

## RAPID COMMUNICATION

# Finding new posttranslational modifications in salivary proline-rich proteins

Rui Vitorino<sup>1</sup>, Renato Alves<sup>1</sup>, António Barros<sup>1</sup>, Armando Caseiro<sup>1</sup>, Rita Ferreira<sup>1</sup>, Maria Calheiros Lobo<sup>2</sup>, Ana Bastos<sup>3</sup>, José Duarte<sup>4</sup>, Davide Carvalho<sup>5</sup>, Lúcio Lara Santos<sup>3,6</sup> and Francisco L. Amado<sup>1</sup>

<sup>1</sup> QOPNA, Mass Spectrometry Center, Department of Chemistry, University of Aveiro, Aveiro, Portugal

<sup>2</sup> CESPU, Dental School, Northern Health Institute, Porto, Portugal

<sup>3</sup> IPO-Porto, Portuguese Institute of Oncology Francisco Gentil, Porto, Portugal

<sup>4</sup> CIAFEL, Faculty of Sports Science, University of Porto, Porto, Portugal

<sup>5</sup> Endocrinology Department, S. João Hospital, University of Porto Medical School, Porto, Portugal

<sup>6</sup> University Fernando Pessoa, Porto, Portugal

Proline-rich proteins (PRPs) are the most complex family of salivary peptides with distinct isoforms and PTMs. Up to date, only the serine phosphorylation at positions 8, 17, and 22 have been experimentally observed on acidic PRP (aPRPs), and at position 8 on basic PRP1 and 2. The presence of a glucuronyl group at Ser17 was also noticed on aPRP. The main goal of this study was to identify new PTMs and distinct isoforms of salivary PRPs using LC-MALDI-TOF/TOF. Through the salivary peptidome characterization of 20 different subjects from Control, Diabetic, and Head and Neck Cancer groups, it was possible to identify the following species: (i) *N*-glycosylation sites: two in basic proline-rich protein 2 (bPRP2), one in bPRP3 and one in bPRP4; (ii) *O*-glycosylation sites: two in bPRP2 and one in aPRP; (iii) other terminal monosaccharide sites: six in bPRP1, two in bPRP2 and two in bPRP3; (iv) other modifications such as *N*-terminal pyro-Glu (two in bPRP1, six in bPRP2, eight in bPRP3 and nine in bPRP4); (v) phosphorylation in serine, three in bPRP1, one in bPRP2, one in bPRP3 and one in aPRP1; (vi) bPRP1 (allele S, allele M and variant CP5) and bPRP4 (allele M). In summary, salivary peptidome data analysis allowed the identification of 45 new PRP-modified residues, mainly due to glycosylation, phosphorylation and conversion of Gln to pyro-Glu. Moreover, comparing all subject groups, it was noticed a predominance of *N*-acetyl hexosamine modification on bPRPs in the Head and Neck Cancer patients.

Received: April 21, 2010

Revised: July 9, 2010

Accepted: August 3, 2010



## Keywords:

Biomedicine / LC-MS/MS / MALDI-TOF/TOF / Peptidomics / PTM / Saliva

Salivary secretions, similarly to other body fluids, contain several low-molecular-weight protein species with the unique distinction that peptide species with an *m/z* up to 20 kDa are responsible for about 40% of the total secreted

proteins [1]. Salivary peptides are mostly present as protein families; those that are closely related structurally [2, 3] have been grouped since the 1970s into a few major classes, namely, histatins, statherin, cystatins and proline-rich proteins (PRPs).

PRPs, which comprise 20–30% of all saliva protein content [4], are divided into two different groups: acidic proline-rich proteins (aPRPs) and basic proline-rich proteins (bPRPs). The bPRPs are typically divided into glycosylated PRPs and nonglycosylated species [5].

The aPRPs complex is expressed by the *PRH1* and *PRH2* genes, which express the forms PRP1, PRP2, PIFs and Dbs that are proteolytically cleaved at the RPPR motif located at

**Correspondence:** Professor Francisco L. Amado, Department of Chemistry, University of Aveiro, 3810-193, Aveiro, Portugal

**E-mail:** famado@ua.pt

**Fax:** +351234370084

**Abbreviations:** aPRP, acidic proline-rich protein; bPRP, basic proline-rich protein; FDR, false discovery rate; HexNac, *N*-acetyl hexosamine; HNC, head and neck cancer; PRP, proline-rich protein

residue Arg123 of the primary sequence [6–8]. The N-terminal cleavage of aPRPs originates the species PRP3, PRP4, Piff and the Dbf, whereas the C-terminal fragment corresponds to the P-C peptide. Beside this cleavage, all these species may undergo different PTMs, including phosphorylation at Ser8, Ser17 and Ser22 residues [6, 8] and the addition of a glucuronyl group at Ser17 on aPRPs [9].

bPRPs are expressed by four different genes named *PRB1-PRB4*. Several studies [5, 10–15] have pointed to a multitude of bPRPs sequences, classified as P-D, P-E, P-F and P-H by Isemura *et al.* [10–16], a designation which other researchers adopted [12–14]. Meanwhile, more species were identified and classified as IB1, IB4 (P-H), IB5 (P-D), IB6, IB7, IB8a, IB8b, IB8c (P-F), IB9 (P-E), II-1 and II-2 [5]. All these species result from proteolytic cleavage, which might occur during or before granule maturation as observed by Messana *et al.* [7]. In addition to the number of detected forms of bPRPs in saliva and similarly of aPRPs, these species undergo PTMs such as phosphorylation at Ser8 on bPRP1 [15]. As stated above, it has been suggested that some bPRPs should present glycosylated forms regarding the existence of potential glycosylation sites. However, there is no experimental evidence of glycosylated bPRPs. Besides PTMs, little is known concerning the bulk of bPRPs isoforms that could result from the cleavage of precursors proteins originating from the translation of many alleles associated with each *PRB* gene [7, 16, 17].

Considering the lack of information regarding salivary PTMs and PRPs isoforms, the main goal of this study was to look for evidence of their existence in healthy subjects and in patients with alterations in the carbohydrate metabolism. To achieve this goal, whole saliva was collected from 20 different subjects (aged 23–54) grouped into three distinct sets according to their clinical evaluation: ten with head and neck cancer (HNC), five with diabetes mellitus type 1 (DM1) and five classified as the control (Cont), corresponding to subjects with no evidence of salivary gland oral pathologies or inflammatory processes. This study was approved by the local ethics committee and all participating individuals gave their informed consent.

