Published in final edited form as:

J Proteome Res. 2011 November 4; 10(11): 5031–5046. doi:10.1021/pr200505t.

Characterization of the human submandibular/sublingual saliva glycoproteome using lectin affinity chromatography coupled to Multidimensional Protein Identification Technology

Mireya Gonzalez-Begne¹, Bingwen Lu³, Lujian Liao³, Tao Xu³, Gurrinder Bedi¹, James E. Melvin^{2,a}, and John R. Yates III^{3,*a}

¹Center for Oral Biology, University of Rochester Medical Center; Rochester, New York 14642

²Department of Pharmacology and Physiology, University of Rochester Medical Center; Rochester, New York 14642

³Department of Chemical Physiology, The Scripps Research Institute; La Jolla, CA 92037

Abstract

In-depth analysis of the salivary proteome is fundamental to understanding the functions of salivary proteins in the oral cavity and to reveal disease biomarkers involved in different pathophysiological conditions, with the ultimate goal of improving patient diagnosis and prognosis. Submandibular and sublingual glands contribute saliva rich in glycoproteins to the total saliva output, making them valuable sources for glycoproteomic analysis. Lectin-affinity chromatography coupled to mass spectrometry-based shotgun proteomics was used to explore the submandibular/sublingual (SM/SL) saliva glycoproteome. A total of 262 *N*- and *O*-linked glycoproteins were identified by multidimensional protein identification technology (MudPIT). Only 38 were previously described in SM and SL salivas from the human salivary *N*-linked glycoproteome, while 224 were unique. Further comparison analysis with SM/SL saliva of the human saliva proteome, revealed 125 glycoproteins not formerly reported in this secretion. KEGG pathway analyses demonstrated that many of these glycoproteins are involved in processes such as complement and coagulation cascades, cell communication, glycosphingolipid biosynthesis neolactoseries, *O*-glycan biosynthesis, glycan structures-biosynthesis 2, starch and sucrose metabolism, peptidoglycan biosynthesis or others pathways. In summary, lectin-affinity chromatography coupled to MudPIT mass spectrometry identified many novel glycoproteins in SM/SL saliva. These new additions to the salivary proteome may prove to be a critical step for providing reliable biomarkers in the diagnosis of a myriad of oral and systemic diseases.

Keywords

Submandibular/Sublingual saliva; MudPIT; lectin-affinity chromatography; glycoproteins; biomarkers

*Correspondence: John R. Yates III., The Scripps Research Institute, Department of Chemical Physiology, North Torrey Pines Road, SR-11, La Jolla, CA 92037, Phone: (858)784-8862, Fax: (858)784-8883, jyates@scripps.edu.

^aSenior authors

Supporting information:

Detailed information summarizing annotation of SM/SL glycoproteins, protein evidence corresponding to common glycosylated proteins and KEGG annotations are available free of charge at <http://pubs.acs.org>.

All the raw mass spectra files corresponding to the lectin-affinity *N*- and *O*-linked glycoproteome are available to the scientific community at <http://fields.scripps.edu/published/salivaglycoproteome2011/>

Introduction

Saliva plays a vital role in the maintenance of oral health ^{1,2}. Three major pairs of salivary glands and various minor salivary glands secrete more than 1,000 different proteins with numerous functions, e.g., cleansing of the oral cavity, lubrication, digestion, tooth mineralization, anti-viral and anti-bacterial activities, among others ^{1,3–8}. In fact, the multiple functions of saliva are derived from its specific components and unique gland fluid characteristics ^{6,9}.

Submandibular salivary glands secrete a unique mixture of mucous and serous components, while the sublingual glands secrete mainly a mucous-containing fluid ^{6,10,11}. The anatomical and physiological characteristics of these two glands make them attractive fluid sources for analyzing glycoprotein content.

Many posttranslational modifications (PTMS) have been reported in salivary proteins ^{12–14}. Among them glycosylation has been considered a common PTM that regulates many cellular processes ^{15–25}. Different analytical procedures have been used to profile the salivary glycoproteome to date. Using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) in combination with electroblotting and a battery of seven lectins Carpenter and colleagues ²⁶ found that many parotid salivary proteins were *N*-glycosylated, while only IgA- α chain was *O*-glycosylated. In addition, they discovered the presence of salivary glycoprotein polymorphisms in different individuals. In a high-throughput glycoproteomic study conducted on human whole saliva (which contains mainly saliva from the major and minor salivary glands, but also gingival crevicular fluid, nasal and bronchial secretions, food debris, bacteria, desquamated epithelial cells and blood/serum components), using a hydrazide chemistry and release method ²⁵ coupled to liquid chromatography tandem mass spectrometry (LC-MS/MS), 84 *N*-glycosylated peptides were identified from 45 proteins ²³. Moreover, 16 new salivary proteins and 44 new glycosylation sites were reported. Larsen and collaborators ¹⁹ utilizing high-affinity-sialic acid TiO₂ chromatography in combination with nanoscale liquid chromatography-tandem mass spectrometry found 97 *N*-linked glycosylation sites corresponding to 45 proteins in whole saliva. More recently, a modified hydrazide capture method ²⁷ in conjunction with LC-MS/MS and 2D LC-MS/MS revealed a total of 156 formerly *N*-glycosylated peptides matching a total of 77 unique *N*-glycoproteins in different types of saliva. Of these, 122 peptides and 62 proteins were identified in whole saliva, 62 peptides and 34 proteins were found in parotid saliva, 80 peptides and 44 proteins were detected in submandibular saliva and 98 peptides and 53 proteins were distinguished in sublingual saliva ²². Further analysis conducted by Sondej and collaborators provided the first glycosylation map showing not only the widespread glycosylation profile of unstimulated whole saliva but also the individual oligosaccharide variability using a panel of 15 lectins in conjunction with 2-D gel electrophoresis, lectin blotting and mass spectrometry techniques ²⁸. Lastly, utilization of a combination of a dynamic range compression method based on hexapeptide libraries (ProteoMiner and Library-2) coupled to hydrazide chemistry and μ LC-MS/MS identified 268 *N*-glycosites in 193 glycoproteins, expanding the whole saliva glycoproteome catalog ¹².

