Published in final edited form as:

Anal Biochem. 2010 December 1; 407(1): 19–33. doi:10.1016/j.ab.2010.07.012.

Large-Scale Phosphoproteome of Human Whole Saliva Using Disulphide-Thiol-Interchange Covalent Chromatography and Mass Spectrometry

Erdjan Salih^{*}, Walter L. Siqueira, Eva J. Helmerhorst, and Frank G. Oppenheim

Department of Periodontology and Oral Biology, Henry M. Goldman School of Dental Medicine, Boston University Medical Center, 700 Albany Street, Boston, MA 02118-2648

Abstract

Thus far only a handful of phosphoproteins with important biological functions have been identified and characterized in oral fluids and these include some of the abundant protein constituents of saliva. Whole saliva (WS) samples were trypsin digested followed by chemical derivatization using dithiothreitol (DTT) of the phospho-serine/threonine containing peptides. The DTT-phosphopeptides were enriched by covalent disulphide-thiol-interchange chromatography and analysis by nano-flow LC-ESI-MS/MS. The specificity of DTT chemical derivatization was evaluated separately under different base-catalyzed conditions with NaOH and Ba(OH)₂, blocking cysteine residues by iodoacetamide and enzymatic O-deglycosylation prior to DTT reaction. Further analysis of WS samples which were subjected to either of these conditions provided supporting evidence for phosphoprotein identifications. The combined chemical strategies and mass spectrometric analyses identified 65 phosphoproteins in WS of which 28 were based on two or more peptide identification criteria with high confidence, and 37 were based on a single phosphopeptide identification. Most of the identified proteins, ~80%, were hitherto unknown phosphoprotein components. This study represents the first large-scale documentation of phosphoproteins of WS. The origins and identity of WS phosphoproteome suggest significant implications for both basic science and the development of novel biomarkers/diagnostic tools for both systemic and oral disease states.

Keywords

Phosphoproteomics; mass spectrometry; covalent chromatography; saliva; oral; biomarkers; diagnostics

INTRODUCTION

To date, all of the known salivary secretory phosphoproteins whose biological functions are established represent an abundant group of proteins isolated, purified and characterized by classical protein chemistry approaches. The advent of mass spectrometry having the ability to identify a large number of proteins from very small biological samples without extensive

^{*}Address correspondence to: Dr. Erdjan Salih, 700 Albany Street, Boston, MA 02118-2648 Phone: 617-638-2308; Fax: 617-638-4924; Salih@bu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

purification steps, led to the emergence of the global proteomics era. This, combined with the significant interest in developing saliva-based biomarkers for diagnostics, has fueled the establishment of saliva proteome projects [1–6]. While salivary proteome studies are ongoing their focus is predominantly in defining the global proteome. The phosphoproteome analysis of oral fluids containing both abundant and low abundant phospho- proteins/peptides remains to date a totally uncharted territory. Aside from their involvement in the regulation of various aspects of mineral homeostasis, phosphoproteins of low abundance in oral fluid and those derived from cellular origin have the potential to contribute to the repertoire of novel biomarkers for diagnostics in both systemic and oral diseases.

Most of the biological phosphorylation events are related to critical cellular pathways or cell communication [7–9]. These processes include a multitude of regulatory mechanisms of metabolism, cell division, cell growth, and differentiation. In order to understand the biological role of phosphoproteins, in addition to charting the phosphoproteome, phosphoprotein content and expression, the sites of phosphorylation need to be defined. The most common amino-acid residues modified by phosphorylation are serine (P-Ser, ~90%), and to a lesser extent threonine (P-Thr, ~ 10%) and tyrosine (P-Tyr, ~0.05%) [10]. A classical example of phosphoproteins are those involved in the homeostasis of mineralized tissues such as bone and teeth. In this class of phosphorylated proteins, the phosphate plays an intricate function, and loss of the phosphate leads to loss of function. Many of the bone mineral phosphoproteins and their structure-function relationship have been established in previous studies [11–14]. Saliva, which surrounds teeth, has been shown to contain a variety of phosphoproteins critically important for the maintenance of enamel homeostasis [15–20]. Enamel differs from bone in that, in the mature state after eruption, it has no cells which can spring into action for biological repair. For its repair therefore, enamel is entirely dependent on other organ systems such as the salivary glands to provide a continuous supply of phosphorylated proteins. The biochemistry of the oral fluid dictates the processes occurring in oral health and disease and proteins representing major components of secretions are actually phosphoproteins. The best known and characterized phosphoproteins are the acidic proline-rich proteins (PRPs) [15,16], statherin [20], and histatin 1 [18,19]. Salivary phosphoproteins such as PRPs and statherin are known for their inhibitory effect on primary and/or secondary calcium phosphate (Ca-P) precipitation [21,23]. By being inhibitors of Ca-P precipitation in solution, they prevent Ca-P from precipitation even under supersaturated conditions, hence generating a thermodynamic driving force which drive Ca and P into enamel, providing a mechanism for enamel re-mineralization.

Despite significant progress in the development of methods for selective isolation of phosphopeptides, analyses of comprehensive phosphoproteomes of biological samples remain challenging [24–27]. This is due to their low abundance, and the unique and labile chemistry of phospho-Ser/Thr residues. The latter renders MS/MS analysis and protein identification by collision-induced-dissociation (CID) especially difficult because the phosphate groups become rapidly dissociated during the CID fragmentation step. This leads to complications both in fragmentation patterns for identification of the phosphorylated peptides within a complex mixture of global peptide mass spectra, as well as loss of phosphate precluding the localization of the site of phosphorylation. To overcome these limitations, specific phosphopeptide enrichment methods have been developed and utilized. These include a number of strategies, reagents, and chemistries introduced to enable both qualitative and relative quantitative MS analysis of phospho-proteins/peptides. One general approach uses immobilized metal affinity chromatography (IMAC), which exploits the high affinity of phosphate groups on the phosphopeptide to cations such as Fe^{3+} , Ga^{3+} and titanium oxide (TiO_2) [28,29,30,31] followed by MS analysis. However, this approach does suffer from the above disadvantages since it utilizes the native phosphopeptides for the MS/MS analysis. In addition, there is also the lack of specificity because of its capacity to bind and enrich “non-phosphopeptides” which contain

acidic residues. Nevertheless, IMAC-based approaches have been used in recent times to establish comprehensive phosphoproteome analyses for yeast, *Saccharomyces cerevisiae* [32] and the fruit fly, *Drosophila melanogaster* [25]. An alternative approach is based on strategies to enrich phospho- proteins/peptides by covalent-modifications incorporating affinity tags utilizing the physico-chemical properties of phosphoamino acids. The P-Ser/P-Thr containing proteins/peptides are derivatized under base-catalyzed conditions by thiol agents and studies using both mono- and di-thiol reagents [14,27,33–41]. The current manuscript describes the application of this latter technology to the saliva field to establish the global phosphoproteome of this body fluid.

MATERIALS AND METHODS

Preparation of the Sepharose-4B-Glutathione-2-Pyridyl Disulphide Conjugate

Essential steps of covalent-chromatography gel preparation are summarized in Scheme 1. 10 ml of Sepharose-4B-Glutathione (Amersham Pharmacia Biotech, Piscataway, NJ, U.S.A.) was washed, equilibrated with 0.1 M Tris-HCl buffer, pH 8.0, containing 0.3 M NaCl and 1mM EDTA and reduced by 10 ml of 20 mM dithiothreitol (DTT) in the same buffer for 30 min with gentle stirring. The excess DTT was removed by repeat washing with above buffer containing no DTT. The fully reduced gel was allowed to react with 20 ml of 15 mM 2,2'-dipyridyl disulphide (2PDS) (Sigma-Alrich, Co., St Louis, MO, U.S.A.) in 50% ethanol and 0.1M NaHCO₃ buffer, pH 8.0, with gentle stirring at room temperature overnight. The gel was then washed extensively on a sintered glass funnel until no 2PDS could be detected as judged by measuring absorbance at 281nm [42]. An aliquot of the synthesized Sepharose-4B-glutathione-2-pyridyl disulphide conjugate gel was titrated with mercaptoethanol to determine its thiol-binding capacity, as described previously [42]. Titration and measurement of absorbance at 343 nm for the release of 2-thiopyridone provided disulphide-interchange capacity for free thiol-containing proteins/peptides of 0.3 μ moles/ml of gel.

Human Whole Saliva Collection

Saliva collection protocols were approved by the Institutional Review Board of Boston University Medical Center, and informed consent was obtained from each subject participating in the study. Inclusion criteria included overall systemic health, no current or recent medications, and no impairment in salivary gland function. Whole saliva was collected from five individuals, three females and two males, mean age 27 ± 2 years and pooled. To minimize the effects of circadian rhythm, saliva samples were consistently collected in the morning under masticatory stimulation (8 to 10 a.m.). Also, participants were instructed to not smoke, eat or drink, brush during the 2 hr period before saliva collection. All analyses were carried out with whole saliva pools derived from the same five individuals. For triplicate experiments 3 separate 5 ml WS pools were used and each analyzed in duplicates. For chemical derivatization both NaOH and Ba(OH)₂ base-catalyzed conditions were used for the experimental native saliva proteins. For control experiments assessing the interference of O-glycosylation samples were treated both by after O-deglycosylation and after O-deglycosylation combined with dephosphorylation. In addition, free cysteine residues in tryptic peptide samples were alkylated to establish whether such a treatment will have an impact on the identified phosphoproteins of WSS.

