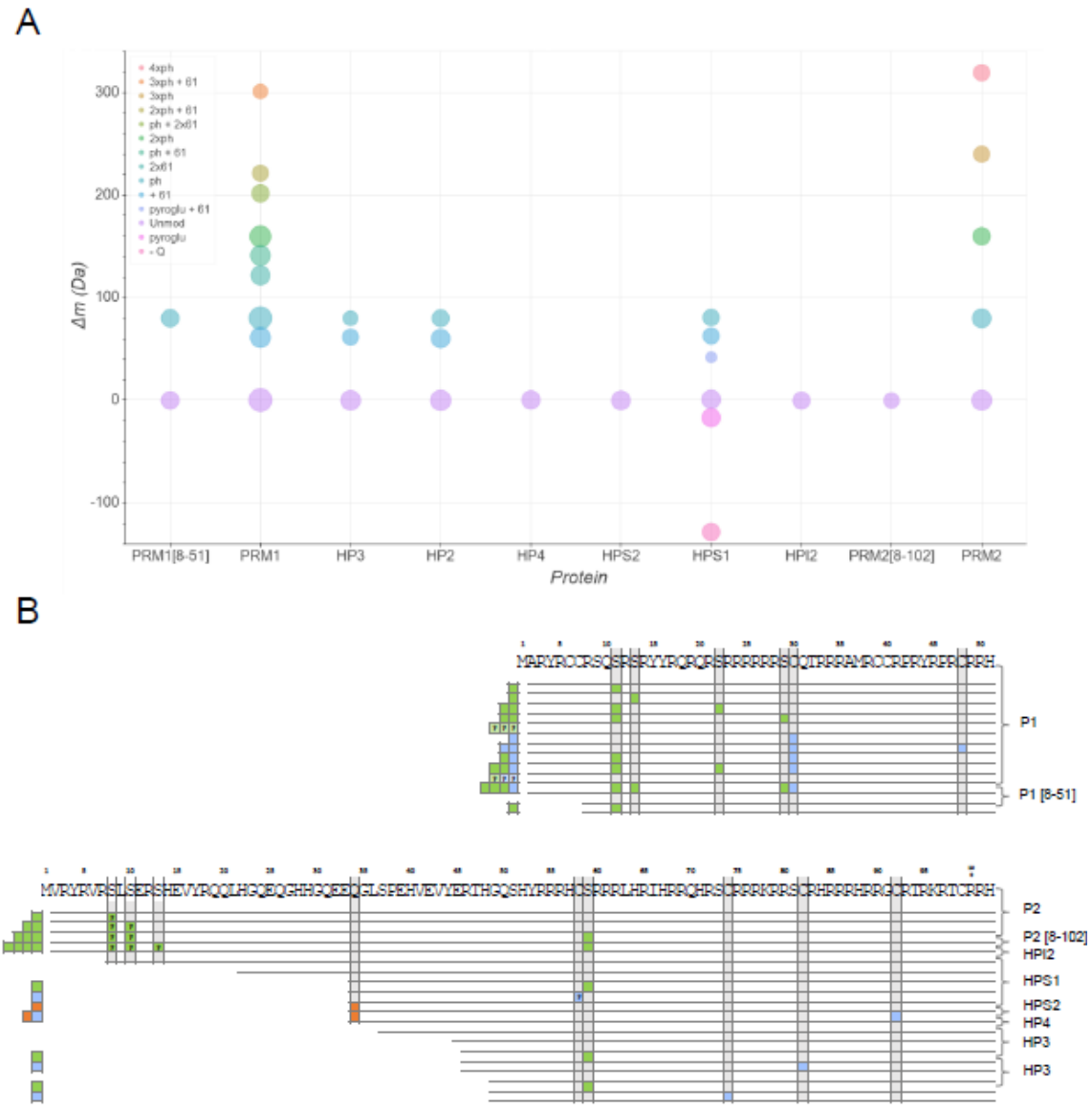


Figure 3. Protamine 1 (P1) and protamine 2 (P2) proteoforms identified by the top-down mass spectrometry approach. A) Graphical representation of the different protamine proteoforms identified with the ProSight engine, in which native proteoforms are located at point 0 and modifications are represented as an increase in the difference between the modification and the intact proteoform (Δm). B) Representation of P1 and P2 proteoforms identified and their associated PTMs with high confidence. Phosphorylated residues, +61 Da modified residues and pyroglutamic acid modified residues are indicated in green, light purple and orange, respectively. Scissors indicate the proteolytic site for mature forms (HP2 in yellow, HP3 in orange, and HP4 in blue).