Unstimulated whole saliva was collected at 10:00 a.m. by direct draining into a saliva collection tube, after a fasting period of at least 2 h. Salivary peptide isolation was immediately performed following the procedure described by Vitorino *et al.* [18]. After passing the supernatant through 10 kDa filters, enriched peptides fractions were separated using an Ultimate 3000 (Dionex, LCPackings) onto a 150 mm × 75 μm Pepmap100 capillary analytical C18 column with 3 μm particle size (Dionex, LCPackings) at a flow rate of 300 nL/min. A linear gradient of 5% Buffer B (85% ACN, 5% 2-propanol, 0.04% TFA) to 50% Buffer B was run over a period of 35 min. Peptides eluting from the capillary column were mixed with a continuous flow of α-CHCA matrix solution (270 nL/min, 2 mg/mL in 70% ACN/0.1% TFA and internal standard Glu-Fib at 15 fmol) and were directly deposited onto the LC-MALDI plates at

20-s intervals for each spot (100 nL/fraction). For every separation run, 156 fractions were collected. The MALDI-TOF/TOF MS analysis was performed using a 4800 MALDI-TOF/TOF Analyzer (Applied Biosystems, Foster City, CA, USA). An *S/N* threshold of 50 was used to select peaks for MS/MS analyses. A fragmentation voltage of 2 kV was used throughout the automated runs. The spectra were processed and analyzed by the Global Protein Server Workstation (Applied Biosystems), which uses internal MASCOT software (v.2.1.0.4, Matrix Science, UK) for protein/peptide identification based on the peptide mass fingerprints and MS/MS data.

An initial search was performed against the Swiss-Prot protein database (March 2009, 428 650 entries) for *Homo sapiens*. An MS tolerance of 30 ppm was found for precursor ions and 0.3 Da for fragment ions, as well as two missed cleavages. Protein identifications were considered as reliable when the MASCOT score was >70 (MASCOT score was calculated as  $-10 \times \log P$ , where *P* is the probability that the observed match is a random event.). This is the lowest score indicated by the program as significant ( $p < 0.05$ ) and indicated by the probability of incorrect protein identification. For PTM characterization, all data were reprocessed in an internal database created with all entries retrieved from the initial search including all PRP variants described in the literature [19] and in Uni-Prot [20]. In order to estimate the false discovery rate (FDR) and considering the repetitive PRP motif, a random decoy database was created for all Swiss-Prot and internal database entries resulting in 12 and 5% of FDR, respectively (false-positive peptides/(false-positive peptides+total peptides)) × 100. Unique peptides retrieved from FDR search were considered for PTM characterization.

Searched PTMs (Supporting Information Table 1) were considered as reliable when the program indicated the MASCOT score as significant ( $p < 0.05$ ) and the individual ion score (calculated as described above for protein search) was higher than 20. Each PTM was further manually validated through PTM-specific neutral loss and a, b and/or y fragmentation ions series. Saliva presents highly proteolytic cascade events promoted by different enzymes that enlarge the spectrum of peptides identified [18, 21]. Using LC-MALDI for the analysis of salivary peptidome, we identified 1391 fragments from bPRPs (334, 298, 122 and 139 fragments of bPRP1, bPRP2, bPRP3 and bPRP4, respectively, covering 97, 94, 61 and 100% of the protein main sequence) and from aPRPs (498 fragments covering 100% of main sequence). Sample data processing, including mining 38 different PTMs (Supporting Information Table 1), allowed the identification of 75 unique modified peptides (12 different modifications) mainly due to glycosylation, phosphorylation and conversion of Gln to pyro-Glu (Table 1) as well as three different distinct isoforms. Taking into consideration this last issue, PRPs' polymorphism could raise the complexity of distinct isoforms deriving from the presence of different common alleles [22]. Although several

**Table 1.** List of all identified PTMs on PRPs

| Species<br>(Uni-Prot<br>accession<br>number) | Calc<br>mass | Obs<br>mass | Match<br>error<br>(Da) | Match<br>error<br>(ppm) | Start<br>seq pos | End<br>seq pos | Sequence                            | Residue<br>modified | Uni-Prot<br>indication | Literature<br>indication | Prediction<br>site   | Ion<br>score | Ion<br>score<br>C.I. (%) | Modification  | No. of<br>individuals<br>(20) | Group (No. of<br>individuals) |     |   |
|--|--------------|-------------|------------------------|-------------------------|------------------|----------------|-------------------------------------|---------------------|------------------------|--------------------------|--|--------------|--------------------------|---|-------------------------------|-------------------------------|-----|---|
|  |              |             |                        |                         |                  |                |                                     |                     |                        |                          |  |              |                          |   |                               | (10)                          | (5) |   |
| <b>bPRP1</b>                                 | 1380.68      | 1380.72     | -0.04                  | -29                     | 302              | 314            | PKGQGGPPAQQGS                       | Ser314              |                        |                          | YinOYang<br>prediction<br>hSer8,<br>hSer71,<br>hSer132,<br>hSer183,<br>hSer254,<br>hSer259,<br>hSer314   | 34           | 98                       | HexNAc (S)[13]  | 7                             | 5                             | 1   | 1 |
| <b>P04280</b>                                | 3154.60      | 3154.61     | 0.00                   | -1                      | 43               | 72             | GPPPOGKPKQGGPPPP<br>GKQGGPPQGGDKSR  | Ser71               |                        |                          |  | 39           | 99                       | HexNAc (S)[29]  | 3                             | 2                             | 1   | 0 |
|  | 2097.07      | 2097.08     | -0.01                  | -4                      | 54               | 72             | GPPPPGKPOGGPPPOGD<br>KSR            | Ser71               |                        |                          |  | 63           | 100                      | HexNAc (S)[18]  | 7                             | 4                             | 1   | 2 |
|  | 1653.81      | 1653.86     | -0.05                  | -30                     | 120              | 134            | GKQGGPPQGGDKSQS                     | Ser134              |                        |                          |  | 27           | 91                       | dHex (S)[15]  | 4                             | 1                             | 2   | 1 |
|  | 2027.02      | 2026.97     | 0.05                   | 26                      | 24               | 42             | SPOGGKPOGGPPPPPG<br>KPK             | Ser25               |                        |                          |  | 43           | 98                       | Hex (S)[1]  | 1                             | 1                             | 0   | 0 |
|  | 2077.99      | 2077.94     | 0.06                   | 27                      | 54               | 71             | GPPPPGKPOGGPPPOG<br>DKS             | Lys70               |                        |                          |  | 34           | 98                       | Glucosylgalactosyl<br>(K)[17]                         | 3                             | 3                             | 0   | 0 |
|  | 2220.09      | 2220.07     | 0.02                   | 7                       | 121              | 141            | KQGGPPQGGDKSQSFR                    | Lys131              |                        |                          |  | 32           | 97                       | Galactosyl (K)[11]                                    | 3                             | 3                             | 0   | 0 |
|  | 3135.52      | 3135.59     | -0.06                  | -20                     | 43               | 71             | GPPPOGKPKQGGPPPP<br>GKQGGPPQGGDKS   | Lys70               |                        |                          |  | 73           | 100                      | Glucosylgalactosyl<br>(K)[28]                         | 4                             | 3                             | 1   | 0 |
|  | 3138.53      | 3138.56     | -0.03                  | -8                      | 287              | 316            | GPPPOGKPKQGGPPPP<br>GKQGGPPAQQGGSKS | Lys313              |                        |                          |  | 35           | 98                       | Glucosylgalactosyl<br>(K)[29]                         | 4                             | 3                             | 1   | 0 |
|  | 1663.86      | 1663.84     | 0.01                   | 8                       | 26               | 42             | QGGKPKQGGPPPPGKPKQ                  | Gln25               |                        |                          |  | 50           | 100                      | Gln->pyro-Glu<br>(N-term O)[0]                        | 9                             | 5                             | 2   | 2 |
|  | 3379.51      | 3379.51     | 0.00                   | 0                       | 1                | 32             | QNLEDVSOEESPSLIAG<br>NPOGFPQGGKPKQ  | Gln1,<br>Ser8       |                        | pSer8                    | NetPhos<br>prediction<br>pSer8,<br>pSer73,<br>pSer76,<br>pSer134,<br>pSer137,<br>pSer185,<br>pSer198,<br>pSer254,<br>pSer256,<br>pSer259,<br>pSer318 | 34           | 97                       | Gln->pyro-Glu<br>(N-term O)[0],<br>Phospho<br>(ST)[8] | 3                             | 2                             | 1   | 0 |

