Journal of Proteomics 127 (2015) 89–95

Contents lists available at ScienceDirect

Journal of Proteomics

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

# Human salivary proteome — a resource of potential biomarkers for oral cancer☆

Priya Sivadasan [a,b], Manoj Kumar Gupta [b], Gajanan J. Sathe [b], Lavanya Balakrishnan [b], Priyanka Palit [a], Harsha Gowda [b], Amritha Suresh [a,c], Moni Abraham Kuriakose [a,c,*], Ravi Sirdeshmukh [b,c,**]

[a] Head and Neck Oncology, Mazumdar Shaw Medical Center, Narayana Health, Bangalore 560099, India
[b] Institute of Bioinformatics, International Tech Park, Bangalore 560066, India
[c] Mazumdar Shaw Center for Translational Research, Mazumdar Shaw Medical Foundation, Narayana Health, Bangalore 560099, India

ARTICLE INFO

*Article history:*
Received 2 February 2015
Received in revised form 9 May 2015
Accepted 26 May 2015
Available online 12 June 2015

*Keywords:*
Saliva
Salivary proteome
Proteomics
Mass spectrometry
Oral cancer

ABSTRACT

Proteins present in human saliva offer an immense potential for clinical applications. However, exploring salivary proteome is technically challenged due to the presence of amylase and albumin in high abundance. In this study, we used four workflows to analyze human saliva from healthy individuals which involved depletion of abundant proteins using affinity-based separation methods followed by protein or peptide fractionation and high resolution mass spectrometry analysis. We identified a total of 1256 human salivary proteins, 292 of them being reported for the first time. All identifications were verified for any shared proteins/peptides from the salivary microbiome that may conflict with the human protein identifications. On integration of our results with the analyses reported earlier, we arrived at an updated human salivary proteome containing 3449 proteins, 808 of them have been reported as differentially expressed proteins in oral cancer tissues. The secretory nature of 598 of the 808 proteins has also been supported on the basis of the presence of signal sequence, transmembrane domain or association with exosomes. From this subset, we provide a priority list of 139 proteins along with their proteotypic peptides, which may serve as a reference for targeted investigations as secretory markers for clinical applications in oral malignancies.
This article is part of a Special Issue entitled: Proteomics in India.



## 1. Introduction

Human saliva is a complex biological fluid that bathes the oral cavity and is critical to the preservation and maintenance of oral health [1,2]. Composed of more than 99% water [3], it contains secretions from the salivary glands (parotid, submandibular, sublingual glands, and minor salivary glands) and non-salivary components including the gingival crevicular fluid, nasal and bronchial secretions, blood derivatives, desquamated epithelial linings, food components and micro-organisms [2]. The chemical composition of saliva, which primarily includes proteins, peptides, nucleic acids and enzymes, suggests it to be an informative biological fluid useful for diagnosis, prognosis and post-treatment surveillance of patients with oral cancers as well as other diseases [4,5]. Amylase and albumin account for approximately 60% of the salivary proteome [6–9]. In addition, the abundant protein portfolio includes proline-rich proteins, mucins, cystatins and statherins along with other plasmatic proteins [10–12]. A comprehensive cataloguing of the lesser abundant salivary proteome would hence be important and help in the identification of disease-specific biomarkers.

Mass spectrometry (MS)-based proteomics has been earlier employed to explore the salivary proteome under normal and pathological conditions. After the initial studies using 2D-MS approach that normally accessed highly abundant proteins, the first high throughput proteomic analysis of saliva using LC–MS/MS approach was published by Xie et al. [9] revealing 437 proteins. This was followed by several other reports. Denny et al., using a combination of multiple depletion and fractionation strategies reported a total of 1166 salivary proteins, with a high proportion of these proteins being also present in blood plasma [13]; this result was also supported by another study wherein many salivary proteins were found to originate from plasma [4]. A study using capillary isotachophoresis-based multi-dimensional separation platform coupled with tandem mass spectrometry identified a total of 1479 salivary proteins [14]. The use of hexapeptide libraries for dynamic range compression coupled with three dimensional peptide fractionation using preparative isoelectric focusing, SCX and capillary-reversed-phase HPLC, followed by LC–MS/MS analysis resulted in the identification of 2340

☆ This article is part of a Special Issue entitled: Proteomics in India.