Chemical Derivatization of the Human Whole Saliva (WS) Phosphoserine- and Phosphothreonine-Containing Phosphopeptides

The freshly collected WS samples were centrifuged at 12,000 x g to remove particulate material, host and microbial cells. On average 5 ml of WS sample would have yielded 4–6 mg total protein as determined by modified Lowry's assay using bicinchoninic acid. To ensure and

evaluate the specificity of the base-catalyzed DTT derivatization of the phosphopeptides, a number of different conditions were used as outlined below.

For base-catalyzed DTT-derivatization WS samples buffered in NH_4HCO_3 at a final concentration of 50 mM, pH 8.0, were subjected to TPCK-trypsin, 2% (w/w), (Sigma-Aldrich, Co., St Louis, MO, U.S.A.) treatment and incubated at 37°C overnight. The digests were then treated with 10 mM DTT in the presence of 0.3 M NaOH or 0.1 M $\text{Ba}(\text{OH})_2$ at 50 °C for 1 hr to derivatize the phosphoserine (P-Ser) and phosphothreonine (P-Thr) containing phosphopeptides [14,37,38]. For control experiments aliquots of WS proteins were first O-deglycosylated by incubation with a combination of O-glycosidase and neuraminidase as described previously [14] followed by trypsin digestion and base-catalyzed reaction with DTT. Additional controls comprised of O-deglycosylation and phosphatase treatment prior to trypsin digestion and chemical derivatization (14). This latter treatment also probes for possible interference by presence of sulfonated Ser/Thr residues even though such modifications are rare. To achieve protein dephosphorylation the enzymes used were alkaline phosphatase (from bovine intestinal mucosa, Sigma-Aldrich Co, St. Louis Missouri, U.S.A.) 100 units/mg of sample protein in 20 mM NH_4HCO_3 , pH 8.0. Incubations were carried out for 6–8 hrs at room temperature, followed by the addition of acid phosphatase (from potato, Sigma-Aldrich Co, St. Louis Missouri, U.S.A.) 15 units/mg of sample protein, 0.1M acetate buffer, pH 5.0, and incubation for 16 hr at 37°C. These steps were followed by trypsin digestion and base-catalyzed reaction. For alkylation of cysteine residues aliquots of WS samples were rendered 8 M with respect to urea and reduced by 4 mM DTT in 50 mM NH_4HCO_3 , pH 8.0, at 50°C for 1 hr followed by alkylation of the free cysteine residues by 9 mM iodoacetamide for 1 hr in the dark at room temperature. After dialysis 1 mM CaCl_2 was added, the proteins were trypsin digested, followed by the base-catalyzed reaction.

To remove excess DTT samples were neutralized with diluted HCl followed by addition of $\text{H}_2\text{O}/0.1\%$ (v/v) $\text{F}_3\text{CCO}_2\text{H}$ and subjected to RP-HPLC on a TSK-gel 5 μm , ODS-120T C_{18} 25 X 0.46 cm column (TOSOHaas, Montgomeryville, PA, USA). The two step gradient used for the development of the C_{18} column consisted of buffer A ($\text{H}_2\text{O}/0.1\%$ v/v $\text{F}_3\text{CCO}_2\text{H}$) and buffer B (80% $\text{CH}_3\text{CN}/0.1\%$ v/v $\text{F}_3\text{CCO}_2\text{H}$) at a flow rate of 1 ml/min. The CH_3CN and $\text{F}_3\text{CCO}_2\text{H}$ were removed from the samples using a speed-vac (Savant, Farmingdale, New York, NY).

Covalent Chromatography Using Sepharose-4B-Glutathione-2-Pyridyl Disulphide to Specifically Capture and Enrich Phosphopeptides Through Thiol-Interchange

The tryptic mixture of peptides including DTT-derivatized phosphopeptides without free DTT were suspended in 0.1 M NaHCO_3 buffer, pH 8.0, containing 0.15 M NaCl and 1 mM EDTA. 3–5 mls of the phosphopeptide solution were mixed with 1–2 ml of Sepharose-4B-glutathione-2-pyridyl disulphide gel in a 15 ml Falcon tube and left shaking at room temperature for 16 h. The following day the samples were centrifuged at 4,500 x g and the resin was subjected to extensive repeat washings with the same NaHCO_3 buffer. Each ml of resin was washed 5x with 10 ml. This series of washes removed almost all of the non-phosphopeptides, while the DTT-derivatized phosphopeptides remained covalently bound to the gel. The DTT-derivatized phosphopeptides were eluted by incubation with 2 ml of 5 mM DTT in NH_4HCO_3 buffer, pH 8.0, at 40 °C for 4 hrs. The excess DTT was removed by RP-HPLC as described above and the captured and enriched DTT-phosphopeptides were freeze dried using a speed-vac (Savant, Farmingdale, New York, NY).

Nano-Flow Liquid-Chromatography and Electrospray-Ionization Tandem Mass Spectrometric (LC-ESI-MS/MS) Analysis

LC-MS/MS analyses were carried out using an LTQ-linear ion trap mass spectrometer (Thermo Electron, San Jose, CA). Samples were suspended in 97.4% H_2O : 2.5% CH_3CN :0.1%

HCO₂H and LC-ESI-MS/MS analyses were carried out using an on-line autosampler (Micro AS, ThermoFinnigan, CA) with auto-injections of 3 µl onto an in-line fused silica microcapillary column, (75 µm X 10 cm), packed in-house with C₁₈ resin (Micron Bioresource, Inc. Auburn, CA) and developed at a flow rate of 250 nl/min. Peptides were separated by a 55 min elution comprising of multi-step-linear gradient using solvent A, H₂O/2.5 % CH₃CN/0.1 % HCO₂H, and solvent B, CH₃CN/0.1 % HCO₂H. The gradient steps were from 100 % solvent A to 8 % solvent B in 5 min, to 15 % solvent B in 10 min, to 25 % solvent B in 10 min, to 50 % solvent B in 20 min, and to 95 % solvent B in 10 min using a Surveyor MS Pump Plus (ThermoFinnigan, CA). The eluted peptides were directly nano-electrosprayed and the MS/MS data were generated using data-dependent acquisition with a MS survey scan range between 390 and 2000 m/z. This data-dependent acquisition begins with the LC separation which generates a total ion chromatogram in a survey scan followed by selection of specific ions for collision induced dissociation (CID, MS/MS) in descending order of signal intensity. Each survey scan (MS) was followed by automated sequential selection of five peptides for CID, at 35% normalized collision energy, with dynamic exclusion of the previously selected ions. This process was continuously alternated between MS survey scan and five tandem MS throughout the nano-LC chromatography.

Database Search and Phosphopeptide Identification—All MS/MS spectra from LC-ESI-MS/MS were searched against the human database, Uniprot (Universal Protein Resource, Version 9.0), which combines the data from Swiss-Prot (Version 51), TrEMBL (Version 34) and PIR using Bioworks 3.3.1 software and SEQUEST search engine [43]. The data were searched against 241,242 entries. To determine the “false positive rate” the data were searched against a concatenated human sequence database containing both the forward and the reverse sequence version, kindly provided by Dr Steven Gygi (Harvard Medical School, Boston MA). The false positive rate was calculated by using the number of matches to the reverse database multiplied by 2 and divided by the total number of matches (forward plus reverse), as described by Peng et al. [44]. The DTA files were generated with the following settings: precursor-ion tolerance of 1.5 amu, fragment ion tolerance of 1.0 amu, and automated calculated charged states +1, +2, and +3 which also included 5 point smoothing. The searches were performed with parameters: partial trypsin, 2 miscleavages, modifications of serine and threonine residues by DTT (+136.2 Da; dynamic modification). The 136.2 Da modification is a unique mass addition to a given peptide which was originally phosphorylated. This is as a result of the reaction of DTT (+154.2 Da) with dehydroalanine for P-Ser and dehydroamino butyric acid for P-Thr. The loss of phosphate group as phosphoric acid (−H₃PO₄, −98 Da) includes an 18 Da contribution from the hydroxyl group of Ser/Thr residue during base-catalyzed β-elimination [37,38]. For database search the modification of 136.2 Da is used since human database contains only the primary protein sequences with no posttranslational modifications such as phosphorylation. Hence a peptide which is derivatized by DTT on Ser or Thr residues can be identified in the database by matching the mass of the peptide plus 136.2 Da. This is specific to Ser and Thr residue modifications and excludes automatically during the search identification of peptides with cysteine residues if modified by DTT since mass addition in these cases will be only 120.2 Da. Both full and partial tryptic peptides were used to create the phosphoproteome list in Table 1. The inclusion of partial tryptic peptides was to minimize the exclusion of peptides generated by the oral proteolytic activity known to occur in WS, [1,45–47]. The database search results were filtered using the criteria: ΔCn ≥ 0.1; probability ≤ 0.1; for full tryptic peptides, XCorr ≥ 1.7, 2.0, 3.5 for Z= +1, +2, +3; and for half-tryptic peptides, XCorr ≥ 1.9, 2.2, 3.75 for Z= +1, +2, +3. In addition to the search parameters and criteria used, the identified phosphopeptide sequences and the specific site(s) of phosphorylation were stringently evaluated and assessed manually by examining each of the identified phosphopeptide MS/MS data for the quality and the confidence through the *b* and *y* ion fragment series.