The above studies demonstrated the power of using mass spectrometry in conjunction with other chromatography chemistries as a strategy for the global analysis of the saliva glycoproteome ²¹. However, due to the complex nature of salivary proteins, multiple methodologies are needed to complete the characterization of the salivary proteome, an important consideration if saliva will be used in the future as a tool for diagnosing and monitoring health and disease status. To enrich the existing salivary glycoproteome catalogue, multidimensional protein identification technology (MudPIT) mass spectrometry

was used in conjunction with lectin-affinity chromatography. The analytical methodology used here, was shown to be robust and selective enough to achieve a purification efficiency of more than 90% in spite of the broad specificities and non-glycosylated co-purification problems associated with the use of lectin fractionation. We found 125 glycoproteins not previously identified in the SM/SL human saliva proteome, highlighting the value of combining methodologies to carry out an in-depth exploration of the salivary proteome.

Materials and Methods

Purified lectin affinity chromatography kits containing immobilized lectins [*Canavalia ensiformis* (Con A), *Ricinus communis* (RCA-I) and *Ulex europaeus* (UEA-I)], equilibration/washing buffers, and specific carbohydrates for elution were purchased from EY Laboratories Inc. (San Mateo, CA). Tris base (Promega Co., Madison, WI), ϵ -Amino-*n*-caproic acid (Calbiochem/EMD Biosciences Inc., La Jolla, CA), ethylene diamine tetraacetic acid (EDTA; J.T. Baker, Phillipsburg, NJ), leupeptin (Bachem Inc., Torrance, CA), trypsin and lysine C (Roche Applied Science, Indianapolis, IN) were also acquired. The remaining chemicals were bought from Sigma-Aldrich (St. Louis, MO), unless otherwise indicated.

Saliva collection

The protocol for the collection of human submandibular/sublingual saliva was approved by the University of Rochester Institutional Review Board. A written informed consent was obtained from the donors before salivary collection.

Non-cannulated submandibular/sublingual (SM/SL) saliva (without the insertion of a tapered polyethylene tubing in the secretory duct)²⁹ was obtained from two healthy, non-medicated and non-smoking Caucasian male donors (48–50 years old). Saliva collection was performed under standard conditions between 8:00 a.m. and 10:00 a.m. at the University of Rochester, Center for Oral Biology. Subjects were asked not to eat, drink, or perform any oral hygiene measures at least one hour before salivary collection. Stimulated SM/SL saliva was collected on ice by applying to the tongue 0.4% citric acid at a 30 seconds interval during a 5 minute collection period²⁹. A custom-fitted Block and Brotman collector³⁰ was used since secretions from both glands are mostly voided in the oral cavity via a common duct^{29,31–33} and cannulation can tear the secretory duct wall if not done properly²⁹. An equal volume of 2x protease inhibitor cocktail (0.1 M Tris-HCl (pH7.4), 0.1 M epsilon amino caproic acid, 0.01 M sodium EDTA, 5 mg/L of pepstatin A, 0.005 M benzamidine-HCl, 0.05 mg/L of leupeptin and 0.002 M phenylmethylsulfonyl fluoride) was added immediately to the final volume of collected saliva to ensure protein integrity preservation. After collection, saliva was centrifuged at 10,000 rpm for 20 minutes (Sorvall RC6 Plus; Thermo Fisher Scientific Inc. Asheville, NC), dialyzed using a 3.5 kDa molecular weight cutoff dialysis membrane (Spectra/Por®3; Spectrum Laboratories, Inc.; Rancho Dominguez, CA) and lyophilized (Heto PowerDry LL3000 Freeze Dryer; Thermo Scientific Inc., Asheville, NC).

Lectin-Affinity Chromatography

The experimental strategy used in this study involved the utilization of Con A, RCA-I and UEA-I lectin-affinity chromatographies for enrichment of salivary glycoproteins. Con A is well recognized to specifically bind α -mannosidic structures (high-mannose type, hybrid-type and biantennary complex type *N*-glycans) and α -glucose^{16,18,28,34–36}. Meanwhile, RCA-I lectin binds oligosaccharides ending in β -D galactose and *N*-acetyl- α -D-galactosamine^{35–39}, whilst UEA-I lectin binds specifically to Fuc α 1-2Gal-R^{18,28,35,36}.