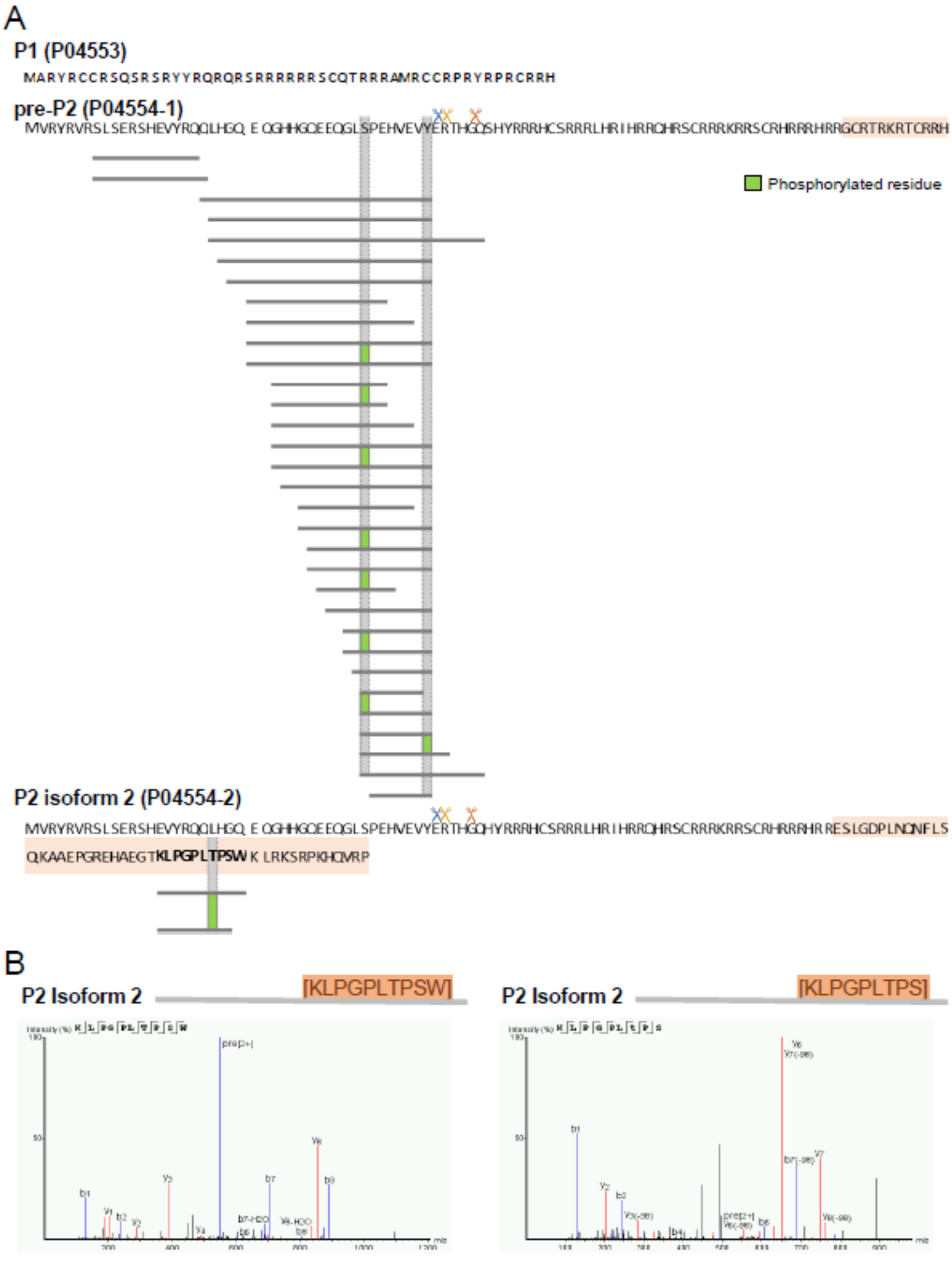


Figure 4. Protamine peptides identified by the bottom-up mass spectrometry (MS) approach. A) No peptides have been identified in the P1 amino acid sequence (upper box). The lower box shows the Identified peptides located into the amino acid sequence of pre-P2 and P2 isoform 2 including their associated post-translational modifications (PTMs) represented in green