Evaluation of the effects of base treatment with NaOH of a model synthetic cysteine containing peptide before and after carboxyamidomethylation

Synthetic peptide containing a single native cysteine residue with a free thiol group, EAGDDIVPCSMSYTWGTA (American peptide company, Inc., Sunnyvale, CA), was used to determine the effects of base treatment. 5 µg of this peptide was used for each of the following treatments followed by mass spectrometric (MS) analysis: (i) original peptide in H₂O alone, (ii) peptide incubated in 0.3 M NaOH at 50° C for 1 hr, (iii) peptide carboxyamidomethylated by 1 mM iodoacetamide in 50 mM NH₄HCO₃, pH 8, in the dark at room temperature for 1 hr, and (iv) carboxyamidomethylated peptide incubated with 0.3 M NaOH at 50° C for 1 hr. After above treatments those samples with base and salts were cleaned by passage through C-18 micro-spin columns, followed by MS analysis. For these pure peptide samples MS analysis was carried out using “static” sample analysis method, direct electrospray of the sample using Glass Pico-Tip emitters (New Objective, Inc., U.S.A) without liquid chromatography step. The peptide masses were determined by using the raw MS data and “deconvolution” software, ProMass Deconvolution^{TM} version 2.5 (Thermo Scientific, CA).

Protein Annotations and Protein Kinases That are Involved in Specific Phosphorylation

The identified phosphoproteins were classified and assigned by molecular function, biological process and cellular component using three web-based applications: Babelomics database (<http://babelomics.bioinfo.cipf.es/index.html>), AmiGO database (http://amigo.geneontology.org/cgi-bin/amigo/go.cgi?advanced_query=yes) and Swiss protein database (<http://ca.expasy.org/>).

Protein kinases involved in the phosphorylation of each of the phosphoproteins and the specific aminoacid sequence recognition motifs utilized by different classes of protein kinase(s) were defined based on kinase recognition templates described in our previous work [11,12,14,37, 38], reviews by other investigators [48,49] and the web-base application of Human Protein Reference Database (http://www.hprd.org/PhosphoMotif_finder).

RESULTS

Scheme 1 summarizes the work flow for the preparation of the covalent chromatography resin, the steps to derivatize the phosphopeptides, and methods to isolate the derivatized peptides from the overall peptide mixture. As indicated in Scheme 1, removal of excess DTT after the peptide derivatization step is carried out by RP-HPLC. The tryptic peptides (native and derivatized) eluted at 26 min (Fig. 1). After covalent exchange chromatography of the derivatized peptides and their elution from the resin with excess DTT, the peptides were again subjected to RP-HPLC. From the peak areas in the two chromatograms, it could be estimated that the derivatized (phospho) peptides constitute approximately 0.7% of the tryptic peptides (Fig.1). Evidence for the efficient separation of non-phosphorylated peptides was manifested in the loss of these peptides screened by the SEQUEST searches under dynamic modifications. These dynamic modification search conditions revealed identification of a very few non-phosphopeptides.

Figure 2 shows a typical base-peak-ion chromatogram of WS supernatant phosphopeptides obtained by LC-ESI-MS. The elution diagram shows the presence of low, medium and high abundance phosphopeptides eluting between 5 and 60 minutes from the microcapillary C₁₈ column. The effectiveness of experimental approaches to capture phosphorylated peptides was evaluated first using 1 µg (~60 pmole) of a purified preparation of acidic proline-rich-protein-1 (aPRP-1). This salivary phosphoprotein contains two fully and one partially phosphorylated serine residues within the N-terminal 30 residue domain [21,50]. LC-ESI-MS/MS analysis identified succinctly this phosphopeptide and their previously established sites of

phosphorylation, at Ser residues 8 and 22 (fully phosphorylated) and the more recently discovered Ser-17 residue (partly phosphorylated) in the peptide exhibiting the sequence QDLDEDVS*QEDVPLVIS*DGGDS*EQFIDEER.Q. Partially overlapping phosphopeptides were also identified such as E.DVS*QEDVPLVIS*DGGDS*EQFIDEER.Q, L.VIS*DGGDS*EQFIDEER.Q, V.ISDGGDS*EQFIDEER.Q, and P.LVIS*DGGDSEQFIDEER.Q. These results indicated that the approach correctly identified and in this case verified the phosphorylation of salivary phosphoproteins.

The technique was subsequently applied to obtain data on the whole saliva phosphoproteome. Figure 3A shows a typical MS/MS spectrum from a peptide derived from acidic PRPs identifying *b* and *y* ions and the localization of the phosphorylation site (Ser-22) between *y*₈ and *y*₉ specifically highlighted by a signature loss of 223 Da (dehydroalanine-DTT adduct).

Apart from identifying well known phosphorylation, the present work also uncovered novel phosphorylation sites of abundant salivary proteins. For example the highly abundant basic proline rich proteins (bPRPs) were considered to be non-phosphorylated. The data in Figure 3B show for the first time one of the phosphopeptide regions containing two phosphorylation sites within non-glycosylated bPRP2. In addition the N-glycosylated bPRP3 was also found to contain phosphate. This protein contained only a single phosphopeptide region with multiple phosphorylation sites.

Among the smaller salivary phosphoproteins, statherin and histatin-1 phosphorylation sites were also investigated. As expected, Ser-2 and Ser-3 in statherin were found to be phosphorylated. For histatin-1 the phosphorylation at Ser-2 position has been well established. The peptide region containing phosphorylation site at position 2 of this protein was not identified due to the fact that trypsinization of histatin 1 generates an N-terminal peptide containing only 5 a.a. and a *M_r* of ~500 Da. This type of peptide with multiple charged states is below the typical operating survey scan range. We found a novel phosphorylation site at Ser-32. The MS/MS data of this identification is shown in Fig. 4A. It was surprising that no phosphopeptides were found of the cystatin family which is known to exhibit heterogeneous phosphorylation in both N-terminal and C-terminal domains. The N-terminal phosphopeptides of cystatins could have escaped identification due to the presence of phosphorylated serines in a very short N-terminal tryptic peptide. Any C-terminally located phosphopeptides are in close vicinity to any of the four cysteine residues, and those peptides were excluded.

Aside from the small number of well known phosphoproteins in WS, we have established the presence of a large number of phosphoproteins which were not known to be present in this oral fluid. Overall, using the covalent chromatography enrichment strategies, we have identified 65 phosphoproteins and their specific phosphorylation sites of which 28 were identified by 2 or more peptide identification criteria and summarized in Table 1. Additional 37 phosphoproteins were identified based on a single phosphopeptide identification and are reported as the “supplemental data”. It is important to note that there has been reports of the possibility that non-phosphorylated Ser/Thr residues may derivatize by DTT under base catalyzed conditions to a very small extent <1% relative to corresponding phosphorylated forms. However, in the present work any of the phosphorylation sites we found are reported only when they are present relative to their being absent in the control experiments.

A commonly applied requirement for *bona fide* identifications of proteins in proteomics studies is the presence of at least two unique peptides [51,52]. This criterion would also be applicable to phosphoprotein identification if the enrichment was at the protein level. However, if the enrichment is at the phosphopeptide level then it is not as easy to apply the same criteria since most of the phosphoproteins contain only a single phosphorylation site or contain more than one phosphorylation site within the same tryptic peptide. In this latter case protein identification

is restricted to the capturing and characterization of a single peptide. The “false positive rate” for specific filtering parameters was established by searching the raw data against a concatenated human database containing both the forward and the reverse sequences. Each of the raw data files generated were subjected to such searches. Use of the filtering criteria chosen we obtained a “false positive rate” of less than 2% for the 28 phosphoproteins identified based on two or more peptides, Table 1, and a 6% “false positive rate” for the 37 phosphoproteins identified by a single phosphopeptide, supplemental data. To further verify that the identified phosphoproteins were indeed present and part of the WS proteome, additional confirmation was obtained by considering the non-phosphorylated peptides derived from these proteins. Among the 65 phosphoproteins identified, 28 were confirmed with additional presence of non-phosphorylated peptides (Table 2), providing additional support for the phosphoprotein identifications made. Most of these are of the medial to high abundant phosphoproteins indicating that the low level phosphoproteins require phosphopeptide enrichment.