Table 1. Continued

| Specie<br>(Uni-Prot<br>accession<br>number) | Calc<br>mass | Obs<br>mass | Match<br>error<br>(Da) | Match<br>error<br>(ppm) | Start<br>seq pos | End<br>seq pos | Sequence                         | Residue<br>modified | Uni-Prot<br>indication       | Literature<br>indication | Prediction<br>site  | Ion<br>score | Ion<br>score<br>C.I. (%) | Modification                   | No. of<br>individuals<br>(20) | Group (No. of<br>individuals) |                 |                |  |
|---|--------------|-------------|------------------------|-------------------------|------------------|----------------|----------------------------------|---------------------|------------------------------|--------------------------|---|--------------|--------------------------|--------------------------------|-------------------------------|-------------------------------|-----------------|----------------|--|
|   |              |             |                        |                         |                  |                |                                  |                     |                              |                          |   |              |                          |                                |                               | Cancer<br>(10)                | Diabetes<br>(5) | Control<br>(5) |  |
|   | 2375.16      | 2375.22     | -0.06                  | -24                     | 61               | 82             | POGPPPOGDKRSRSPPPG<br>KPK        | Ser76               |                              |                          |   | 65           | 99                       | Phospho (ST)[16]               | 6                             | 1                             | 2               | 3              |  |
|   | 2726.30      | 2726.38     | -0.08                  | -29                     | 123              | 148            | QGPFGGDKSOSRSPGK<br>POGPPPK      | Ser134              |                              |                          |   | 43           | 99                       | Phospho (ST)[12]               | 6                             | 2                             | 1               | 3              |  |
|   | 1766.87      | 1766.88     | -0.01                  | -4                      | 76               | 93             | SPPGKPOGPPPOGGNPPQ               |                     | Allele M                     |                          |   | 33           | 100                      |                                | 6                             | 3                             | 1               | 2              |  |
|   | 2719.38      | 2719.30     | 0.08                   | 28                      | 76               | 103            | SPPGKPOGPPPOGGNPPQ<br>GPPPPGKPK  |                     |                              |                          |   | 54           | 100                      |                                | 5                             | 2                             | 1               | 2              |  |
|   | 2205.15      | 2205.21     | -0.06                  | -27                     | 115              | 136            | GPPPFKPOGPPPOGDKSR<br>SPQ        |                     | Allele S                     |                          |   | 67           | 100                      |                                | 7                             | 3                             | 1               | 3              |  |
|   | 2896.49      | 2896.44     | 0.05                   | 16                      | 115              | 143            | GPPPFKPOGPPPOGDKSR<br>SPQSPGKPK  |                     |                              |                          |   | 45           | 100                      |                                | 6                             | 3                             | 0               | 3              |  |
|   | 1662.85      | 1662.84     | 0.01                   | 7                       | 296              | 313            | POGPPPKPOGPPAOGG                 |                     | CF5                          |                          |   | 32           | 100                      |                                | 4                             | 2                             | 0               | 2              |  |
| <b>bPRP2</b>                                | 2125.07      | 2125.04     | 0.03                   | 13                      | 249              | 268            | GPPPOGDKSOGPPPPG<br>KPK          | Asn256              | Asn152,<br>Asn214,<br>Asn256 |                          | NetNGlyc<br>prediction  | 83           | 100                      | HexNAc (N)[8]                  | 5                             | 5                             | 0               | 0              |  |
| <b>P02812</b>                               | 2183.07      | 2183.04     | 0.03                   | 13                      | 198              | 217            | GPPPFKPOGPPPOGDN<br>KSK          | Asn214              |                              |                          | hAsn152,<br>hAsn214,<br>hAsn256   | 31           | 95                       | HexNAc (N)[17]                 | 6                             | 4                             | 1               | 1              |  |
|   | 2333.15      | 2333.15     | 0.00                   | -2                      | 207              | 227            | GPPPOGDKSOSARSPPG<br>KPK         | Asn214              |                              |                          |   | 36           | 99                       | HexNAc (N)[8]                  | 2                             | 2                             | 0               | 0              |  |
|   | 2126.05      | 2126.03     | 0.02                   | 11                      | 198              | 217            | GPPPFKPOGPPPOGDNK<br>SQ          | Ser232              |                              |                          |   | 34           | 95                       | dHex (S)[19]                   | 4                             | 2                             | 2               | 0              |  |
|   | 3135.52      | 3135.59     | -0.06                  | -20                     | 64               | 92             | GPPPOGDKSOGPPPPGK<br>POGPPPOGDKS | Lys107              |                              |                          |   | 73           | 100                      | Glucosylgalactosyl<br>(K)[28]  | 2                             | 2                             | 0               | 0              |  |
|   | 1084.55      | 1084.54     | 0.02                   | 15                      | 382              | 391            | QFGPPRRPQ                        | Gln397              | pyro-Gln1                    |                          |   | 45           | 100                      | Gln->pyro-Glu<br>(N-term Q)[0] | 4                             | 1                             | 2               | 1              |  |
|   | 1313.67      | 1313.66     | 0.02                   | 12                      | 384              | 396            | QGPFRPPOGGRPS                    | Gln399              | pSer8                        |                          | NetPhos<br>prediction   | 38           | 99                       | Gln->pyro-Glu<br>(N-term Q)[0] | 8                             | 4                             | 2               | 2              |  |
|   |              |             |                        |                         |                  |                |                                  |                     |                              |                          | pSer8,<br>pSer36,<br>pSer94,<br>pSer97,<br>pSer156,<br>pSer157,<br>pSer159,<br>pSer218,<br>pSer221,<br>pSer258, |              |                          |                                |                               |                               |                 |                |  |