Enrichment of salivary glycoproteins was achieved following the manufacturer's protocols for Con A, RCA-I, and UEA-I lectin-affinity chromatographies at 4 ° C. Briefly, chromatography columns containing 1ml of settled agarose-lectin beads covalently linked to Con A, RCA-I and UEA-I were washed with 10x the gel volume of an equilibration/washing buffers containing either: 0.05 M Tris-0.15 M NaCl-0.004 M CaCl₂, pH 7.0 for Con A, or 0.01 M Phosphate-0.15 M NaCl pH 7.2 for RCA-I and UEA-I. 10 mg of dry-weight lyophilized SM/SL saliva dissolved in 5 ml of the specific equilibration/washing buffers was applied to each lectin column. Lectin columns containing the salivary mixture were incubated over night at 4 ° C. After incubation, salivary flow-through was collected and the remaining unbound material was washed off from each column using 7 ml of each equilibration/washing buffers. Collected Con A, RCA-I, and UEA-I unbound fractions (1 ml each) were monitored at 280 nm absorbance to ensure that all unbound salivary proteins were removed. Elution of bound salivary glycoproteins was accomplished as mentioned by the manufacturer's instructions with the following modifications, Con A: 0.2 M methyl α -D-mannopyranoside pH 7.2; RCA-I: 0.1 M α -lactose pH 7.2; and UEA-I: 0.05 M α -L-fucose pH 7.2 buffers. Strongly-bound glycoproteins were eluted with the following buffers: Con A: 0.5 M methyl α -D-mannopyranoside pH 2.5; RCA-I: 0.25 M α -lactose pH 2.5; and UEA-I: 0.5 M α -L-fucose pH 2.5 buffers. 1 ml fractions were collected and the elution of salivary glycoproteins was monitored at 280 nm. Following the elution step, fractions were immediately neutralized with 50 μ l of 1 M Tris buffer pH 9.5. Subsequently, the columns were regenerated with 10 times (gel volume) of 1.4 M NaCl and then re-equilibrated using 50 times (gel volume) of equilibration/washing buffer.

To potentially capture a higher number of glycoproteins, the fractions corresponding to the two elution steps (from neutral and acidic pHs) were pooled together for each Con A, RCA-I and UEA-I lectin experiment, dialyzed against 50 mM ammonium bicarbonate to remove salts and carbohydrates and then freeze-dried.

Sample processing and Digestion for proteomic analysis

Eluted Con A, RCA-I and UEA-I sample solutions were precipitated by adding ice-cold trichloroacetate (TCA) to a final concentration of 30%. After isolating the protein precipitate by centrifugation (14,000g, 15 min), the pellets were washed twice with ice-cold acetone, and subsequently dried by SpeedVac. The protein pellets were resuspended with 8 M urea in 5x Invitrosol protein solubilization buffer (Invitrogen, CA), reduced by 10 mM Tris(2-rboxyethyl)phosphine hydrochloride (TCEP) for 15 minutes at room temperature, and cysteine alkylated by 15 mM iodoacetamide for 30 minutes in the dark at room temperature. The samples were then diluted 4 fold with 100 mM Tris-HCl buffer, pH 8.5; a final concentration of 1 mM CaCl₂ was added and afterward the samples were digested with trypsin at a substrate/enzyme ratio of 100:1 at 37 °C for 16 hours. The digestion was terminated by adding 90% formic acid to a final concentration of 4%.

Analysis of protein digests with Multi-dimensional Liquid Chromatography-Tandem Mass Spectrometry

The protein digests were pressure-loaded onto a triphasic chromatography column. The triphasic column consisted of a 100- μ m i.d. capillary with a 5- μ m pulled tip and was packed in the following order from the tip: 1) 10 cm of 5- μ m Aqua C₁₈ material (Reversed Phase, or RP), 2) 3 cm of 5- μ m Partisphere strong cation exchange material (SCX), and 3) 3 cm of 5- μ m Aqua C₁₈ material.

After loading the peptide digests, the column was washed for 15 minutes with buffer A (5% acetonitrile/0.1% formic acid), and then placed inline with an Agilent 1100 quaternary HPLC (Agilent, Palo Alto, CA) or Eksigent nano 2D-LC system (Eksigent, Dublin, CA)

and analyzed using a modified 12-step separation. The buffer solutions were 5% acetonitrile/0.1% formic acid (buffer A), 80% acetonitrile/0.1% formic acid (buffer B), and 500 mM ammonium acetate/5% acetonitrile/0.1% formic acid (buffer C). Step 1 consisted of an 85 min gradient from 0–100% buffer B. Steps 2–11 had the following profile: 3 min of 100% buffer A, 5 min of X% buffer C, a 10 min gradient from 0–15% buffer B, and a 97 min gradient from 15–45% buffer B. The 5 min buffer C percentages (X) were 5, 10, 15, 20, 25, 30, 35, 40, 55, and 75% respectively for the 10-step analysis (Steps 2–11). The final step, the gradient, contained: 3 min of 100% buffer A, 20 min of 100% buffer C, a 10 min gradient from 0–15% buffer B, and a 107 min gradient from 15–70% buffer B. The eluted peptides were directly introduced into either a Thermo Finnigan LCQ Deca XP, Thermo Finnigan LTQ linear ion trap or a Thermo Scientific hybrid LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA), by nano-electrospray ionization. The spray voltage was 2.5 kv and the temperature was 200°C. For LTQ-Orbitrap analysis, the mass spectrometry (MS) scan was at the resolution 60,000, followed by 6 data dependent MS/MS scan of the 6 most abundant precursor ions. Dynamic exclusion was applied to exclude 150 ions in the list for 120 minutes.