In order to provide additional confidence to the identified phosphoproteins in this study, it was of interest to investigate the effects of base-catalyzed conditions on the cysteine containing peptides. For this purpose a model synthetic cysteine containing peptide, $M_r = 1903$ Da, EAGDDIVPCSMSYTWGA, was used. The MS analysis of the peptide without any treatment provided the expected $M_r = 1903$ and its mono- Na^+ adduct 1925 ($1903 + 23 - \text{H}^+$), Fig. 5A. This peptide after incubation with 0.3 M NaOH for 1 hr at 50°C , showed no evidence of loss of mass or $-\text{SH}$ group, but rather addition of NaOH led to multiple Na^+ adducts, original peptide $M_r = 1903$ Da, and mono- ($M_r = 1925$ Da), di- ($M_r = 1948$ Da), tri- ($M_r = 1970$ Da) and tetra- ($M_r = 1992$ Da) sodium adducts of the peptide, Fig. 5B. Carboxyamidomethylation of the peptide using iodoacetamide showed complete alkylated form with $M_r = 1960$ Da and its mono- Na^+ -adduct $M_r = 1982$ Da, Fig. 5C. However, MS analysis of the carboxyamidomethylated peptide after incubation with 0.3 M NaOH for 1 hr at 50°C showed loss of mercaptoacetamide $\text{HS-CH}_2\text{CO.NH}_2$ (-92 Da), peptide $M_r = 1868$ and its mono- (1892.5 Da), di- (1913.5 Da) and tri- (1936 Da) Na^+ adducts, Fig. 5D. Interestingly, loss of the $-\text{SH}$ group also accompanied with one less Na^+ adduct, indicating that one of the Na^+ adducts was with the thiol group of the cysteine and the others are with the carboxy groups of Glu, and two Asp residues.

DISCUSSION

Mass spectrometry (MS) has become the technique of choice for large-scale proteome analysis because of its sensitivity and high-throughput [5,25,32,53,54], whereas large scale phosphoproteome analysis is much more challenging and therefore only a few studies in this area have been reported. Application of MS to salivary proteomic studies has provided significant contributions to our knowledge of the overall protein composition [1–6]. While these studies have focused on the global proteome analysis, the present study adds not only another dimension to salivary proteomics by documenting for the first time the large-scale phosphoproteome of WS but also provides the basis for additional functional characterization of salivary proteins. We have previously employed DTT and its analogues for identifying and determining state and sites of phosphorylation by qualitative [37,38], relative quantitative [14,55] and absolute quantitative [14,37] approaches. Our study with human WS indicate that a large majority of the identified phosphoproteins, $\sim 80\%$, in human WS were novel and constituted medium to low level abundance phosphoproteins, Table 1. These were neither previously reported in any of the general saliva proteomics publications [1–4,6,56,57] nor listed under the very recently generated databases for salivary components (<http://fields.scripps.edu/public/project/saliva/>, <http://www.hspp.ucla.edu/cgi-bin/display.cgi?page=news>).

The most extensive proteomics studies were carried out so far have been restricted to pure exocrine secretions derived from parotid and submandibular/sublingual glands. In difference

to glandular secretions, WS contains many additional components which are contributed by minor glands, gingival crevicular fluid, and oral epithelial cells. These additional protein sources are likely to add many components not readily detected in pure glandular secretions. The identified number of phosphoproteins, 65 (28 based on two or more, Table 1, and 37 based on a single phosphopeptide, supplemental Table data), in WS exceeds the number of the well known salivary phosphoproteins present in glandular secretions. The notion that the identified phosphoproteins in the present work includes low level phosphoproteins derive from the fact that in biological systems typically protein kinases, cellular phosphoproteins, cytokines, growth factors, transcription factors, cell receptors and hydrolases all exist at very low levels. An example in this category is the vaccinia-related kinase, commonly known as Ser/Thr-VHK2 kinase, based on the peptide comprising residues D.FTSPDIFKKS*R.S, (Fig. 4B) which is the newest class of kinases exhibiting a low level occurrence. About 20% of the identified WS phosphoproteins had already been reported but interestingly, in some of these proteins, additional phosphorylation sites were found.

Some limitations and complexities of analyzing the native phospho-Ser/Thr peptides on a large-scale for phosphoproteomics by MS/MS approaches without chemical derivatization have been highlighted recently [58,59]. The chemical derivatization approach used in the present work has several clear advantages. Firstly, phosphopeptides are highly enriched with increased selectivity by virtue of the covalent-chromatography step. Secondly, the DTT-phosphoserine-derivatives are stable to CID during MS/MS analysis. Thirdly, the MS/MS spectra provides non-ambiguous results for phosphate localization and peptide identifications since they are not complicated by loss of phosphate groups. Fourthly, the modification of +136.2 Da by DTT-derivatization is a unique differential mass addition in the identification of the peptide amino-acid sequence during the database search which originally contained the phosphate group and pinpoints unequivocally the precise site(s) of phosphorylation. Furthermore, the +136.2 Da modification on S/T residues is diagnostic for the phosphopeptide since covalent-chromatography can also capture free cysteine containing peptides, because of their free thiol group reaction with the chemistry of the solid support. Such peptides were used to provide additional evidence for the *bona fide* identification of some of the phosphoproteins. Table 2 shows that among the 65 phosphoproteins found, seven were identified on the basis of one DTT-derivatized phosphopeptide and an additional native cysteine containing peptide. It is noteworthy that inherently the identification of phosphoproteins is restricted to only S/T-DTT derivatized peptides with specific modification of 136.2 Da during the database search. Any peptide containing cysteine residue (without P-Ser/P-Thr) that may have undergone DTT-derivatization by loss of -SH or the loss of carboxyamidomethyl-SH groups under base-catalyzed conditions can not be mis-identified as phosphopeptide since they will not be identified at all during the database search using 136.2 Da mass addition. This is because the mass addition by DTT derivatization in those cases where -SH was lost is distinctly different and would be 120.2 Da. Assessment of a model cysteine containing peptide, EAGDDIVPCSMSYTWGA, and treatment of such peptide with NaOH under conditions similar to those used in the present work showed no evidence for the conversion of the cysteine residue to dehydroalanine, Fig. 5B. This is consistent with the enrichment and elution of seven cysteine peptides in their native states during the covalent-chromatography step of DTT-derivatized phosphopeptides of WSS, Table 2. Figure 6 shows MS/MS data for two such native cysteine containing peptides with no phosphorylation and classic loss of 103 Da during CID reflecting native cysteine residue. Interestingly, when the model cysteine containing peptide was carboxyamidomethylated, Fig. 5C, and subjected to NaOH treatment, there was a loss of mercaptoacetamide leading to formation of dehydroalanine from alkylated cysteine, Fig. 5D. However, as detailed above such peptides would not be identified as phosphopeptides since the search algorithm uses 136.2 Da mass addition and not 120.2 Da. It has also been reported that under base-catalyzed conditions a small percentage of O-glycosylation sites may also undergo conversion to dehydroalanine which can react with DTT. While this reaction can

theoretically occur, it has been established that the rate of this conversion is only ~1% that of P-Ser/P-Thr in NaOH [60] and therefore this would hardly affect the outcome of the studies as presented. To verify our results we have also acquired data using Ba(OH)₂ instead of NaOH base-catalyzed conditions [61]. In addition we enzymatically deglycosylated O-linked carbohydrates prior to DTT-derivatization and MS/MS analysis. All these additional control experiments provided essentially the same results depicted in Table 1 and Supplemental data. Another posttranslational modification to be taken into account in the derivatization approach is O- sulfonation. The most common and well known protein sulfonations occur on hydroxyl groups of tyrosine residues [62,63]. While the possibility of sulphate groups occupying the Ser/Thr residues have to be considered, to date there is a single recent report with only three proteins with Ser/Thr sulphonation [64]. Our experimental procedure of combined O-deglycosylation and dephosphorylation of WS prior to DTT treatment led to no identification of DTT-derivatized proteins indicating that the results in Table 1 and Supplemental data are not complicated by the presence of sulphonation. Furthermore, this approach also demonstrated lack of conversion of Ser/Thr residues to DTT reactive form when they are not posttranslationally modified.

Almost all of the identified phosphopeptides, with the exception of a few, contained amino-acids flanking the phosphorylated residue(s) with consensus sequences that reflect classically known recognition a.a. sequence(s) to one or more of the well known protein kinase families. These included most frequently CKII, CKI, cGMP-dependent kinase and PKC and occasionally cyclin- and cAMP-dependent kinases (Table 1). Often, phosphoproteins with or without multiple phosphorylation sites show a heterogeneous state of phosphorylation, whereby the potential phosphorylation site may not be phosphorylated on every molecule of a given protein leading to an overall less than stoichiometric phosphate content. Such phosphorylation heterogeneity has been well described for extracellular matrix phosphoprotein [11–14]. Examples of heterogeneous state of phosphorylation were clearly found among the identified phosphoproteins in the present work.