Table 1. Continued

| Specie<br>(Uni-Prot<br>accession<br>number) | Calc<br>mass | Obs<br>mass | Match<br>error<br>(Da) | Match<br>error<br>(ppm) | Start<br>seq pos | End<br>seq pos | Sequence                                       | Residue<br>modified | Uni-Prot<br>indication   | Literature<br>indication | Prediction<br>site   | Ion<br>score | Ion<br>score<br>C.I. (%) | Modification                   | No. of<br>individuals<br>(20) | Group (No. of<br>individuals) |                 |                |
|---|--------------|-------------|------------------------|-------------------------|------------------|----------------|--|---------------------|--|--------------------------|--|--------------|--------------------------|--------------------------------|-------------------------------|-------------------------------|-----------------|----------------|
|   |              |             |                        |                         |                  |                |  |                     |  |                          |  |              |                          |                                |                               | Cancer<br>(10)                | Diabetes<br>(5) | Control<br>(5) |
|   | 1605.81      | 1605.80     | 0.01                   | 6                       | 376              | 391            | QAPPAGPOGPPRPQ                                 | Gln391              |  |                          |  | 30           | 95                       | Gln->pyro-Glu<br>(N-term O)[0] | 6                             | 1                             | 3               | 2              |
|   | 1663.82      | 1663.85     | -0.04                  | -21                     | 47               | 63             | QGGNQPGPPPPGKPKQ                               | Gln62               |  |                          |  | 98           | 100                      | Gln->pyro-Glu<br>(N-term O)[0] | 6                             | 4                             | 1               | 1              |
|   | 1830.92      | 1830.81     | 0.11                   | 29                      | 374              | 391            | QPOAPPAGPOGPPRPQ                               | Gln389              |  |                          |  | 36           | 99                       | Gln->pyro-Glu<br>(N-term O)[0] | 2                             | 2                             | 0               | 0              |
|   | 2721.35      | 2721.32     | 0.03                   | 10                      | 47               | 74             | QGGNQPGPPPPGKPKQ<br>GPPQGGNKPKQ                | Gln62               |  |                          |  | 50           | 100                      | Gln->pyro-Glu<br>(N-term O)[0] | 5                             | 1                             | 4               | 0              |
|   | 3567.71      | 3567.46     | 0.25                   | 71                      | 355              | 390            | QOEGNPNQGGPPPPAGG<br>NPOQPOAPPAGQPKQ<br>GPPRPQ | Gln370              |  |                          |  | 34           | 98                       | Gln->pyro-Glu<br>(N-term O)[0] | 1                             | 1                             | 0               | 0              |
|   | 2916.35      | 2916.38     | -0.03                  | -11                     | 9                | 37             | QEEPSLAGNPOGAPP<br>OGGNKPOGPPSP                | Ser36               |  |                          |  | 36           | 100                      | Phospho (ST)<br>(S)[36]        | 6                             | 2                             | 1               | 3              |
| <b>bPRP3</b>                                | 1663.83      | 1663.85     | -0.03                  | -15                     | 229              | 242            | KPQGGPPPOEGNKPKQ                               | Asn239              |  |                          |  | 38           | 100                      | Hex (N)[11]                    | 5                             | 3                             | 1               | 1              |
| <b>Q04118</b>                               | 2097.04      | 2097.08     | -0.04                  | -21                     | 54               | 73             | GPPRRPGKPEGPPPPQGG<br>NOS                      | Ser72               |  |                          |  | 36           | 100                      | dHex (S)[20]                   | 6                             | 1                             | 3               | 2              |
|   | 2282.12      | 2282.10     | 0.01                   | 6                       | 52               | 72             | SQGPPRRPGKPEGPPPPQ<br>GGNQ                     | Asn71               | hAsn50,<br>hAsn71,<br>hAsn92,<br>hAsn113,<br>hAsn134,<br>hAsn155,<br>hAsn176,<br>hAsn197 |                          | NetNGlyc<br>prediction<br>hAsn50,<br>hAsn71,<br>hAsn92,<br>hAsn113,<br>hAsn134,<br>hAsn155,<br>hAsn176,<br>hAsn197 | 51           | 100                      | HexNAc (N)[20]                 | 6                             | 4                             | 1               | 1              |
|   | 1112.61      | 1112.60     | 0.01                   | 6                       | 262              | 272            | QLPPPPAGKPKQ                                   | Gln277              |  |                          |  | 43           | 100                      | Gln->pyro-Glu<br>(N-term O)[0] | 6                             | 1                             | 4               | 1              |
|   | 1142.60      | 1142.58     | 0.01                   | 10                      | 32               | 42             | QGPPIRRPGKPE                                   | Gln47               |  |                          | NetPhos<br>prediction<br>pSer8,<br>pSer12,   | 36           | 99                       | Gln->pyro-Glu<br>(N-term O)[0] | 3                             | 1                             | 2               | 0              |