Interpretation of tandem mass spectra data sets

The human protein database used was the EBI International Protein Index (IPI) protein database version 3.48 (September 1, 2008). The reverse protein sequences of the IPI database were used as the decoy database. The ProLuCID program⁴⁰ was used for MS/MS database search. The validity of peptide/spectrum matches was assessed in DTASelect⁴¹ using two SEQUEST⁴² defined parameters, the cross-correlation score (a.k.a. X_{Corr}) and normalized difference in cross-correlation scores (a.k.a. ΔCn). Distribution of cross-correlation score and normalized difference in cross-correlation scores values for direct and decoy database hits was obtained, and the two subsets were separated by quadratic discriminant analysis. Full separation of the direct and decoy subsets is generally not possible. Therefore, the discriminant score was set such that a false-discovery rate of 1% was determined based on the number of accepted decoy database peptide/spectrum matches. This procedure was independently done on data subsets for charge states +1, +2, and +3. Proteins were considered detected if they were identified by at least two peptides. DTASelect assembles identified peptides into proteins and protein groups by using a parsimony principle in which the minimum set of proteins accounts for all the observed peptides.

Data analysis

Protein function annotation using Gene Ontology (GO) was carried out via GoAssigner, developed in the Yates' lab. The GoAssigner program assigned cellular component, molecular function, and biological process GO terms based on GOA gene association files for human (released on July 28, 2011) by the European Bioinformatics Institute at: <http://www.ebi.ac.uk/GOA/>. Additionally, in order to be able to find the pathways that are significantly represented by the identified glycoproteins a pathway analysis was performed using KEGG (Kyoto Encyclopedia of Genes and Genomes at: <http://www.genome.jp/kegg/pathway.html>). Lastly, confirmation of N- or O-linked glycosylation was achieved by using the European Bioinformatics Institute (EBI) (<http://www.ebi.ac.uk/>) or ExPASy-UniProt-Swiss-Prot and TrEMBL (<http://ca.expasy.org/sprot/>) knowledgebases. In the case of non-annotated proteins, verification of glycosylation was determined using previously published results.

Comparison of human salivary lectin-affinity N- and O-linked glycoproteome to the human saliva proteome, human salivary N-linked glycoproteomes/Glycoprofile of the human

salivary proteome, human salivary sialome, hexapeptide *N*-linked glycoproteome and whole saliva proteome

The databases corresponding to the SM/SL human saliva proteome (917 proteins), the human salivary *N*-glycoproteomes/glycoproteome of the human salivary proteome (100 proteins) and the whole saliva proteome (2,290 proteins) were downloaded from the central repository hosted at <http://www.hspp.ucla.edu>. The Scripps Research Institute (TSRI) database that correspond to male SM/SL human saliva proteome (695 proteins) was downloaded from <http://fields.scripps.edu/public/project/saliva>, while the salivary sialome (45 proteins) and the hexapeptide *N*-linked glycoproteome (187 proteins) were obtained from the supplemental tables published elsewhere^{12,19}. The hexapeptide *N*-linked glycoproteome was mapped to the IPI v.3.48 for comparison purposes with all the proteomes.

Gel electrophoresis and Western blot

50 µg of SM/SL salivary protein was heated at 70 °C for 10 minutes prior to separation in NuPAGE® Novex 4–12% Bis-Tris polyacrylamide gradient gels (Invitrogen, Carlsbad, CA). Protein was transferred onto polyvinylidene (PVDF) membranes (Invitrogen, Carlsbad, CA) for 1 hr using 1x NuPAGE transfer buffer. Membranes were blocked overnight at 4 °C with 5% non-fat dry milk in 25 mM Tris-HCl pH 7.5, 150 mM NaCl (TBS) and then incubated with goat polyclonal antibody to isoform 1 of *N*-acetylated- α -linked acidic dipeptidase-like protein (Santa Cruz Biotechnology, Inc., Santa Cruz, CA), rabbit polyclonal antibody against α -(1,3)-fucosyltransferase (Abgent, Inc., San Diego, CA), 2ndMet rabbit antibody against amino acids 209–285 of rat NKCC1 fusion protein generously provided by Dr. R. J. Turner, NIDCR⁴³, or polyclonal rabbit to Isoform TGN46 of Trans-Golgi network integral membrane protein 2 (Abcam Inc., Cambridge, MA); 1:200; 1:200; 1:3,000; 1:2,000 dilutions, respectively in 2.5% non-fat dry milk solution at 4 °C overnight. After washing with TBS containing 0.1% Tween-20 (TBS-T), the membranes were incubated either with horseradish peroxidase-conjugated goat anti-rabbit IgG secondary antibody or with rabbit anti-goat IgG H&L (HRP) antibody at a dilution of 1:2,500 respectively (Thermo Scientific Pierce Antibodies; Rockford, IL) in TBST/2.5% non-fat dry milk for 1 hr at room temperature. Labeled proteins were visualized using enhanced chemiluminescence (ECL detection kit, GE/Amersham Biosciences; Piscataway, NJ).