* Correspondence to: M.A. Kuriakose, Mazumdar Shaw Medical Center, Narayana Health, Bangalore 560099, India.
** Correspondence to: R. Sirdeshmukh, Institute of Bioinformatics, Bangalore, 560066, India.
*E-mail addresses:* makuriakose@gmail.com, moni.abraham@ms-mf.org (M.A. Kuriakose), ravisirdeshmukh@gmail.com, ravi@ibioinformatics.org (R. Sirdeshmukh).

http://dx.doi.org/10.1016/j.jprot.2015.05.039
1874-3919/

human salivary proteins [15], the largest number identified in any one study. These earlier studies differed with respect to saliva sampling — glandular or whole saliva, sample processing and analytical platforms.

In the present study, we carried out proteomic analysis of saliva from healthy individuals by using variations of depletion and fractionation strategies followed by high resolution mass spectrometry. Our analysis resulted in the identification of 1256 human proteins that were exclusive of any protein/peptides of microbial origin present in the saliva. The identified human proteins include 292 novel identifications. By integrating our results with earlier reports, we present an updated salivary proteome as a useful reference for developing clinical applications for oral malignancies.

## 2. Materials and methods

### 2.1. Sample collection and processing

The study was approved by the Institutional Ethics Committee. The procedure for collection and processing of saliva was adapted from earlier reports [16,17]. Briefly, unstimulated saliva samples (5 ml) were collected from healthy subjects of either sex in the age group between 20–50 years, with written informed consent. The individuals selected were without any risk habits like tobacco chewing, smoking or alcohol abuse. Samples were collected in the morning after rinsing the mouth with water and with subjects refraining from food/drink for at least 1 h prior to the collection. All the samples were centrifuged at 2000 rpm, at 4 °C for 10 min to remove the cells. The supernatant was then collected and centrifuged at 14,000 rpm to remove any debris. Protein estimation was carried out using RC-DC protein assay (Bio-Rad, USA) as per the manufacturer's guidelines and the samples were stored at $-80^0$ C until further use.

### 2.2. Depletion and fractionation methods

Equal volumes of saliva were pooled based on the age groups and pooled saliva samples were processed further. One pool included samples from individuals of 30–50 years of age, (Pool A) and the other pool included samples from individuals of 20–30 years of age (Pool B). We adopted two strategies to deplete abundant proteins. Depletion of amylase alone was carried out by using starch affinity-based amylase capture and depletion of amylase and plasmatic proteins by amylase capture followed by antibody-based depletion of plasma proteins such as albumin, immunoglobulins and others. The depleted protein fraction was then subjected to fractionation on SDS-PAGE and in-gel tryptic digestion or in solution digested with trypsin and tryptic peptides were fractionated by SCX chromatography (Fig. 1; workflow 1–3 respectively). In another strategy, compression of the protein dynamic range of total salivary proteins was carried out using hexapeptide library enrichment kit (ProteoMiner, Bio-Rad, CA, USA). The tryptic digest of the enriched protein fraction was then subjected to fractionation by SCX chromatography (Fig. 1; workflow 4).

For amylase depletion, 5 ml of pooled saliva (approximately 5 mg of protein) was mixed with 1.5 g of potato starch (Sigma Aldrich, MO, USA) [previously washed 3 times with water (3000 rpm, 5 min)] and incubated for 1 h in a rotating shaker, at room temperature. The mixture was then centrifuged at 3000 rpm for 5 min and the supernatant was collected. The pellet was washed again to recover trapped saliva. Protein estimation was then carried out as mentioned above. Depletion of albumin, immunoglobulins and any other abundant plasma proteins (transferrin, fibrinogen, immunoglobulin A, haptoglobin, alpha antitrypsin, alpha 2 macroglobulin, immunoglobulin M, apolipoprotein A1, alpha1 acid glycoprotein, Complement C3, apolipoprotein A11 and transthyretin) was carried out using Human MARS-14 spin cartridge (Agilent Technologies, CA, USA) as per manufacturers' instructions. The protein sample after amylase depletion was passed through the MARS-14 cartridge and the unbound protein was collected. The procedure was repeated multiple times to collect approximately 500 μg of depleted protein fraction for further experiments. Flow through fractions were collected, concentrated and desalted using a 5 kDa MW cut off ultracentrifugal filter device (Amicon, Millipore, Billerica, MA). The protein concentration of the sample was determined as mentioned above.