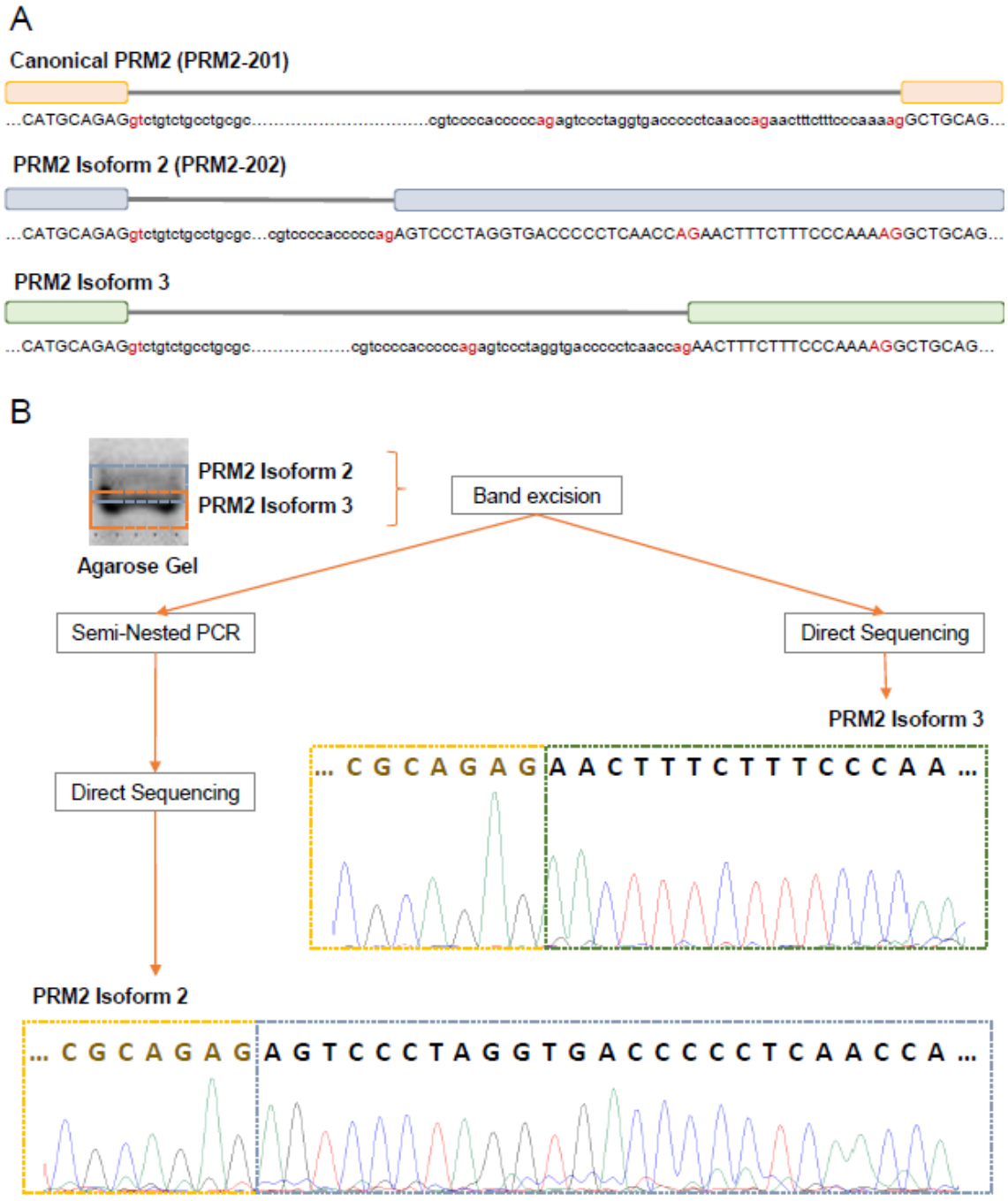
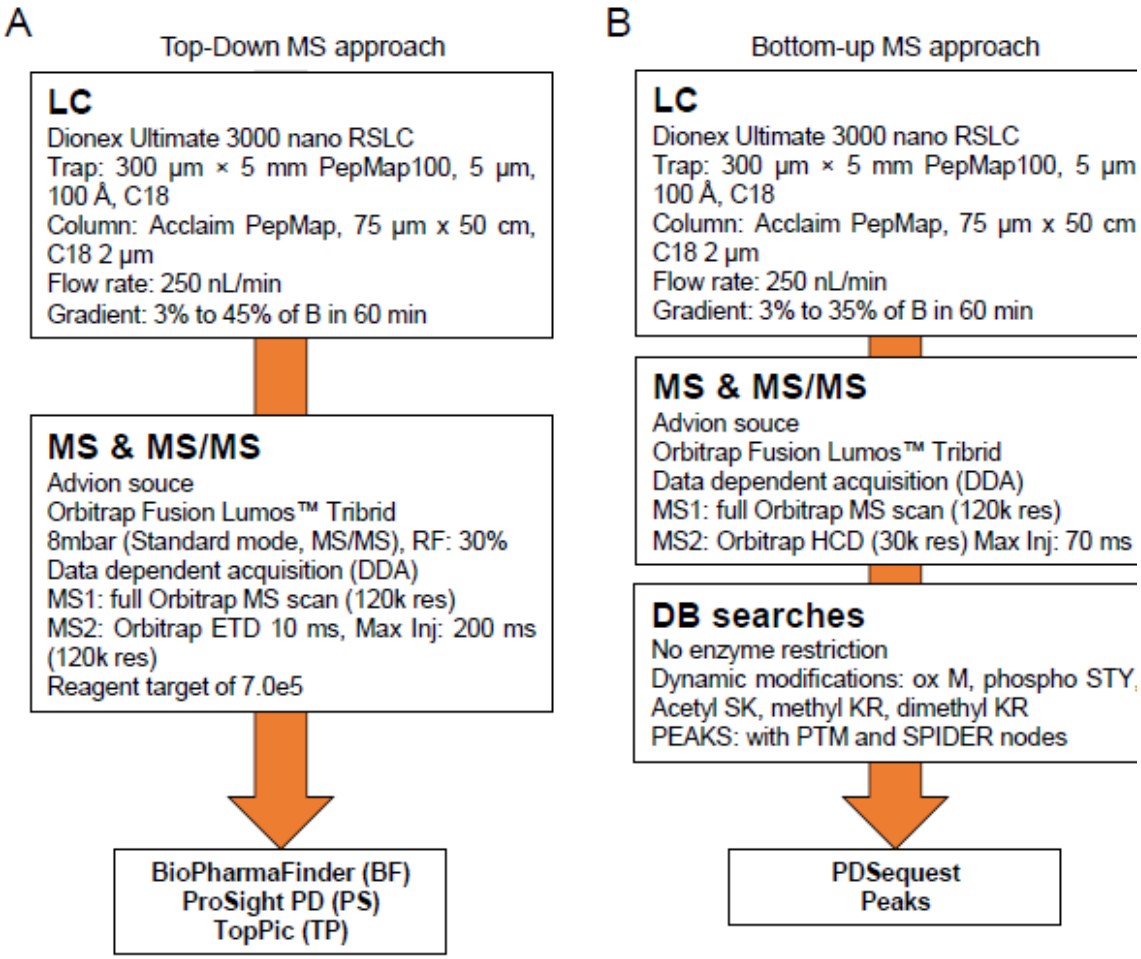