Results

Lectin-affinity chromatography analysis of SM/SL glycoproteins

Glycosylation is a common posttranslational modification having critical roles in many biological processes in the cell^{15–25}. Analytical studies of glycoproteins are very complicated due to protein microheterogeneity of the carbohydrate chains,^{15,16,19–21,44}. Therefore, purification of glycosylated proteins involving the use of different methodologies and chemistries^{19,22,23,25,27,28,44–52} is highly desirable.

The analytical methodology proposed in this study involved the use of lectin-affinity chromatography as an enrichment step. Subsequent to the initial washing, flow through step, Con A chromatogram [Fig # 1A] was performed with two elution steps. The first smaller elution peak corresponded to the bound proteins that desorbed from Con A with a modest concentration of competitive eluent (0.2 M methyl α -D-mannopyranoside). The second bigger peak represented the strongly-bound proteins which required high concentration of eluent and a different pH in order to desorb. This elution pattern was previously reported, highlighting the requirement for different elution conditions such as eluent concentration, length pause, pH, and NaCl concentration for efficient glycoprotein desorption⁵⁰. In addition, Koyama and collaborators demonstrated that sugar chains with biantennary

complex-type structures are eluted with low concentration of elution buffer while, high mannose type and hybrid-type sugar chains require a high concentration elution buffer ⁵³.

The RCA-I chromatogram displayed an initial larger peak corresponding to proteins eluted with the 0.1 M α -lactose elution buffer, followed by a smaller peak eluted with the 0.25 M α -lactose elution buffer, representing the retarded galactose-containing proteins ³⁸. It has been previously reported that retarded oligosaccharides require more stringent conditions for their elution ³⁷ [Fig # 1B].

A similar pattern to the RCA-I chromatogram was observed for UEA-I lectin, i.e., the majority of the proteins were desorbed in the first elution step (initial peak) using a mild elution buffer of 0.05 M α -L-fucose [Fig # 1C]. Overall, the elution patterns suggested that factors such as hydrophobic interactions, ionic interactions, hydrogen bonding, hydrophilic interactions and molecule's size, charge, composition, structure, linkage and oligosaccharide branching play an important role in analyte retention ^{20,54,55}.

Proteomic analysis of human salivary lectin-affinity *N*-and *O*-linked glycoproteome

Multidimensional protein identification technology (MudPIT) mass spectrometry, a robust approach for analyzing complex protein mixtures ^{56,57}, was coupled to lectin-affinity chromatography to obtain a comprehensive catalog of SM/SL glycoproteins. A total of 277 proteins were identified using an inclusion criterion of a minimum of two peptides for each recognized protein. Only 262 were *N*-or *O*-linked glycosylated, while the remaining 15 proteins were non-annotated as such. Of the 262 proteins, 109 (41.6%) unique proteins were isolated by the Con A affinity chromatography (e.g. fibromodulin and isoform 1 of kallikrein-11), 41 (15.6%) by RCA-I (e.g. furin and isoform B of Protein FAM3B) and 25 (9.5%) by UEA-I (e.g. isoform 1 of podocalyxin-like protein 2 and beta-hexosaminidase subunit beta).

The remaining 87 proteins (33.2%) were found either in two or three glycoprotein lectin-based isolation techniques, including proteins such as IGHG4 protein, alpha-2-macroglobulin and SPARC-like protein 1 [Figure #2 and Supplemental table #1]. These results demonstrate that many of these salivary proteins express multiple types of glycan modifications ^{28,36,58,59}.

It is well known that human saliva is a rich source of *N*- and *O*- linked glycosylated proteins ^{13,23,28,60}. The selectivity of the methodology used here allowed the detection of both *N*- and *O*-linked salivary glycoproteins not previously attained using other glycochemistries ^{19,22,23,28,35}.

N-linked glycosylated proteins

N-linked glycosylation occurs when carbohydrates are linked to the peptide backbone via asparagine residues and is prevalent in proteins destined for extracellular environments ^{17,20,21,25,46,54,61–64}. More than 50% of human proteins are conjugated to glycans ^{18,28,44,47,58,65,66} with *N*-linked glycosylation being the most prevalent form of posttranslational modification among proteins targeted to extracellular environments ^{25,47,67}. Therefore, it was possible to identify 166 annotated proteins presenting only *N*-linked glycosylation and 50 annotated proteins showing both *N*- and *O*-linked glycosylations [Supplemental table #1]. Examples of these proteins are both extracellular, plasma membrane-associated proteins (ELA2 leukocyte elastase) and proteins buried in the plasma membrane with the *N*-glycosylation sites present in the loop flanking its extracellular site (e.g. isoform 2 of solute carrier family 12 member 2) ⁶⁸, as well as secreted proteins (ribonuclease pancreatic).