Table 1. Continued

| Specie<br>(Uni-Prot<br>accession<br>number) | Calc<br>mass | Obs<br>mass | Match<br>error<br>(Da) | Match<br>error<br>(ppm) | Start<br>seq pos | End<br>seq pos | Sequence                    | Residue<br>modified | Uni-Prot<br>indication      | Literature<br>indication    | Prediction<br>site                                    | Ion<br>score | Ion<br>score<br>C.I. (%) | Modification                   | No. of<br>individuals<br>(20) | Group (No. of<br>individuals) |                 |                |
|---|--------------|-------------|------------------------|-------------------------|------------------|----------------|-----------------------------|---------------------|-----------------------------|-----------------------------|---|--------------|--------------------------|--------------------------------|-------------------------------|-------------------------------|-----------------|----------------|
|   |              |             |                        |                         |                  |                |                             |                     |                             |                             |   |              |                          |                                |                               | Cancer<br>(10)                | Diabetes<br>(5) | Control<br>(5) |
|   | 1338.64      | 1338.64     | 0.01                   | 6                       | 72               | 84             | QSOGPPHPGKPE                | Gln72               |                             |                             | NetPhos<br>prediction<br>pSer8,<br>pSer22,<br>pSer225 | 73           | 100                      | Gln->pyro-Glu<br>(N-term O)[0] | 4                             | 3                             | 1               | 0              |
|   | 1406.72      | 1406.71     | 0.01                   | 4                       | 177              | 189            | QSHRPPPPGKPE                | Gln177              |                             |                             |   | 46           | 100                      | Gln->pyro-Glu<br>(N-term O)[0] | 2                             | 2                             | 0               | 0              |
|   | 1705.87      | 1705.84     | 0.03                   | 17                      | 51               | 67             | QSOGPPPPGKPEGRPP            | Gln51               |                             |                             |   | 62           | 100                      | Gln->pyro-Glu<br>(N-term O)[0] | 3                             | 1                             | 1               | 1              |
|   | 1785.90      | 1785.85     | 0.05                   | 28                      | 72               | 88             | QSOGPPHPGKFERPPP            | Gln72               |                             |                             |   | 55           | 100                      | Gln->pyro-Glu<br>(N-term O)[0] | 1                             | 1                             | 0               | 0              |
|   | 1814.88      | 1814.83     | 0.05                   | 28                      | 197              | 214            | QSOGPPHPGKPEGPPPO           | Gln197              |                             |                             |   | 36           | 99                       | Gln->pyro-Glu<br>(N-term O)[0] | 1                             | 1                             | 0               | 0              |
|   | 2330.12      | 2330.09     | 0.03                   | 12                      | 197              | 219            | QSOGPPHPGKPEGPPPO           | Gln197              |                             |                             |   | 29           | 96                       | Gln->pyro-Glu<br>(N-term O)[0] | 1                             | 1                             | 0               | 0              |
|   | 2486.22      | 2486.20     | 0.02                   | 7                       | 197              | 220            | QSOGPPHPGKPEGPPPO<br>EGNKS  | Gln197              |                             |                             |   | 33           | 98                       | Gln->pyro-Glu<br>(N-term O)[0] | 2                             | 2                             | 0               | 0              |
|   | 1597.82      | 1597.82     | 0.00                   | 2                       | 94               | 109            | SOGTPPPGKPEGRPP<br>EGNKS    |                     | Allele M                    |                             |   | 32           | 99                       |                                | 3                             | 2                             | 0               | 1              |
| <b>aPRP</b>                                 | 1443.58      | 1443.61     | -0.03                  | -21                     | 21               | 30             | DSEQFIDEER                  | Ser22               | gSer17                      | gSer17                      |   | 45           | 100                      | Glucuronyl (S)[2]              | 5                             | 1                             | 2               | 2              |
| <b>P02810</b>                               | 2705.18      | 2705.10     | 0.08                   | 29                      | 21               | 43             | DSEQFIDEERQGPPLGGQ<br>OSQPS | Ser22               | gSer17                      | gSer17                      |   | 48           | 100                      | Glucuronyl (S)[2]              | 6                             | 1                             | 2               | 3              |
|   | 2306.01      | 2306.97     | 0.03                   | 15                      | 21               | 39             | DSEQFIDEERQGPPLGG<br>QQ     | Ser22               | gSer17                      | gSer17                      |   | 52           | 100                      | Glucuronyl (S)[2]              | 6                             | 1                             | 2               | 3              |
|   | 1461.54      | 1461.55     | 0.00                   | -3                      | 19               | 30             | GGDSEQFIDEER                | Ser22               | pSer8,<br>pSer17,<br>pSer22 | pSer8,<br>pSer17,<br>pSer22 | NetPhos<br>prediction<br>pSer8,<br>pSer22,<br>pSer43  | 70           | 100                      | Phospho (ST)<br>(S)[4]         | 16                            | 10                            | 2               | 4              |
|   | 1576.57      | 1576.57     | 0.00                   | 3                       | 18               | 30             | DGGDSEQFIDEER               | Ser22               |                             |                             |   | 64           | 100                      | Phospho (ST)<br>(S)[5]         | 12                            | 4                             | 4               | 4              |
|   | 1589.60      | 1589.60     | 0.01                   | 4                       | 19               | 31             | GGDSEQFIDEERQ               | Ser22               |                             |                             |   | 50           | 100                      | Phospho (ST)<br>(S)[4]         | 11                            | 4                             | 3               | 4              |
|   | 1704.63      | 1704.63     | 0.00                   | 1                       | 18               | 31             | DGGDSEQFIDEERQ              | Ser22               |                             |                             |   | 63           | 100                      | Phospho (ST)<br>(S)[5]         | 8                             | 2                             | 4               | 2              |
|   | 1875.75      | 1875.74     | 0.01                   | 5                       | 15               | 30             | VISDGDSEQFIDEER             | Ser22               |                             |                             |   | 47           | 100                      | Phospho (ST)<br>(S)[8]         | 1                             | 1                             | 0               | 0              |
|   | 1953.81      | 1953.83     | -0.01                  | -8                      | 19               | 35             | GGDSEQFIDEERQGPLL           | Ser22               |                             |                             |   | 78           | 100                      | Phospho (ST)<br>(S)[4]         | 7                             | 1                             | 4               | 2              |