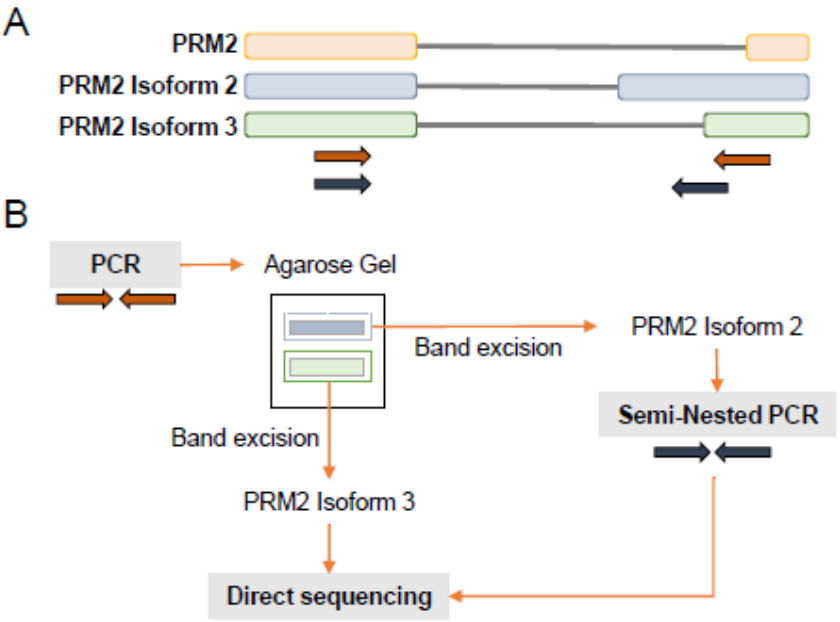