It is important to note that the majority of the *N*-linked proteins identified were annotated as glycosylated proteins in the European Bioinformatics Institute (EBI) or ExPASy-UniProt-Swiss-Prot and TrEMBL knowledgebases. However, some of them were not (15.3%), even though their *N*-glycosylation sites have been previously identified by other groups^{18,19,22,23,54,58,61,64}. Furthermore, significant advances have been accomplished in the study of glycoproteins^{20,21,45,52,69}; indeed, thousands of *N*-glycosylation sites have been identified^{25,46,47,54,61,64}. A few examples of these previously identified, but non-annotated proteins that may be clinically important include LTF Truncated lactoferrin,⁶⁴ vitamin D-binding protein precursor¹⁸, and IGHM FLJ00385 protein (Fragment)²². Other examples can be found in Supplemental table #1.

O-linked glycosylated proteins

Conversely, *O*-linked glycoproteins contain *O*-glycans attached to the OH group of hydroxyl amino acids^{62,63}. The most common modification found is the attachment of the carbohydrate moiety to serine or threonine residues^{20,21,62}. A total of 6 annotated *O*-linked proteins were detected. These proteins are also observed intracellularly (e.g. isoform GN-1 of glycogenin-1) besides being secreted^{15,70}. *O*-linked glycoproteins appear to participate in the regulation of signaling pathways^{19,49}.

As expected, there was also some non-specific binding (5.4%; data not shown) associated with the use of the lectin chromatography^{15,19,58}. Yet, the results obtained with our robust methodology enhanced the glycoprotein capture as indicated by the achievement of a purification efficiency of 94.6%. Purification efficiency was estimated by subtracting the percentage of non-specific binding (15 proteins) from the total of 277 identified proteins as previously described by Larsen et al¹⁹.

Proteins not reported to be glycosylated, like glyceraldehyde-3-phosphate dehydrogenase, endoplasmic reticulum mannosyl-oligosaccharide 1,2- α -mannosidase, and *N*-acetylgalactosaminyltransferase 7, were identified. This could be explained by: 1) binding of these proteins to other glycoproteins bound to the lectins^{34,58}, 2) presence of a hydrophobic binding by Con A lectin which allowed the capturing of proteins that don't fit the criteria of canonical *N*-linked glycosylation^{15,36}, or 3) presence of potential glycosylation sites that need further analysis and confirmation.

Overall Con A, RCA-I and UEA-I lectin chromatographies captured salivary proteins presenting α -mannose, α -glucose, β -galactosyl, α -fucosylated, sialylated and *N*-linked complex-structures. These results are consistent with other findings^{28,35,53,71}. For example, we detected both the MUC5B and MUC7 mucins, which exhibit carbohydrate moieties consisting primarily of mannose, galactose, fucose, *N*-acetyl-glucosamine, *N*-acetyl-galactosamine, and sialic acid. MUC7 *per se*, displays different amounts of fucose and sialic acid components^{63,71}.

Gene Ontology (GO) analysis

The glycoproteins obtained from the human salivary lectin-affinity *N*- and *O*-linked glycoproteome and the proteins from the whole saliva proteome were classified according to their biological significance by GoAssigner. Although, the distribution of the categories was equivalent among categories, some areas of enrichment were observed between both proteomes.

In regarding to the “*molecular function*” classification, these proteomes were over-represented in the binding category (48%/54%; glycoproteome/proteome) and under-represented in antioxidant activity (2%/1%; glycoproteome/proteome). This binding mapping has been reported previously in whole saliva^{22,72} and also in other fluids like

plasma⁶¹. Additionally, the tendency of whole saliva proteins and *N*-glycosylated proteins to map to catalytic activity followed by enzyme regulator activity has been demonstrated elsewhere^{22,61,72}. This is not surprising since saliva carries out physiological and digestion functions that require the presence of peptidases, hydrolases and nucleases^{22,72}. The other “*molecular function*” categories of salivary proteins can be found in [Fig # 3].

With regard to the “*biological process*”, the trend observed across the functional categories was very similar. However, there was enrichment of some glycosylated proteins in categories such as immune system process (10%/6%; glycoproteome/proteome) and biological adhesion (6%/2%; glycoproteome/proteome), among others. Different lectin fractionation methods coupled to mass spectrometry have been used for the identification of different immune components based on their ability to bind certain oligosaccharides structures^{16,51,58,73}. Other glycochemistries have been used as well in different fluids^{12,19,22,23,47,61,74}. These results demonstrated the importance of pre-fractionation steps for reducing sample complexity^{15,25} allowing the detection of low-abundant proteins such as complement components¹⁶. Additionally, our results showed that the majority of proteins in both proteomes mapped to cellular process followed by metabolic process. Proteins previously identified in whole saliva and in sera presented this tendency of mapping to the metabolic process^{18,22,72}. For other examples of the “*biological process*” classification of salivary proteins, see [Fig # 3].