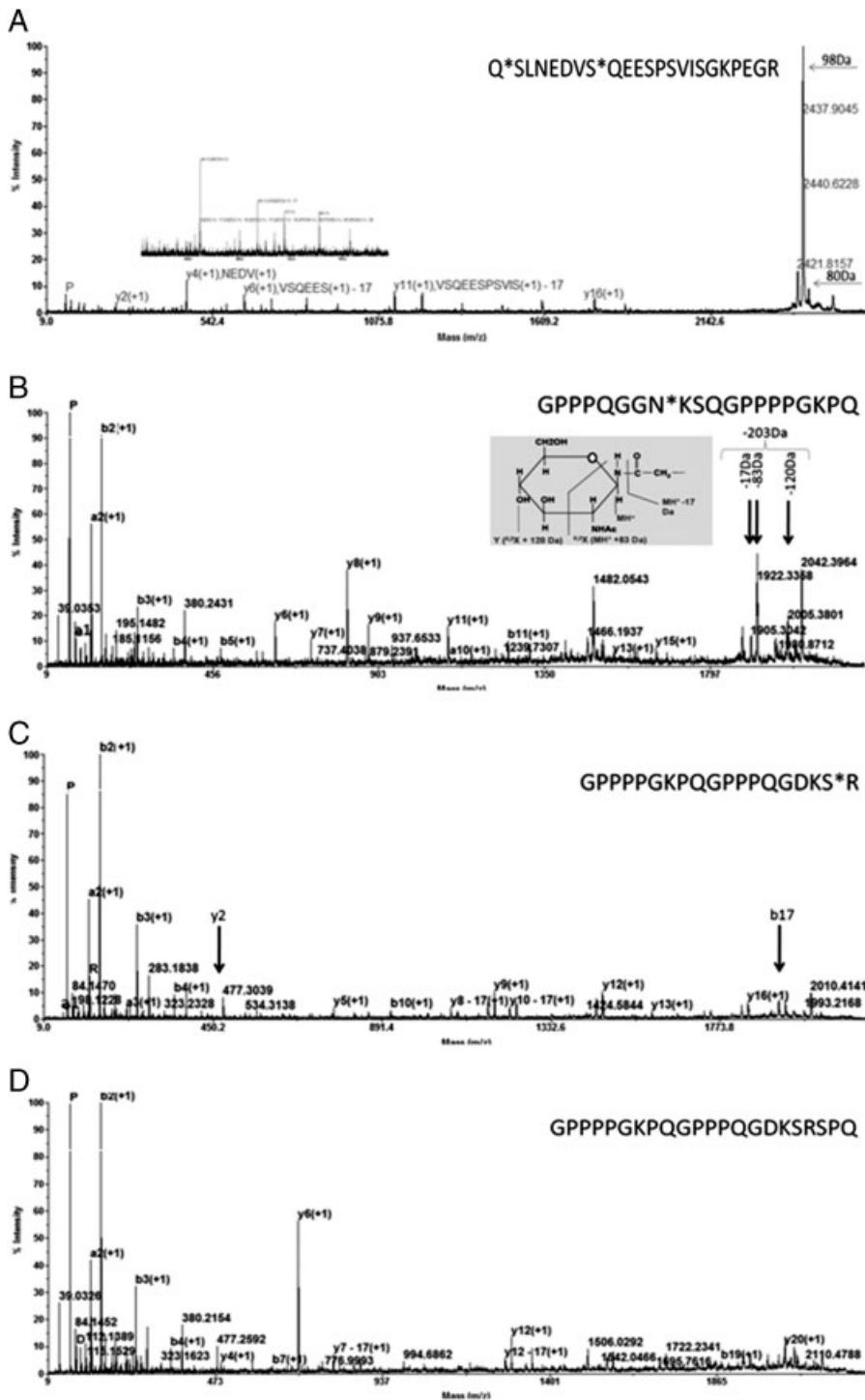
Table 1. Continued

| Specie<br>(Uni-Prot<br>accession<br>number) | Calc<br>mass | Obs<br>mass | Match<br>error<br>(Da) | Match<br>error<br>(ppm) | Start<br>seq pos | End<br>seq pos | Sequence                       | Residue<br>modified | Uni-Prot<br>indication | Literature<br>indication | Prediction<br>site | Ion<br>score | Ion<br>score<br>C.I. (%) | Modification            | No. of<br>individuals<br>(20) | Group (No. of<br>individuals) |                 |                |
|---|--------------|-------------|------------------------|-------------------------|------------------|----------------|--------------------------------|---------------------|------------------------|--------------------------|--------------------|--------------|--------------------------|-------------------------|-------------------------------|-------------------------------|-----------------|----------------|
|   |              |             |                        |                         |                  |                |                                |                     |                        |                          |                    |              |                          |                         |                               | Cancer<br>(10)                | Diabetes<br>(5) | Control<br>(5) |
| 2003.81                                     | 2003.82      | 2003.82     | -0.01                  | -4                      | 15               | 31             | VISDGDSEQFIDEERO               | Ser22               |                        |                          |                    | 72           | 100                      | Phospho (ST)<br>(S)[8]  | 1                             | 1                             | 0               | 0              |
| 2010.83                                     | 2010.85      | 2010.85     | -0.01                  | -6                      | 19               | 36             | GGDSEQFIDEERGGPPLG             | Ser22               |                        |                          |                    | 78           | 100                      | Phospho (ST)<br>(S)[4]  | 4                             | 1                             | 2               | 1              |
| 2067.86                                     | 2067.85      | 2067.85     | 0.01                   | 4                       | 19               | 37             | GGDSEQFIDEERGGPPL<br>GG        | Ser22               |                        |                          |                    | 80           | 100                      | Phospho (ST)<br>(S)[4]  | 17                            | 9                             | 5               | 3              |
| 2195.91                                     | 2195.88      | 2195.88     | 0.03                   | 14                      | 19               | 38             | GGDSEQFIDEERGGPPL<br>GGQ       | Ser22               |                        |                          |                    | 69           | 100                      | Phospho (ST)<br>(S)[4]  | 9                             | 4                             | 4               | 1              |
| 2299.99                                     | 2300.01      | 2300.01     | -0.02                  | -10                     | 11               | 30             | DVPLVISDGGDSEOFID<br>EER       | Ser22               |                        |                          |                    | 66           | 100                      | Phospho (ST)<br>(S)[12] | 4                             | 2                             | 2               | 0              |
| 2323.97                                     | 2323.91      | 2323.91     | 0.06                   | 25                      | 19               | 39             | GGDSEQFIDEERGGPPLG<br>GQQ      | Ser22               |                        |                          |                    | 82           | 100                      | Phospho (ST)<br>(S)[4]  | 15                            | 6                             | 4               | 5              |
| 2439.00                                     | 2438.99      | 2438.99     | 0.01                   | 3                       | 18               | 39             | DGGDSEQFIDEERGGPPL<br>GGQQ     | Ser22               |                        |                          |                    | 51           | 100                      | Phospho (ST)<br>(S)[5]  | 9                             | 4                             | 4               | 1              |
| 2723.15                                     | 2723.07      | 2723.07     | 0.08                   | 28                      | 19               | 43             | GGDSEQFIDEERGGPPL<br>GGQQSQPS  | Ser22               |                        |                          |                    | 28           | 100                      | Phospho (ST)<br>(S)[4]  | 7                             | 1                             | 4               | 2              |
| 2738.18                                     | 2738.19      | 2738.19     | -0.01                  | -3                      | 15               | 39             | VISDGDSEQFIDEERQG<br>PPLGGQQ   | Ser22               |                        |                          |                    | 50           | 100                      | Phospho (ST)<br>(S)[8]  | 3                             | 1                             | 2               | 0              |
| 2838.18                                     | 2838.10      | 2838.10     | 0.08                   | 28                      | 18               | 43             | DGGDSEQFIDEERGGPPL<br>GGQQSQPS | Ser22               |                        |                          |                    | 38           | 100                      | Phospho (ST)<br>(S)[5]  | 4                             | 1                             | 2               | 1              |