Figure 5. P2 isoform 2 (ENST00000435245.2) validated at the RNA level by direct sequencing. A) Graphical representation of the two exons comprising PRM2 gene with the two different transcripts derived from splice variants. The red letters indicate the splice sites. B) P2 isoforms 2 and 3 separated by a 3 % agarose gel were excised and while direct sequencing directly sequenced P2 isoform 3, a semi-nested PCR followed by direct sequencing was necessary to obtain P2 isoform 2 sequence. The yellow box and brown letters indicate the final part of exon 1 common in all transcripts and the P2 isoform 3 and isoform 2 splice sites are shown in the green and grey boxes, respectively.



Supporting Figure 1. Workflow applied for the top-down MS and the bottom-up MS approaches. A) Top-down liquid chromatography, MS&MS/MS parameters and software used. B) Bottom-up liquid chromatography, MS&MS/MS parameters, database searches and software used.

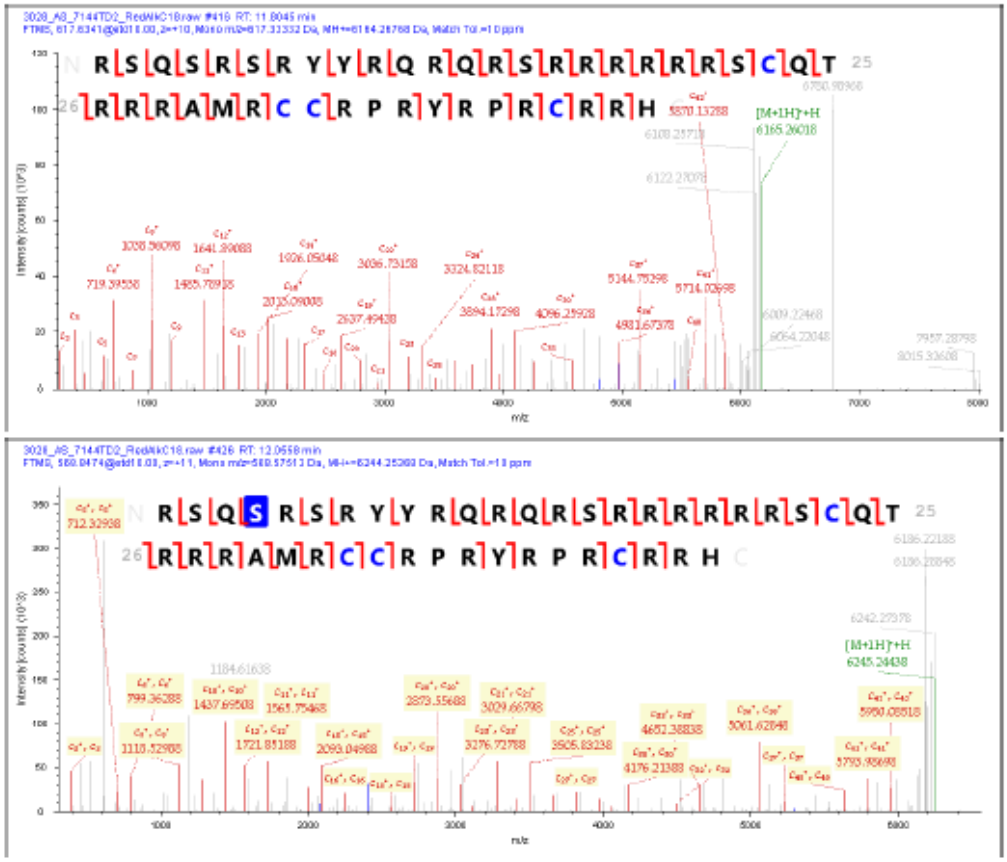


Supporting Figure 2. Overall Strategy employed to identify PRM2 isoform 2 at the RNA level. A) Schematic representation of canonical PRM2 (in orange), PRM2 isoform 2 (in grey) and PRM2 isoform 3 (in green) genes and the primer sets