Two hundred micrograms of above mentioned depleted saliva protein was resolved on a 10% SDS-PAGE (16X18cm) and gel was stained using colloidal Coomassie blue. Twenty-five gel slices were excised and destained using 40 mM ammonium bicarbonate in 40% acetonitrile (ACN). The sample was then subjected to reduction using 5 mM DTT (60 °C for 45 min) followed by alkylation using 20 mM iodoacetamide (10 min. at room temperature). In-gel digestion with trypsin was carried out at 37 °C for 12–16 h using modified sequencing grade trypsin (Promega, WI, USA). Peptides were extracted from gel pieces sequentially using 5% formic acid, 5% formic acid in 40% ACN and finally with 100% ACN. The extracted peptides were dried and stored at −80 °C until LC–MS/MS analysis.

Alternatively, depleted protein fraction was subjected to direct in-solution digestion with trypsin and the resulting peptides fractionated by SCX chromatography. Briefly, 200 μg of protein was reduced with 5 mM DTT and alkylated using 10 mM IAA as above. The proteins were then digested with trypsin as above and the digested peptide mix was reconstituted in solvent A (10 mM potassium phosphate, 30% ACN, pH 2.7) and fractionation was carried out on a SCX column (Polysulfoethyl A column; 300 Å, 5 μm, 100 × 2.1 mm; PolyLC, MD, USA) using 1200 HPLC system (Agilent Technologies, CA, USA) coupled with a binary pump, UV detector and a fraction collector. Peptides were eluted using a linear salt gradient (0 to 35%) of solvent B (10 mM potassium phosphate buffer containing 30% ACN, 350 mM KCl, pH 2.7) at a flow rate of 200 μl/min. The adjacent fractions were then pooled based on the chromatographic profile to make the total number to 25. The samples were dried, reconstituted in 0.1% TFA and desalted using C18 stage-tip. The desalted samples were dried and stored at −80 °C until further analysis.

For enrichment using ProteoMiner, salivary proteins were subjected to the procedure according to the manufacturers' instructions (ProteoMiner; Bio-Rad, CA, USA). Briefly, 10 mg of salivary protein was added to the ProteoMiner column, incubated in a rotational shaker for 2 h at room temperature and centrifuged at 1000 × *g* for 1 min to discard the unbound fraction. The column was then washed thrice with 200 μl of wash buffer, by centrifugation at 1000 *g* for 1 min. Two hundred microlitres of deionized water was added and centrifuged at 1000 *g* for 1 min. The enriched low abundant proteins bound to the column were eluted with 100 μl of rehydrated elution reagent, desalted using 5 kDa MW cut off ultracentrifugal filter device (Amicon, Millipore, Billerica, MA) and protein estimation was carried out. The enriched protein sample was digested in-solution with trypsin and the tryptic digest was subjected to SCX fractionation as described above.

### 2.3. LC–MS/MS analysis

Fourier-transform LTQ-Orbitrap Velos mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) equipped with Proxeon Easy nLC was used for LC–MS/MS analysis. In house chromatographic capillary columns made up of Magic $C_{18}$ AQ reversed phase material (Michrom Bioresources, 5 and 3 μm, 100 Å) were used for HPLC. Nanospray source with an emitter tip of 10 μm (New Objective, Woburn, MA) was used for ionization with a voltage of 2 kV. Peptides were enriched on trap column (75 mm × 2 cm) at a flow rate of 3 μl/min using Solvent A (0.1% formic acid) followed by fractionation in an analytical column (75 mm × 10 cm) to resolve the peptides. A linear gradient of 7–30% solvent B (0.1% formic acid, 95% ACN) was used at a flow rate of 350 nl/min, for 80 min. The mass spectrometry parameters used are as follows: acquisition of the full scan data was implemented with a mass resolution of 60,000 at 400 m/z, top 20 intense peaks from each MS cycle were selected for MS/MS fragmentation with a mass resolution of 15,000 at 400 m/z. Only multiple charged peptides were selected and 39% normalized collision energy was used for fragmentation with 45 s exclusion