Based on the “*cellular component*” classification, the majority of the proteins were assigned as extracellular (54/29%; glycoproteome/proteome). A major isolation of extracellular and cytoplasmic vesicle part proteins was achieved by our lectin-affinity fraction method. This could be explained not only by the capture efficiency attained with this pre-fractionation strategy but also due to PTM changes (i.e. loss of glycosylation) that occurs in whole saliva caused by the presence of proteases, consequently impairing their identification by mass spectrometry²². These findings were in agreement with other studies in which a higher proportion of *N*-glycoproteins were from the extracellular space or plasma followed by membrane proteins^{12,18,19,22,61}. In contrast, a higher percentage of whole saliva proteins mapped to the cytoskeletal part (4/28%; glycoproteome/proteome) in comparison to the lectin-affinity *N*- and *O*-linked glycoproteome. This trend has been observed by others^{2,72}. Identification of membrane proteins was also seen in our glycoproteome. The possible origin of these membrane glycoproteins could be attributed to cellular debris¹⁹ or exosomes⁴. As demonstrated previously^{19,22,23,28}, we also identified proteins from cellular compartments belonging to the cytoplasm and organelles. Other “*cellular component*” classification of salivary proteins can be found in [Fig # 3].

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis

KEGG pathway analysis was performed on the enriched glycoproteome to evaluate which pathways were significantly represented ($p \leq 0.05$). Nineteen different pathways were linked to the SM/SL glycoproteome. Glycoproteins were most commonly involved in processes such as cell communication (18 proteins), complement and coagulation cascades (17 proteins), regulation of actin cytoskeleton (13 proteins), ECM-receptor interaction (12 proteins), and focal adhesion (12 proteins). Other pathways can be seen in Supplemental table #2.

Comparisons between human salivary lectin-affinity *N*- and *O*-linked glycoproteome with other salivary proteomes

A general overview of all the salivary proteomic projects compared is displayed in table #1. These salivary studies underscore the importance of each saliva type, the fractionation

method utilized and the mass spectrometry approaches used for the enrichment of both the salivary proteome and glycoproteome.

Comparison of human salivary lectin-affinity *N*- and *O*-linked glycoproteome with the human salivary *N*-linked glycoproteome in SM and SL salivas

Supplemental table #1 shows that 38 glycoproteins found in the lectin-affinity *N*- and *O*-linked glycoproteome (262 proteins) overlapped with the human salivary *N*-linked glycoproteome (58 proteins) reported elsewhere²². Proteins such as isoform 1 of Kallikrein-11, Golgi membrane protein 1 and isoform 1 of alpha-1-antitrypsin, were common to the two glycoproteomes. In contrast, hypoxia-upregulated protein, cathepsin L, and biotinidase were unique to the human salivary *N*-linked glycoproteome, while fibromodulin, isoform 1 of sulfhydryl oxidase 1 and isoform H17 of myeloperoxidase, were exclusive to the lectin-affinity *N*- and *O*-linked glycoproteome [Fig. # 4A]. Although the salivary secretions used in these two studies are the same, factors such as glycochemistry, salivary collection device, subject's age, sex, and race might account for the observed differences^{21,22,60,75}. In addition, the saliva used in our study was pooled SM/SL whereas individual salivas were collected from the SM and SL glands by Ramachandran et al²².

Comparison of the four salivary glycoproteomes: human salivary lectin-affinity *N*- and *O*-linked glycoproteome versus human salivary *N*-linked glycoproteomes/Glycoproteome of the human salivary proteome, versus human salivary sialome and versus hexapeptide *N*-linked glycoproteome

An overlapping of 21 proteins was observed among the four glycoproteomes compared. Out of the 262 proteins detected in the present study, 61 overlapped [Supplemental table #1] with the human salivary *N*-linked glycoproteomes/Glycoproteome of the human salivary proteome (100 glycoproteins identified in whole, parotid, SM and SL salivas)^{22,23,28, 35} [Supplemental table #1] were the same as those detected in the human salivary sialome (45 proteins found in whole saliva)¹⁹, and 58 [Supplemental table #1] were identified in common to those catalogued in the hexapeptide *N*-linked glycoproteome (187 proteins detected in whole saliva). This suggests, that the methodology used here enhanced the dynamic range of detection⁶¹; and on the other hand, emphasizes that each salivary gland contributes both unique and common components to the overall saliva secretion^{11,22,60}. The overlapping protein identifications were generally abundant salivary proteins such as zinc alpha-2-glycoprotein 1, bactericidal/permeability-increasing protein-like 1, and polymeric immunoglobulin receptor. Overlapping abundant glycosylated proteins have also been observed in human plasma with proteins like alpha-1-acid glycoproteins, apolipoproteins, coagulation factors and proteases⁶¹. Conversely, other abundant proteins like salivary acidic proline-rich phosphoprotein 1/2, basic salivary proline-rich protein 2 and cystatin-C were unique in this study.