used. B) Graphical representation of the methodology performed to validate PRM2 isoform 2, in which PCR product was run in an agarose gel and the two products were excised. For PRM2 isoform 2 a semi-nested PCR was necessary prior to direct sequencing.

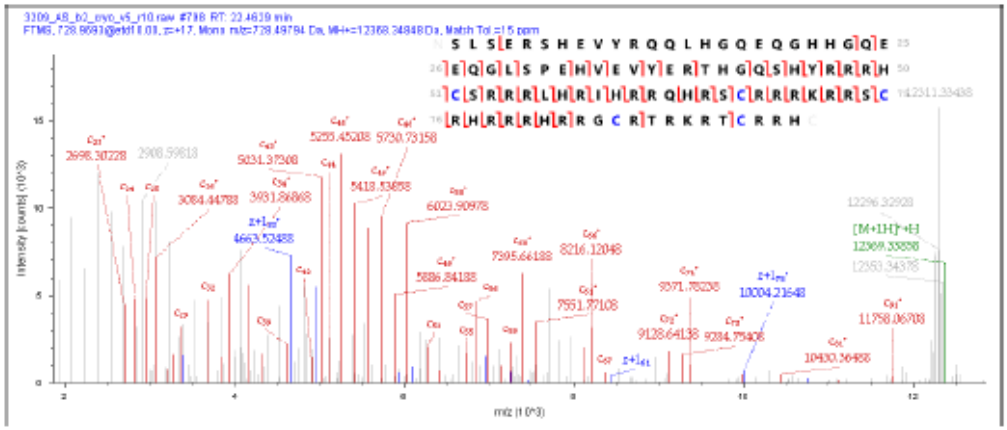
Label	Protein	PTMs	Observed mass (Da)	Theoretical mass (Da)	Mass difference (Da)	Mass difference (ppm)	PCS	P-Score	Fragmets explained (%)	Residue cleavages (%)
a	P1	Unmodified	7029.60	7029.60	0.009	1.27	1037.81	8.50E-88	45	84
b	P1	x1 ph	7109.56	7109.56	-0.004	-0.57	765.22	6.40E-67	26	80
c	P1	x1 ph	7109.61	7109.56	0.050	7.09	518.28	8.00E-48	27	69
d	P1	x2 ph	7189.58	7189.53	0.053	7.36	765.65	5.90E-67	37	80
e	P1	x2 ph	7189.55	7189.53	0.026	3.67	776.21	9.10E-68	32	76
f	P1	x1 +61	7090.62	7090.59	0.024	3.43	830.85	5.80E-72	30	80
g	P1	x2 +61	7151.63	7151.59	0.035	4.85	744.37	2.60E-65	24	67
h	P1	x1 ph + x1 +61	7170.57	7170.56	0.007	0.96	740.51	5.10E-65	28	80
i	P1	X2 ph + x1 +61	7250.55	7250.53	0.028	3.83	531.38	8.00E-49	20	65
j	P1	X3 ph + x1 +61	7330.54	7330.49	0.043	5.90	414.14	1.20E-39	19	51
k	P1 [8-51]	Unmodified	6163.27	6163.23	0.042	6.82	1091.69	6.60E-92	39	86
l	P1 [8-51]	x1 ph	6243.25	6243.20	0.048	7.74	604.46	1.70E-54	26	81