Although it is not possible to elaborate on the biological significance of each of the identified phosphoproteins, certain phosphoproteins such as tuftelin, also known as enamelin, [65] is a surprising finding in the WS phosphoproteome. Originally tuftelin was discovered as an acidic protein synthesized by ameloblasts and functionally involved at the very early stages of tooth enamel development and biomineralization [65]. WS tuftelin can not be related to enamel biomineralization since enamel after eruption contains no tuftelin and therefore must be derived from other cellular sources. Interestingly phosphorylation of tuftelin was only predicted based on the presence of CKII and cGMP consensus sequences regions but not verified experimentally. The current work actually confirmed the presence of phosphate and revealed a total of three phosphorylation sites consistent with earlier predictions [65]. Since its first discovery in ameloblasts, tuftelin has been localized in a wide range of other tissues but in almost all cases it has been related to carcinogenesis. This may be of diagnostic importance reflecting potential origins from oral or even non-oral malignancies. Overall, the phosphoproteome of WS displayed an unexpected diversity with 65 phosphoproteins considering that there were less than ten phosphoproteins indigenous to human salivary gland secretions defined by classical biochemical approaches. The amazing diversity of the WS phosphoproteome is not only related in their multiple cellular origins (Fig. 7A) but also their overall organ distribution based on predictions and some observations (Fig. 7B). The phosphoprotein composition of WS as defined in this study suggests potential to reflect a wide variety of oral and systemic health/disease states. Hence, the present data is not limited to the MS-based technologies, but also opens novel avenues to other high-throughput technologies such as microsensors/microfluidics to expand the scope of target proteins useful for diagnostics. For example presence of a phosphorylated well known ovarian cancer marker/ antigen (also known as antigen CA125 or MUC-16) in WS defined in our current work may

provide such an opportunity. Clearly, quantitative studies will be necessary to utilize this new knowledge on the WS phosphoproteome for the purpose of monitoring health versus disease status. Furthermore, while some salivary phosphoproteins have been already defined to play important functions in host defense mechanisms and mineral homeostasis, the present work forms the basis to explore and expand such functional studies in more detail.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Steven Gygi at Harvard Medical School for providing us with the concatenated human protein sequence database containing both the forward and reverse sequences. This work was supported by grants from NIDCR: DE 018448; DE 05672; DE 07652; DE 018132.

Abbreviations used

WS	whole saliva
ECM	extracellular matrix
aPRP	acidic-proline-rich protein
bPRP	basic-proline-rich protein
CID	collision-induced dissociation
IMAC	immobilized-metal-affinity chromatography
ICAT	isotope-coded-affinity-tag
DTT	dithiothreitol
LC-ESI-MS/MS	liquid chromatography electrospray-ionization-tandem-mass spectrometry
2PDS	2,2'-dipyridyl disulphide
CH ₃ CN	acetonitrile
RP-HPLC	reverse-phase high-performance-liquid-chromatography
CKII	casein kinase II
CKI	casein kinase I
PKC	protein kinase C
cGMP	cyclic-guanidine-mono-phosphate
cAMP	cyclic-adenosine-mono-phosphate
TPCK	tosyl-phenyl-chloroketone

References

1. Hardt M, Thomas LR, Dixon SE, Newport G, et al. m Toward defining the human parotid gland salivary proteome and peptidome: identification and characterization using 2D SDS-PAGE, Ultrafiltration, HPLC, and mass spectrometry. *Biochemistry* 2005;4:2885–2899. [PubMed: 15723531]
2. Hu S, Xie Y, Ramachandran P, Ogorzalek Loo RR, et al. Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry. *Proteomics* 2005;5:1714–1728. [PubMed: 15800970]

3. Wilmarth PA, Riviere MA, Rustvold DL, Lauten JD, et al. Two-dimensional liquid chromatography study of the human whole saliva proteome. *J Proteome Res* 2004;3:1017–1023. [PubMed: 15473691]
4. Wong DT. Towards a simple, saliva-based test for the detection of oral cancer 'oral fluid (saliva), which is the mirror of the body, is a perfect medium to be explored for health and disease surveillance'. *J California Dentl Assoc* 2006;34:283–285.
5. Yates JR 3rd. Mass spectral analysis in proteomics. *Annual Rev Biophys and Biomo Struct* 2004;33:297–216.
6. Denny P, Hagen FK, Hardt M, Liao L, et al. The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. *J Proteome Res* 2008;7:1994–2006. [PubMed: 18361515]
7. Hubbard MJ, Cohen P. On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem Sciences* 1993;18:172–177.
8. Qi M, Elion EA. MAP kinase pathways. MAP kinase pathways. *J Cell Sci* 2005;118:3569–3572. [PubMed: 16105880]
9. Hunter T. Signaling--2000 and beyond. *Cell* 2000;100:113–127. [PubMed: 10647936]
10. Hunter J, Hirst BH, Simmons NL. Epithelial secretion of vinblastine by human intestinal adenocarcinoma cell (HCT-8 and T84) layers expressing P-glycoprotein. *British J Cancer* 1991;64:437–444.
11. Salih E, Ashkar S, Gerstenfeld LC, Glimcher MJ. Identification of the phosphorylated sites of metabolically ^{32}P -labeled osteopontin from cultured chicken osteoblasts. Identification of the phosphorylated sites of metabolically ^{32}P -labeled osteopontin from cultured chicken osteoblasts. *J Biol Chem* 1997;272:13966–13973. [PubMed: 9153260]
12. Salih E, Wang J, Mah J, Fluckiger R. Natural variation in the extent of phosphorylation of bone phosphoproteins as a function of in vivo new bone formation induced by demineralized bone matrix in soft tissue and bony environments. *Biochem J* 2002;364:465–474. [PubMed: 12023890]
13. Salih E, Zhou HY, Glimcher MJ. Phosphorylation of purified bovine bone sialoprotein and osteopontin by protein kinases. *J Biol Chem* 1996;271:16897–16905. [PubMed: 8663267]
14. Salih E, Fluckiger R. Complete topographical distribution of both the in vivo and in vitro phosphorylation sites of bone sialoprotein and their biological implications. *J Biol Chem* 2004;279:19808–19815. [PubMed: 15004024]
15. Bennick A, Connell GE. Purification and partial characterization of four proteins from human parotid saliva. *Biochem J* 1971;123:455–464. [PubMed: 5001593]
16. Oppenheim FG, Hay DI, Franzblau C. Proline-rich proteins from human parotid saliva. I. Isolation and partial characterization. *Biochemistry* 1971;10:4233–4238. [PubMed: 4331333]
17. Oppenheim FG, Salih E, Siqueira WL, Zhang W, Helmerhorst EJ. Salivary proteome and its genetic polymorphisms. *Annals NY Acad Sci* 2007;1098:22–50.
18. Oppenheim FG, Xu T, McMillian FM, Levitz SM, et al. Histatins, a novel family of histidine-rich proteins in human parotid secretion. Isolation, characterization, primary structure, and fungistatic effects on *Candida albicans*. *J Biol Chem* 1988;263:472–477. [PubMed: 2961747]
19. Oppenheim FG, Yang YC, Diamond RD, Hyslop D, et al. The primary structure and functional characterization of the neutral histidine-rich polypeptide from human parotid secretion. *J Biol Chem* 1986;261:1177–1182. [PubMed: 3944083]
20. Schlesinger DH, Hay DI. Complete covalent structure of statherin, a tyrosine-rich acidic peptide which inhibits calcium phosphate precipitation from human parotid saliva. *J Biol Chem* 1977;252:1689–1695. [PubMed: 838735]
21. Hay DI, Carlson ER, Schluckebier SK, Moreno EC, Schlesinger DH. Inhibition of calcium phosphate precipitation by human salivary acidic proline-rich proteins: structure-activity relationships. *Calcif Tissue Internatl* 1987;40:126–132.
22. Schlesinger DH, Buku A, Wyssbrod HR, Hay DI. Chemical synthesis of phosphoserine-phosphoserine, a partial analogue of human salivary statherin, a protein inhibitor of calcium phosphate precipitation in human saliva. *Intertl J Pep and Prot Res* 1987;30:257–262.
23. Schwartz SS, Hay DI, Schluckebier SK. Inhibition of calcium phosphate precipitation by human salivary statherin: structure-activity relationships. *Calcif Tissue Intertl* 1992;50:511–517.
24. Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *Chem Reviews* 2001;101:269–295.

25. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nature Methods* 2007;4:231–237. [PubMed: 17293869]
26. Reinders J, Sickmann A. State-of-the-art in phosphoproteomics. *Proteomics* 2005;5:4052–4061. [PubMed: 16196093]
27. Zhou H, Watts JD, Aebersold R. A systematic approach to the analysis of protein phosphorylation. *Nature Biotech* 2001;19:375–378.
28. Andersson L, Porath J. Isolation of phosphoproteins by immobilized metal (Fe³⁺) affinity chromatography. *Anal Biochem* 1986;154:250–254. [PubMed: 3085541]
29. Posewitz MC, Tempst P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem* 1999;71:2883–2892. [PubMed: 10424175]
30. Larsen MR, Graham ME, Robinson PJ, Roepstorff P. Improved detection of hydrophilic phosphopeptides using graphite powder microcolumns and mass spectrometry: evidence for in vivo doubly phosphorylated dynamin I and dynamin III. *Mol Cell Proteomics* 2004;3:456–465. [PubMed: 14762214]
31. Chicz, RM.; Regnier, FE. Guide to protein purification. Vol. 182. San Diego: Academic Press; Methods of Enzymology; p. 392–321.
32. Li X, Gerber SA, Rudner AD, Beausoleil SA, et al. *J Proteome Res* 2007;6:1190–1197. [PubMed: 17330950]
33. Amoresano A, Marino G, Cirulli C, Quemeneur E. Mapping phosphorylation sites: a new strategy based on the use of isotopically labelled DTT and mass spectrometry. *Europ J Mass Spect (Chichester, England)* 2004;10:401–412.
34. Goshe MB, Veenstra TD, Panisko EA, Conrads TP, et al. Phosphoprotein isotope-coded affinity tags: application to the enrichment and identification of low-abundance phosphoproteins. *Anal Chem* 2002;74:607–616. [PubMed: 11838682]
35. Knights CD, Liu Y, Appella E, Kulesz-Martin M. Phosphospecific proteolysis for mapping sites of protein phosphorylation. *J Biol Chem* 2003;278:52890–52900. [PubMed: 14555661]
36. McLachlin DT, Chait BT. Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal Chem* 2003;75:6826–6836. [PubMed: 14670042]
37. Salih E. In vivo and in vitro phosphorylation regions of bone sialoprotein. *Connect Tissue Res* 2003;44 (Suppl 1):223–229. [PubMed: 12952202]
38. Salih E. Synthesis of a radioactive thiol reagent, 1-S[³H]carboxymethyl-dithiothreitol: identification of the phosphorylation sites by N-terminal peptide sequencing and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Biochem* 2003;31:143–158. [PubMed: 12842118]
39. Vosseller K, Hansen KC, Chalkley RJ, Trinidad JC, et al. Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* 2005;5:388–398. [PubMed: 15648052]
40. Weckwerth W, Willmitzer L, Fiehn O. Comparative quantification and identification of phosphoproteins using stable isotope labeling and liquid chromatography/mass spectrometry. *Rapid Comm Mass Spect* 2000;14:1677–1681.
41. Salih E. Phosphoproteomics by mass spectrometry and classical protein chemistry approaches. *Mass Spectrom Revs* 2005;24:828–846. [PubMed: 15538747]
42. Brocklehurst K, Carlsson J, Kierstan MP, Crook EM. Covalent chromatography. Preparation of fully active papain from dried papaya latex. *Biochem J* 1973;133:573–584. [PubMed: 4733241]
43. Eng J, McCormack A, Yates JR 3rd. An approach to correlate tandem mass spectral data of peptide with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994;5:976–989.
44. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2003;2:43–50. [PubMed: 12643542]
45. Baum BJ, Bird JL, Millar DB, Longton RW. Studies on histidine-rich polypeptides from human parotid saliva. *Arch Biochem and Biophys* 1976;177:427–436. [PubMed: 1015827]
46. Helmerhorst EJ, Alagl AS, Siqueira WL, Oppenheim FG. Oral fluid proteolytic effects on histatin 5 structure and function. *Arch Oral Biol* 2006;51:1061–1070. [PubMed: 16901460]

47. Payne JB, Iacono VJ, Crawford IT, Lepre BM, et al. Selective effects of histidine-rich polypeptides on the aggregation and viability of *Streptococcus mutans* and *Streptococcus sanguis*. *Oral Microbiol and Immunol* 1991;6:169–176. [PubMed: 1945500]
48. Kemp BE, Pearson RB. Intracellular regulation of protein kinases and phosphatases. *Biochim Biophys Acta* 1991;1094:67–76. [PubMed: 1653024]
49. Pearson, RB.; Kemp, BE. *Methods Enzymology*. Academic Press; 1991. Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations; p. 62-81.
50. Inzitari R, Cabras T, Onnis G, Olmi C. Different isoforms and post-translational modifications of human salivary acidic proline-rich proteins. *Proteomics* 2005;5:805–815. [PubMed: 15693058]
51. Orchard S, Hermjakob H, Apweiler R. Annotating the human proteome. *Mol & Cell Proteomics* 2005;4:435–440.
52. Wilkins MR, Appel RD, Van Eyk JE, Chung MC. Guidelines for the next 10 years of proteomics. *J Proteomics* 2006;6:4–8.
53. Lu B, Ruse C, Xu T, Park SK, Yates J 3rd. Automatic validation of phosphopeptide identifications from tandem mass spectra. *Anal Chem* 2007;79:1301–1310. [PubMed: 17297928]
54. Zhang H, Yan W, Aebersold R. Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr Opin Chem Biol* 2004;8:66–75.
55. Salih, E. Emergence of phosphoproteomics through combination of mass spectrometry and classical protein chemistry. In: Landis, WL.; Sodek, J., editors. *The Chemistry & Biology of Miner Tissues*. Toronto University Press; Canada: 2004. p. 208-211.
56. Xie H, Rhodus NL, Griffin RJ, Carlis JV, Griffin TJ. A catalogue of human saliva proteins identified by free flow electrophoresis-based peptide separation and tandem mass spectrometry. *Mol Cell Proteomics* 2005;4:1826–1830. [PubMed: 16103422]
57. Siqueira WL, Salih E, Helmerhorst EJ, Oppenheim FG. Proteome of human minor salivary gland secretion. *J Dent Res* 2008;87:445–450. [PubMed: 18434574]
58. Villen J, Beauvoile SA, Gygi SP. Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis. *Proteomics* 2008;8:4444–4452. [PubMed: 18972524]
59. Boersema PJ, Mohammed S, Heck AJR. Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* 2009;44:861–878. [PubMed: 19504542]
60. Kall L, Storey JD, MacCoss MJ, Noble WS. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res* 2008;7:29–34. [PubMed: 18067246]
61. Byford MF. Rapid and selective modification of phosphoserine residues catalyzed by Ba²⁺ ions for their detection during peptide microsequencing. *Biochem J* 1991;280:261–265. [PubMed: 1741751]
62. Krishna, R.; Wold, F. In *proteins analysis and design*. Ageletti, RH., editor. San Diego: Academic press; 1998. p. 121-206.
63. Cabras T, Fanali C, Monteiro JA, Amado F, Inzitari R, Desiderio C, Scarano E, Giardino B, Castagnola M, Messana I. Tyrosine polysulfonation of human salivary histatin 1. A posttranslational modification specific of the submandibular gland. *J Proteome Res* 2007;6:2472–2480. [PubMed: 17503797]
64. Medzihradszky KF, Darula Z, Perlson E, Fainzilber M. O-sulfonation of serine and threonine: mass spectrometric detection and characterization of a new posttranslational modification in diverse proteins throughout the eukaryotes. *Mol Cell Proteomics* 2004;3:429–440. [PubMed: 14752058]
65. Mao Z, Shay B, Hekmati M, Fermon E. The human tuftelin gene: cloning and characterization. *Gene* 2001;279:181–196. [PubMed: 11733143]

Figure 1. Removal of excess dithiothreitol (DTT) by RP-HPLC

A 25 X 0.46 cm C₁₈ column (TOSOH, Montgomeryville, PA, USA) was eluted with a flow rate of 1 ml/min and the eluate monitored by following absorbance at 219 nm. The CH₃CN gradient used is indicated by a solid line. (A) Elution of all tryptic peptides including DTT-derivatized phosphopeptides at 26 min. (B) Elution of DTT-derivatized phosphopeptides after covalent-chromatography at 26 min. Note that the peak area in (B) represents only 0.7 % that of the peak area in (A).

Figure 2. Nanoflow liquid chromatography electrospray-ionization tandem-mass spectrometry (LC-ESI-MS/MS)

Base-peak ion-chromatogram of dithiothreitol (DDT)-phosphopeptides of WS. Note the existence of many phosphopeptides present at low, medium and high abundance.

Figure 3. MS/MS sequence analyses of covalent chromatography enriched DTT-derivatized phosphopeptides of aPRP and bPRP from whole saliva supernatant

A: MS/MS spectra of one of the phosphopeptides captured from the tryptic digest of WS. This phosphopeptide derived from the WS was found to exhibit part of the N-terminal phosphopeptide of acidic PRPs. The series of *b* and *y* fragmentation ions are labeled and the precise site of phosphorylation (Ser-22, indicated by an asterisk) indicated by a loss of 223 Da (dehydroalanine-DTT adduct) between *y*₉ and *y*₈ which is a signature for phosphoserine derivatized by DTT. **B:** MS/MS spectrum of one of the bPRP-2 phosphopeptides. The series of *b* and *y* fragmentation ions are labeled and the precise sites of phosphorylation (Ser-136 and Ser-138) are indicated by asterisks within the identified sequence.

Figure 4. MS/MS sequence analyses of covalent-chromatography enriched DTT-derivatized phosphopeptide of Histatin-1 and Ser/Thr-VHK2 kinase from whole saliva supernatant tryptic digest

A: MS/MS spectrum of Histatin-1 phosphopeptide. The series of b and y fragmentation ions are labeled and the precise site of phosphorylation (Ser-32) indicated by a loss of 223 Da between y_2 and y_3 . An asterisk on S following the residue in the amino-acid sequence denotes the phosphoserine residue. **B:** MS/MS spectrum of Ser/Thr-VHK2 kinase phosphopeptide captured from the tryptic digest of WS after DTT-derivatization and covalent-chromatography. The series of b and y fragmentation ions are labeled and the precise site of phosphorylation. An asterisk following the residue denotes the phosphoamino-acid residue.