distinct isoforms of three alleles (PIF, Db and Pa) associated to *PRH1* had been identified in saliva [23], to date there is no evidence of distinct isoforms of other bPRPs allele variants. The major characteristic of these distinct isoforms is the depletion of internal sequences that remain conserved both at N- and C-terminal. For instance, the bPRP1 allele S

isoform presents a deletion on the main sequence between residues 118 and 239. Using this strategy, positive identifications were obtained from searching all obtained spectra for bPRP1 alleles S, M, CP5 and bPRP4 allele M. For instance, the *m/z* 2205.21 specie (Table 1) was assigned to bPRP1 allele S which corresponds to the sequence

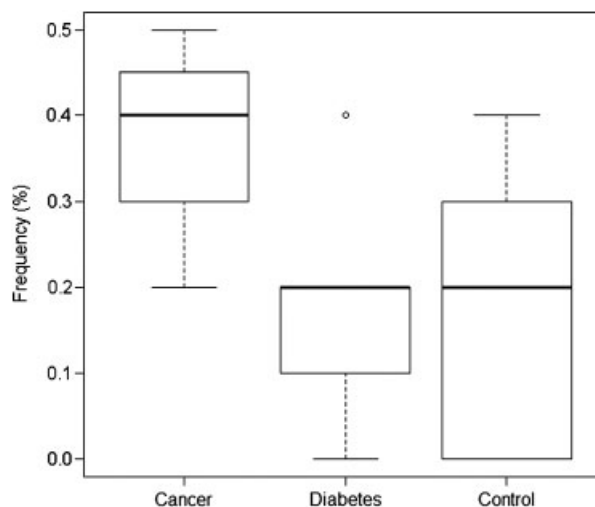


**Figure 1.** Representative MS/MS spectra for major PTMs identified in PRP. (A) Identification of phosphorylation on Ser 8 and conversion of Gln to pyro-Glu from peptide with *m/z* 2535.11 which belongs to bPRP3; (B) Identification of N-glycosylation on Asn8 from peptide with *m/z* 2125.04 which belongs to bPRP2; (C) Identification of O-glycosylation on Ser18 from the peptide with *m/z* 2097.08 which belongs to bPRP1; (D) MS/MS spectrum of peptide *m/z* 2205.92 corresponding to the alternative splicing variant BPRP1(S).

GPPPPGKPPQGGDKSRSPQ (115–136) (Fig. 1D), covering the sequence gap originating from the internal deletion.

When searching for phosphorylated species with the exception of bPRP4, all other forms present at least one phosphorylated serine residue. Overall, 21 different peptides were identified as presenting one or more phosphorylated serine residues, mostly corresponding to aPRPs (16/21) phosphopeptides. Although Ser8, Ser17 and Ser22 have been described in the literature [24, 25] as aPRPs phosphorylated sites, we only detected this modification at Ser22. The remaining peptides belong to bPRP1, bPRP2 and bPRP3. Potential phosphorylated sites can be retrieved from NetPhos v2.0 prediction [26], based on the main sequence analysis. Identified potential hits were assigned for scores higher than 0.9 resulting in 15, 18 and 5 potential Ser residues, respectively. From this analysis, we detected the phosphorylation at Ser8 on bPRP1, as mentioned previously [15], in addition to the phosphorylation at Ser134 on bPRP2 and Ser8 on bPRP3. In Fig. 1A, it is shown that the MS/MS spectra of the specie at  $m/z$  2535.11 are assigned to the sequence QSLNEDVSQEESPSVISGKPEGR from bPRP3. Through fragmentation spectra analysis, it was observed that the phosphate group is located at Ser8 while the *N*-terminal presents the conversion of Gln (pyro-Glu). It is noteworthy that the presence of phosphorylated Ser8 on bPRP3 and Ser134 on bPRP2 has never been detected or even suggested. Similarly to aPRPs, the presence of additional phosphorylation sites in bPRPs may suggest a possible participation in calcium homeostasis.

In this study, 21 unique peptides containing terminal monosaccharide units belonging mainly to bPRPs class were identified. Among the widespread modifications found, it must be noted the assignment of *N*-acetyl hexosamine (HexNac) attached to Ser (*O*-glycosylation) and to Asn (*N*-glycosylation) residues. A typical fragmentation profile for *N*-HexNac was observed on MS/MS spectra, and is shown in Fig. 1B. As can be observed, typical fragment ions derive from the 17, 120 and 203 Da neutral losses corresponding to the loss of ammonia from asparagines, a cleavage of a half cross-ring fragmentation and the *N*-acetyl hexosamine attached to the asparagines, respectively. In the case of *O*-HexNac, the MS/MS identification is achieved by the detection of  $y_2$  or  $b_{17}$  ions plus 203 Da (Fig. 1C). Several potential sites for *N*-glycosylation based on Asn-Xaa-Ser/Thr sequons have been described for bPRPs (three on bPRP2, eight on bPRP3 and eight on bPRP4). From these potential sites, we detected *N*-HexNac at Asn214 and Asn256 on bPRP2, Asn71 on bPRP3 and on bPRP4, whereas *O*-glycosylation sites were observed only on bPRP1 at Ser71 and Ser314. This screening also revealed precursors that lost the terminal sugars fucose, dHex (NL 146) present at Ser134 (bPRP1), Ser232 (bPRP2) and Ser232 (bPRP3) and hexose, Hex (NL 162, referring to mannose, glucose or galactose) at Ser25 (bPRP1) and Asn239 (bPRP3). Previously, *O*-linked



**Figure 2.** Box plot as a function of HexNac.

hexuronic acid was identified in aPRPs at Ser17 along with pyro-Glu and phosphopeptide [9]. In our study, we also identified this modification, although at Ser22, which seems to alternate with phosphorylation, as described by Jonsson *et al.* [9] for Ser17.

The above-mentioned modifications were observed in all groups of subjects. In order to compare the occurrence of these PTMs among the Cont, DM1 and HNC groups, we performed a PTM frequency peptide distribution analysis using the boxplot (Fig. 2), the most frequently used graphical technique for visualizing the distribution of continuous univariate data [27]. It was only observed a predominance of HexNac modification on bPRPs in HNC group which was clearly distinct from the other two groups (Fig. 2). This result is in agreement with the literature findings, namely for lung cancer, where it was observed the accumulation of *O*-GlcNAcylation in the glycan profile of patient tissue samples [28]. Moreover, further investigation is necessary to elucidate the meaning and the role of HexNac modification of bPRPs in oral cavity.