Supporting Figure 3. Fragmentation of the identified proteoforms.

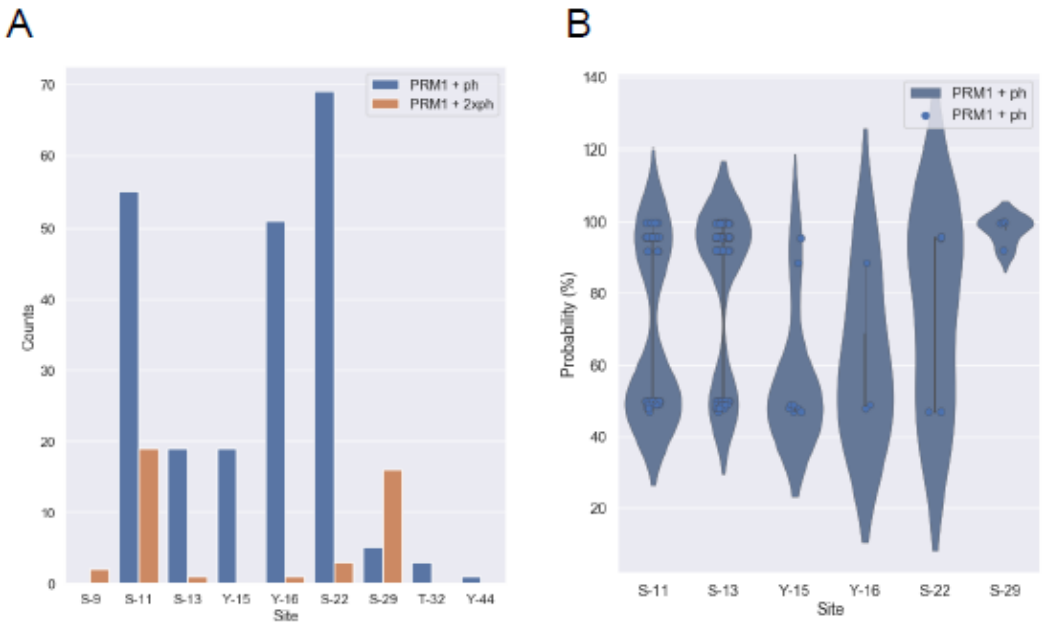
A



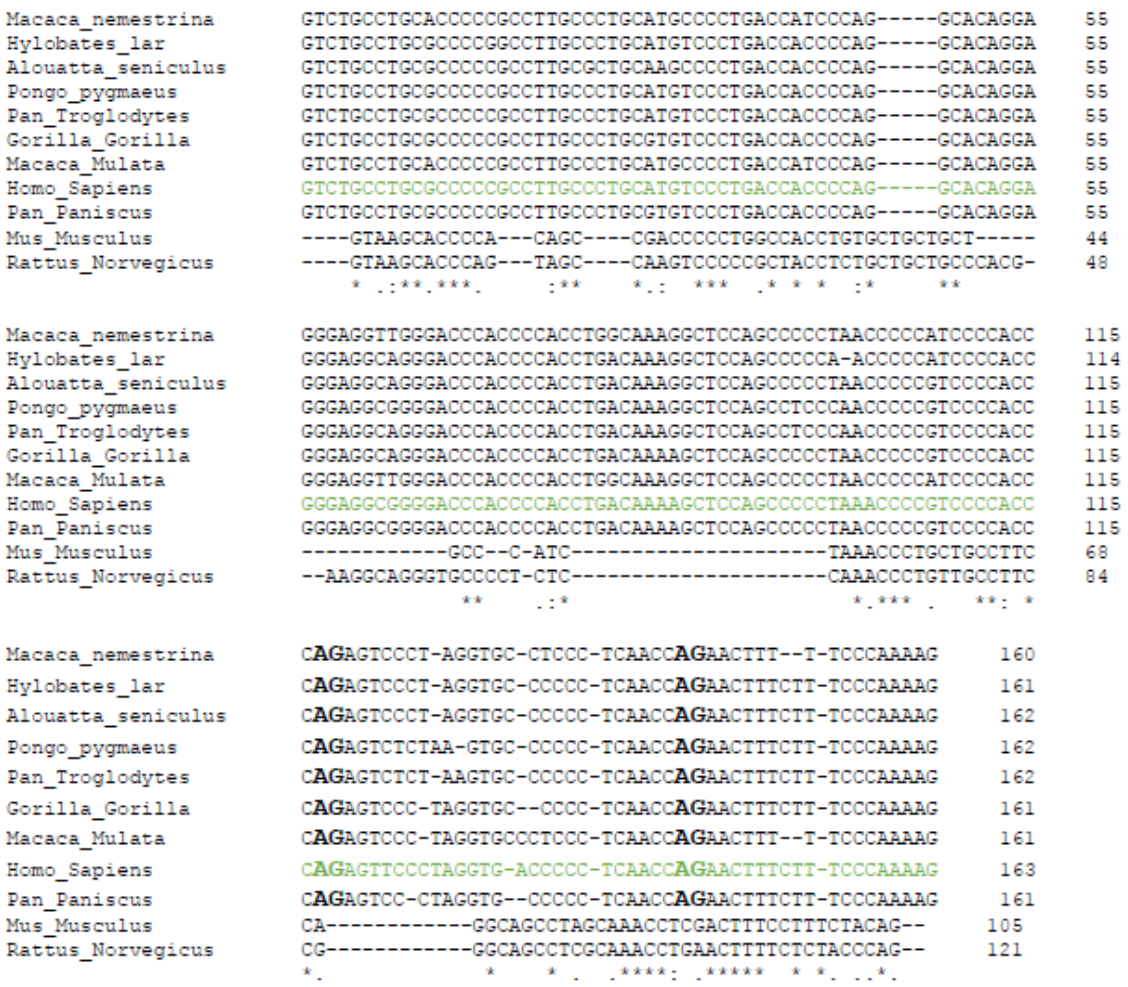
B



Supporting figure 4. Spectra corresponding to the truncated forms of P1 and protamine 2 (P2). A) On the top, MS/MS spectrum of the naïve truncated form of P1 spanning residues 8-51 and the lower spectrum shows the truncated P1 form with S11 residue phosphorylated. B) MS/MS spectrum of the truncated pre-P2 spanning residues 8-102



Supporting figure 5. Specific modified residues in P1 employing two different engines. A) Automated localization of phosphorylated sites using Prosight software. B) Automated localization of phosphorylated residues using TopPic software.



Supporting figure 6. "CLUSTAL O" (http://www.ebi.ac.uk/Tools/msa/clustalo/) alignment corresponding to the intron sequences of the PRM2 gene among different primate species. The Homo Sapiens sequence is shown in green and the splice sites resulting in the splice variant of the P2 isoform 2 and 3, respectively, in bold.

* fully conserved residue, : conservation of strongly similar properties, . conservation of weakly similar properties.