Figure 5. MS analysis of a model synthetic cysteine containing peptide, EAGDDIVPCSMSTWTGA

(A): MS analysis of the original peptide, $M_r = 1903$ and its mono- Na^+ adduct 1925 ($1903 + 23 - \text{H}^+$). (B): MS analysis of the original peptide after incubation with 0.3 M NaOH for 1 hr at 50°C , with no evidence of loss of mass or $-\text{S}$ group, but rather addition of NaOH led to multiple Na^+ adducts, original peptide mass $M_r = 1903$ Da, and mono- ($M_r = 1925$ Da), di- ($M_r = 1948$ Da), tri- ($M_r = 1970$ Da) and tetra- ($M_r = 1992$ Da) sodium adducts of the peptide. (C): MS analysis of the peptide after carboxyamidomethylation (CAM) using iodoacetamide showing complete alkylated form with $M_r = 1960$ Da and its mono- Na^+ -adduct $M_r = 1982$ Da. (D): MS analysis of the carboxyamidomethylated peptide after incubation with 0.3 M NaOH for 1 hr at 50°C showing loss of mercapto-acetamide $\text{HS-CH}_3\text{CO.NH}_2$ (-92 Da), $M_r = 1868$ and its mono- (1892.5 Da) di- (1913.5 Da) and tri- (1936 Da) Na^+ adducts.

Figure 7. Cellular locality and tissue distribution of the identified phosphoproteins

A: Categorization of the identified phosphoproteins based on their classification as present in specific cellular compartments/organelles, extracellular domains. **B:** Tissue distribution of the identified phosphoproteins using phosphoprotein annotations.

Scheme 1.

Flowchart of the chemical steps and reaction pathways for generating a covalent-thiol-interchange solid-phase support and the derivatization of phospho-serine/threonine containing peptides by DTT used for the capture and enrichment of phosphopeptides.

Table 1

The phosphopeptome of human whole saliva supernatant

Accession number	Protein name	Phosphopeptides	Z	XCorr	Protein kinase(s) involved	Molecular function/ Biological process
P02810	Salivary acidic proline-rich phosphoprotein 1/2	I.S*DGGDS*EQFIDEER.Q	2	4.49	CKII/β-adrenergic receptor kinase	G-protein coupled receptor activity, regulation of calcium phosphate precipitation, G-protein coupled receptor protein signaling pathway
		L.VIS*DGGDS*EQFIDEER.Q	2	5.18		
		S.DGGDS*EQFIDEER.Q	2	3.89		
		V.IS*DGGDSEQFIDEER.Q	2	3.50		
		V.ISDGGDS*EQFIDEER.Q	2	5.18		
P1961	Carcinoid alpha-amylase (α-2B)	E.DSVQEDVPLVIS*DG GDS*EQFIDEER.Q	2	4.17	unknown	protein binding, hydrolase, Carbohydrate hydrolysis
		R.IYVDAVINHMS*GN.A	2	3.01		
P04745	Salivary alpha-amylase	N.WEPAGS*KPFYQEVLDLGEPIK.S	2	2.9	CKII / PKC / DNA-dependent protein/CKI	protein binding, hydrolase, Carbohydrate hydrolysis
		K.AHFSIS*NS*AEDPFIH.A	2	4.95		
		K.NWGEWGFMP*DR.A	2	3.40		
		N.SAEDPFIHAE*KL- ,	2	3.19		
		R.TS*IVHLFEWR.W	2	3.52		
		K.IYVS*DDGK.A	2	2.9		
Q99504	Eye absent homolog 3	K.PES*GLIQT*PSPSQH.S	2	2.53	CKI/CKII	Activator, developmental protein, hydrolase, Transcription regulator
P08123	Collagen alpha-2(I)chain	P.PGPAGSRGDDGPPGDMT*GFPGAAGR.T	2	2.85	unknown	Extracellular matrix structural constituent, Odontogenesis, skin morphogenesis, blood vessel development
		.FPGARGPS*GPQGGGPPGPK.	2	2.2		
Q01955	Collagen alpha-3(IV) chain	R.GPPGSRGS*PGAPGPPGPPGSH.V	3	3.60	unknown	Extracellular structural protein
		K.IIS*LPGSPGPPGTPGEP.G	2	2.21	unknown	
Q5T2V0	Rho guanine nucleotide exchange factor (GEF) 7	K.ES*APQVLLPEEK.I	2	2.50	β-adrenergic receptor kinase motif	Ras guanyl-nucleotide exchange factor activity, intracellular signaling cascade, small GTPase mediated signal transduction
		R.SPPET*APEPAGP.E	2	2.42		
P15515	Histatin-1	R.EFPFYGDYGS*NYLYDN-C-terminal	2	3.42	CKI	Protein binding, response to xenobiotic stimulus, response to bacterium & fungus, biomineral formation
		R.EFPFYGDYGS*NY.L	2	3.70		
P026790	Fibrinogen gamma chain	Y.AYFAGGDAGDAFDGDFGDDPS*DK.F	2	2.21	CKI/CKII	Calcium ion binding,protein binding/bridging, Blood coagulation

Accession number	Protein name	Phosphopeptides	Z	XCorr	Protein kinase(s) involved	Molecular function/ Biological process
Q9UEW3	Macrophage receptor	G.AT*GPS*GPQPPGVK. R.DGAT*GPS*GPQPPGVKGEAGLQGPQGPAGK.Q K.QRVS*YVT*ST*E.N	2 3 2	2.4 3.8 2.25	unknown	Motor protein, force generating protein of respiratory cilia, Microtubule-based movement
Q9C0G6	Dynein heavy chain 6					Scavenger receptor activity, calcium-dependent protein binding, induction of bacterial agglutination, epithelial cell differentiation, innate immune response
Q9UGM3	Deleted in malignant brain tumors 1 protein	Q.FGQGS*GPVLDVDR.C	2	3.71	unknown	Precipitation of tannin, lubrication of oral tissue.
Q04118	Basic salivary proline-rich protein 3	N.EDVS*QEESPSVISGKPEGR.P N.EDVSQEES*PSVISGKPEGR.P Q.SLNEDVS*QEESPSVISGKPEGR.P Q.SLNEDVSQEES*PSVISGKPEGR.P	2 2 2 2	4.96 3.82 4.90 3.58	CKI/CKII	nitric oxide transport, oxygen binding, hemoglobin binding, positive regulation of nitric oxide biosynthetic process, oxygen transport, nitric oxide transport
P68871	Hemoglobin subunit beta	R.FFESFGDLS*TPDAVMGNPK.V R.FFES*FGDLSTPDAVMGNPK.V R.FFESFGDLST*PDAVMGNPK.V	2 2 2	2.20 4.75 4.34	CKI/CKII	Precipitation of tannin, lubrication of oral tissue.
P02812	Basic salivary proline-rich protein 2]	K.PQPPPPQGGSKS*RSS*R.S K.PQPPPPQGGSKSRSS*R.S K.QGPPPPQGGSKSRSS*R.S K.S*QS*ARS*PPGKPPQPPQ. H.T*STVLTTTAT*T*TR.T	2 2 2 2 2	3.15 3.21 3.32 2.90 2.33	cGMP / cAMP /CKII	Extracellular structural constituent, proteinase inhibitor activity, Cell adhesion
Q9HC84	Mucin-5B					Protein binding, Signal transduction
Q01484	Ankyrin-2	E.CAEEDDS*ENGEK.K K.GSSEES*LGEDPGL.A	2 2	2.24 2.8	CKI/CKII	Protein antigen binding, Immune response
P01876	Ig alpha-1 chain C region	L.SVTWS*ESGGQVTAR.N R.WLQGS*QELPR.E	2 2	2.40 2.60	CKII	regulation of biological process, cell morphogenesis
Q14563	Semaphorin-3A	N.SS*SYHTFLLDDEER.S N.SSS*YHTFLLDDEER.S S.S*S*YHTFLLDDEER.S	2 2 2	2.54 2.51 2.42	unknown	

Accession number	Protein name	Phosphopeptides	Z	XCorr	Protein kinase(s) involved	Molecular function/ Biological process
P10909	Clusterin precursor	K.QIKT*LEIKT*NEER.K	2	3.15	CKII/	Protein binding, Apoptosis, innate immune response, complement activation
Q9NNX1	Tuftelin	A.RAKTENPGS*IR.I H.SAGHS*LASELVES*H.D	2	2.80	PKC / CK II	structural constituent of tooth enamel, tissue & bone remodeling, odontogenesis
P01842	Ig lambda chain C region:	R.S*YS*QCQTHEGSTVEK.T	2	3.37	unknown	Antigen binding, Immune response
P02808	Statherin	N-term-DS*S*EEKFLR.R	2	2.72	CKI/CKII	hydroxyapatite binding, protein binding, negative regulation of bone mineralization, inhibition of calcium phosphate deposition
Q9UBC2	Epidermal growth factor receptor substrate	K.KTVFPGAVPVLPAAS*PP.P T.PGPDSS*SGSLGSGEFTGVK.E	1	2.2	unknown	Protein binding, Cell adhesion
Q8WX17	Overian cancer marker/antigen (Mucin 16)	R.GSDT*APS*MVTS*PGVDTR. R.SGSS*SSPIS*LSTK.	2	2.2	CKI	
Q95185	Netrin receptor UNC5C precursor	K.GSTHNLRLS*IHD1.A K.VYNT*SGAVSPQDD.L	2	2.21	CKII	
Q96RZ0	Hypothetical protein gs103	R.LTS*PPQPHFEPPPP.T R.LIGWAS*RSLLP	2	2.20	unknown	Developmental protein receptor, glycoprotein, phosphoprotein, apoptosis
Q14679	Tubulin-tyrosine-like protein	E.ILT*KPLSNHEK.V	2	2.54	unknown	unknown
			2	2.21	cGMP	Tubulin-tyrosine ligase activity, Protein modification
			2	2.25	cGMP	

Identified phosphoproteins, their specific phosphopeptide sequence regions and precise site(s) of phosphorylation, physico-chemical properties, molecular function, biological processes they are involved in, and the protein kinases that are involved in their phosphorylation.