*The authors express their appreciation for the financial support provided by the “Fundação para a Ciência e Tecnologia” (FCT- Grant no. PTDC/QUI/72683/2006) and Universidade Fernando Pessoa.*

*The authors have declared no conflict of interest.*

## References

- [1] Amado, F. M., Vitorino, R. M., Domingues, P. M., Lobo, M. J., Duarte, J. A., Analysis of the human saliva proteome. *Expert Rev. Proteomics* 2005, 2, 521–539.
- [2] Helmerhorst, E. J., Oppenheim, F. G., Saliva: a dynamic proteome. *J. Dent. Res.* 2007, 86, 680–693.

- [3] Hu, S., Loo, J. A., Wong, D. T., Human body fluid proteome analysis. *Proteomics* 2006, 6, 6326–6353.
- [4] Huq, N. L., Cross, K. J., Ung, M., Myroforidis, H. *et al.*, A review of the salivary proteome and peptidome and saliva-derived peptide therapeutics. *Int. J. Pept. Res. Ther.* 2007, 13, 547–564.
- [5] Kauffman, D. L., Bennick, A., Blum, M., Keller, P. J., Basic proline-rich proteins from human parotid saliva: relationships of the covalent structures of ten proteins from a single individual. *Biochemistry* 1991, 30, 3351–3356.
- [6] Bennick, A., Salivary proline-rich proteins. *Mol. Cell. Biochem.* 1982, 45, 83–99.
- [7] Messana, I., Cabras, T., Pisano, E., Sanna, M. T. *et al.*, Trafficking and post-secretory events responsible for the formation of secreted human salivary peptides. A proteomic approach. *Mol. Cell. Proteomics* 2008, 7, 911–926.
- [8] Oppenheim, F. G., Hay, D. I., Franzblau, C., Proline-rich proteins from human parotid saliva. I. Isolation and partial characterization. *Biochemistry* 1971, 10, 4233–4238.
- [9] Jonsson, A. P., Griffiths, W. J., Bratt, P., Johansson, I. *et al.*, A novel Ser O-glucuronidation in acidic proline-rich proteins identified by tandem mass spectrometry. *FEBS Lett.* 2000, 475, 131–134.
- [10] Isemura, S., Saitoh, E., Sanada, K., Fractionation and characterization of basic proline-rich peptides of human parotid saliva and the amino acid sequence of proline-rich peptide P-E. *J. Biochem.* 1982, 91, 2067–2075.
- [11] Kauffman, D., Wong, R., Bennick, A., Keller, P., Basic proline-rich proteins from human parotid saliva: complete covalent structure of protein IB-9 and partial structure of protein IB-6, members of a polymorphic pair. *Biochemistry* 1982, 21, 6558–6562.
- [12] Saitoh, E., Isemura, S., Sanada, K., Further fractionation of basic proline-rich peptides from human parotid saliva and complete amino acid sequence of basic proline-rich peptide P-H. *J. Biochem.* 1983, 94, 1991–1999.
- [13] Saitoh, E., Isemura, S., Sanada, K., Complete amino acid sequence of a basic proline-rich peptide, P-F, from human parotid saliva. *J. Biochem.* 1983, 93, 883–888.
- [14] Saitoh, E., Isemura, S., Sanada, K., Complete amino acid sequence of a basic proline-rich peptide, P-D, from human parotid saliva. *J. Biochem.* 1983, 93, 495–502.
- [15] Kauffman, D., Hofmann, T., Bennick, A., Keller, P., Basic proline-rich proteins from human parotid saliva: complete covalent structures of proteins IB-1 and IB-6. *Biochemistry* 1986, 25, 2387–2392.
- [16] Lyons, K. M., Azen, E. A., Goodman, P. A., Smithies, O., Many protein products from a few loci: assignment of human salivary proline-rich proteins to specific loci. *Genetics* 1988, 120, 255–265.
- [17] Azen, E. A., Amberger, E., Fisher, S., Prakobphol, A., Niece, R. L., PRB1, PRB2, and PRB4 coded polymorphisms among human salivary concanavalin-A binding, II-1, and Po proline-rich proteins. *Am. J. Hum. Genet.* 1996, 58, 143–153.
- [18] Vitorino, R., Barros, A., Caseiro, A., Domingues, P. *et al.*, Towards defining the whole salivary peptidome. *Proteomics Clin. Appl.* 2009, 3, 528–540.
- [19] Stubbs, M., Chan, J., Kwan, A., So, J. *et al.*, Encoding of human basic and glycosylated proline-rich proteins by the PRB gene complex and proteolytic processing of their precursor proteins. *Arch. Oral Biol.* 1998, 43, 753–770.
- [20] Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C. *et al.*, UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 2004, 32, D115–D119.
- [21] Helmerhorst, E. J., Sun, X., Salih, E., Oppenheim, F. G., Identification of Lys-Pro-Gln as a novel cleavage site specificity of saliva-associated proteases. *J. Biol. Chem.* 2008, 283, 19957–19966.
- [22] Messana, I., Inzitari, R., Fanali, C., Cabras, T., Castagnola, M., Facts and artifacts in proteomics of body fluids. What proteomics of saliva is telling us? *J. Sep. Sci.* 2008, 31, 1948–1963.
- [23] Castagnola, M., Cabras, T., Inzitari, R., Zuppi, C. *et al.*, Determination of the post-translational modifications of salivary acidic proline-rich proteins. *Eur. J. Morphol.* 2003, 41, 93–98.
- [24] Hay, D. I., Bennick, A., Schlesinger, D. H., Minaguchi, K. *et al.*, The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f). *Biochem. J.* 1988, 255, 15–21.
- [25] Minaguchi, K., Madapallimattam, G., Bennick, A., The presence and origin of phosphopeptides in human saliva. *Biochem. J.* 1988, 250, 171–177.
- [26] Blom, N., Gammeltoft, S., Brunak, S., Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 1999, 294, 1351–1362.
- [27] David, F. N., Exploratory Data-Analysis – Tukey, Jw. *Biometrics* 1977, 33, 768–768.
- [28] Satomaa, T., Heiskanen, A., Leonardsson, I., Angstrom, J. *et al.*, Analysis of the human cancer glycome identifies a novel group of tumor-associated N-acetylglucosamine glycan antigens. *Cancer Res.* 2009, 69, 5811–5819.