Asterisks following the residue denotes phosphorylated amino-acid. CKII, casein kinase II; CKI, casein kinase I; cGMP, cyclic-guanidine-mono-phosphate-dependent kinase; cAMP, adenosine-mono-phosphate-dependent kinase; PKC, protein kinase C. Z = charged state, and XCorr = cross correlation.

The listed phosphoproteins represent common proteins found in all three different derivatization methods as detailed within the text, i.e., (i) derivatization in the presence of NaOH; (ii) first o-deglycosylation followed by derivatization in the presence of NaOH; and (iii) derivatization in the presence of Ba(OH)₂.

Table 2

Identification of proteins in whole saliva augmenting and supporting the one peptide identification criteria during phosphoproteome analysis.

Accession number	Protein name	Identified Peptides	Z	Xcorr
P15515	Histatin 1	H.REPFYGDYGSNYLYDN-C-term	2	5.07
		R.EPFYGDYGSNYLYDN-	2	5.65
		P.FYGDYGSNYLYDN-	2	4.03
		R.EPFYGDYGSNYL.Y	2	3.32
		R.EPFYGDYGSNYL.L	2	3.96
		F.PFYGDYGSNYLYDN-	2	3.43
Q04118	Basic salivary proline-rich protein 3	G.SNYLYDN-	1	1.85
		Q.GPPPPGNGNPQLPPAGKPQPPPPQGGGRPH.R	3	5.13
		Q.GPPPPGNGNPQQLPPAGKPQGGGR.P	3	3.73
		Q.GPPPPGNGNPQQLPPAGKPQ.G	2	4.53
		Q.GPPPPGNGNPQQLPPAGK.P	2	3.41
		K.PQGGPPPPQGGGRPH.R	3	4.10
		Q.SQPPPHPGKPE.G	2	3.22
		R.PHRPQGGPPQ-	2	3.16
		D.VNDWVGPPNDNGVTK.E	2	3.83
		K.LHNLNSNWFEPSK.P	2	4.42
		K.DVNDWVGPPNDNGVTK.E	2	4.45
		K.R.YFENGKDVNDWVGPPNDNGVTK.E	3	5.13
P04745	Salivary alpha-amylase	K.LHNLNSNWFEPSKPF.I	2	4.38
		R.LSGLLDLALGK.D	2	4.36
		K.LHNLNSNWFEPSKPFYQEVIDLGGEPK.S	3	6.64
		N.LNSNWFEPSKPFYQEVIDLGGEPK.S	3	5.53
		R.FGYGYGPYQPVPEQPLYPQPYQPQYQYT.F	2	3.78
		R.FGYGYGPYQPVPEQPLYPQPYQPQYQYT.T	3	5.25
		R.FGYGYGPYQPVPEQPLYPQPY.Q	2	3.74
		R.FGYGYGPYQPVPEQPLYPQPY.Q	3	5.29
		R.FGYGYGPYQPVPEQPLYPQPYQPQYQ.Q	3	5.25
		R.FGYGYGPYQPVPEQPLYPQPYQP.L	2	3.06
		P.EQPLYPQPYQPQYQYT.F	2	3.00
P02808	Statherin			

Accession number	Protein name	Identified Peptides	Z	Xcorr
Q9UGM3	Deleted in malignant brain tumors 1 protein	R.GSWGTVCDSDSWDTSDANVVR.Q	2	5.12
		P.GNAWFGQGGPIALDDVR.C	2	3.61
		Q.FGQGGPIALDDVR.C	2	4.03
		F.GQGGPIALDDVR.C	2	2.85
Q99504	Eye absent homolog 3	A.AAVASISNQDYPTYTIL.G	2	3.83
		K.PESGLIQTPSQH.S	2	2.53
Q8WX17	Ovarian cancer marker/antigen (Mucin 16)	R.GSDTAPSMVTSPGVDTR.	2	2.35
		R.SGSSSSPISLSTEK.	1	2.2
P02812	Basic salivary proline-rich protein 2	S.PPGKQPQQGGNQPGPPPPGKQ.G	3	6.40
		K.PQPPQGGNKPQGGPPPGKQGGPPQ.G	3	3.93
		P.QGPPQGGNQPGPPPPGKQ.G	3	4.78
		K.PQPPQGGNKPQGGPPPGKQ.Q	2	3.02
		Q.GPPQGGNKPQGGPPPGKQGGPPQGDK.S	3	4.91
		K.PQGGPPQGGNKPQGGPPPGK.P	2	2.81
		K.SQGGPPPGKQGGPPPGGSK.S	2	3.90
		R.SPPGKQGGPPQGGNPPQ.G	2	4.28
		R.GPPGAPGEIGPQ.G	1	2.21
		R.GPPGAPGEIGPQGGPEPGRG.A	2	3.50
P08123	Collagen alpha-2(I) chain	R.GPPGSRGS*PGAPGGPPGSH.	3	3.60
		P.PGPAGSRGDGPPGMTGFPGAAGR.T	3	3.1
		G.PIGSAGPPGPPGAPGPK.A	2	2.31
		.PGPPGAVGPAGK.	2	2.99
O15021	Microtubule-associated Serine/Threonine-protein kinase	T.ARSPGTVMESNPQQR.E	2	2.46
		K.PCESDFETIK.L	2	2.51
Q9C0G6	Dynein heavy chain 6	K.QRVSYVTSTE.N	2	2.25
		K.YINNPDFVPEK.V	2	2.21
		R.LGEDLNKWWQALLVQIRK.A	2	2.89
P10909	Clusterin	R.VTTVASHTSDSDVPSGVTEVVVK.L	2	5.51
		K.LFDSDPITVTVPVESR.K	2	4.69
		R.ELDESLQVAER.L	2	3.32
Q9HC84	Mucin-5B	RASSIIDELFQDR.	2	3.54
		H.TSVTLTTTATTTR.T	2	2.33

Accession number	Protein name	Identified Peptides	Z	Xcorr
Q01484	Ankyrin-2	R.AAGGAVCEQPLGLECR.A	2	4.28
		K.AVTLSLDGGDTAIR.V	2	3.54
		R.TGLLVEQSGDYIK.V	2	3.98
		E.CAEEDDSENGEK.K	2	2.24
Q9UEW3	Macrophage receptor	K.GSSEESLGEDPGL.A	2	2.8
		R.DGATGPSGPGPPGVKGEAGLQGPQGAPGK.Q	3	3.8
		G.ATGPSGPGPPGVK.	2	2.4
		G.MPGAPGPPGPAEK.	2	2.23
P01842	Ig lambda chain C region	R.SYSCQVTHEGSTVEK.T	2	3.92
P026790	Fibrinogen gamma chain	K.ATLVCLISDFYPGAVTVAWK.A	2	4.89
		R.LDGSVDFK.K	2	2.23
		K.YEASILTHDSSIR.Y	2	3.6
		K.VGPEDYR.L	2	2.90
O951885	Netrin receptor UNC5C precursor	K.GSTHNLRLSIHDIA	2	2.21
Q96RZ0	Hypothetical protein gs103	K.VYNTSGAVSPQDD.L	2	2.20
		R.LTSPQPHEEPPPP.T	2	2.54
Q14679	Tubulin-tyrosine-like protein	R.LIGWASRSLH.P	2	2.21
		E.ILTKPLSNHEK.V	2	2.25
		R.RPYSCHELFGE.D		
P68871	Hemoglobin subunit beta	R.FFESFGDLSTPDAVMGNPK.V	2	2.20
P02810	Salivary acidic proline-rich phosphoprotein	K.VNVDEVGGEALGR.L	2	3.47
		R.QGPPLGGQSQSPSAGDGNQGGDGPQQGPPQGGGQQQ-	3	6.63
		E.DVPLVISDGGDSEQFIDEER.Q	2	4.50
		R.PQGPQQGGHQGGPPPPPGKPQ.G	3	6.18
		Q.DDGPQQGPQQGGQGGQGGPPPPQGK.P	3	6.12
		Q.GPPPPQGGKPPQGGHQGGHPPPPQGR.P	3	3.61
		K.PQGPQQGGHPPPPQGRPQ.G	3	4.24
		R.PQGPQQGGHPRPP.R	2	3.72
		K.PQGPQQGGHPPPPQGR.P	2	4.31
		Q.GPPQQGGHPPPPQGR.P	2	3.80
P01876	Ig alpha-1 chain C region	L.SVTWSESGQGVTA.N	2	2.40

Accession number	Protein name	Identified Peptides	Z	Xcorr
		R.WLQGSQELPR.E	2	2.60

Z = charged state and Xcorr = cross correlation.