We also detected unique medium and low abundant glycoproteins such as kallikrein-10, isoform 2 of *N*-acetylmuramoyl-L-alanine amidase, and isoform Sap-mu-0 of proactivator polypeptide, highlighting the importance of using pre-fractionation steps in complex mixtures for enriching particular subsets of proteins^{16,21-23,44,46,48,60,76,77}. Salivary glycochemistries different from that used here, detected other distinctive subsets of proteins like integrin beta-2, serine protease DESC1, alpha2,3-sialyltransferase VI, and extracellular matrix protein 1, which supplement our findings^{19,22,23,28} [Fig. # 4B].

Comparison of human salivary lectin-affinity *N*- and *O*-linked glycoproteome with the SM/SL human saliva proteome

To gain information about the efficiency of the glycoprotein enrichment strategy employed in this study, we compared the human salivary lectin-affinity *N*- and *O*-linked

glycoproteome (262 proteins) with the SM/SL human saliva proteome (917 proteins) ¹. SM/SL saliva was collected using essentially the same methods in these two studies. This comparison revealed 137 overlapping proteins [Supplemental table #1] and 125 unique glycoproteins not previously found in the SM/SL human saliva proteome such as FUT3 Galactoside 3(4)-L-fucosyltransferase, isoform 1 of semaphorin-4G and isoform 4 of interleukin-1 receptor antagonist protein, highlighting the importance of enriching subproteome components in complex samples ^{15,25}. There were 780 proteins in the SM/SL human saliva proteome that were not detected by lectin-affinity *N*- and *O*-linked glycoproteome. Together, these data confirms the robust and sensitive nature of combining MudPIT analysis ^{56,57} with lectin-affinity chromatography [Fig. # 4C].

Only male saliva was used in the present study. To eliminate sex variation, the lectin-affinity *N*- and *O*-linked glycoproteome data (262 proteins) were compared to the SM/SL human saliva proteome male data (695 proteins) generated by The Scripps Research Institute/ University of Rochester (TSRI/UR) as part of the larger human saliva proteome project ¹. Similar to the comparison with the SM/SL human saliva proteome (137 proteins), 134 proteins in the human salivary lectin-affinity *N*- and *O*-linked glycoproteome overlapped [Supplemental table #1] with the SM/SL male proteome [figure not shown].

Comparison of human salivary lectin-affinity *N*- and *O*-linked glycoproteome with the whole saliva proteome

To further analyze the enrichment of the methodology used in this study, we compared the human salivary lectin-affinity *N*- and *O*-linked glycoproteome data with the whole saliva proteome. The results revealed the identification of 129 unique proteins not found before in the largest dataset obtained from the whole saliva proteome. Among these proteins identified were SEMA4G protein, mucin-5AC (Fragment), isoform 3 of N-acetylated-alpha-linked acidic dipeptidase-like protein and isoform TGN48 of Trans-Golgi network integral membrane protein ². In addition, an overlap of 133 proteins [Supplemental table #1] was observed between the two proteomes, while 2,157 proteins were unique to the whole saliva proteome demonstrating the complexity of this type of saliva due to the contributions of different salivary gland secretions and the presence of other components from blood, plasma and gingival crevicular fluid ⁷² [Fig. # 4D].

Western Blot analysis

To further validate the results of MudPIT analysis, western blotting was performed on the unbound (UF) and eluted (EF) SM/SL fractions. The selectivity of the lectin-affinity chromatography technique for enriching glycoproteins from complex mixtures was demonstrated by the presence of proteins known to be glycosylated in the elution fractions. MudPIT high accuracy and sensitivity allowed the identification of high-, medium-, and low-abundant proteins, emphasizing the overall detection of a dynamic range of glycosylated proteins. This was confirmed by the presence of newly identified, low abundant proteins such as isoform 1 of *N*-acetylated-alpha-linked acidic dipeptidase-like protein (~85 kDa), FUT5 alpha-(1,3)-fucosyltransferase (~43 kDa), isoform 2 of Solute carrier family 12 member 2 (~160 kDa), and isoform TGN46 of Trans-Golgi network integral membrane protein 2 (~57 kDa) [Fig #5A–5D, respectively].

Discussion

The field of biomarker discovery has increased enormously in the past decade in an attempt to find reliable candidates that will aid in the identification of normal biological process, pathogenic states, and pharmacological responses to specific interventions. Human saliva is a unique fluid composed of a wide range of different analytical molecules ^{6,60}. Many of

these analytes are comparable to those found in serum, making saliva an ideal fluid for studying immunological, hormonal, pharmacological, oncological and microbiological biomarkers^{75,78}. Different proteomic strategies have enabled the detection of candidate biomarkers in cells, body fluids and tissues. However, some of the remaining major challenges yet to overcome are sample complexity, dynamic range of protein abundances and heterogeneity^{15,23,27,28,46,60,61}. Taking into consideration that posttranslational modifications contribute to this complex pattern it is necessary to use analytical methodologies and fractionation methods that will allow the study of the subproteome, thus facilitating medium- and low-abundant protein identifications, potential biomarker candidates due to their biological information and diagnostic value^{15,25,27,76}.

Glycosylation is a posttranslational modification implicated in diverse protein, cellular, and biological functions^{17,27,59,70,76,77}. In the oral cavity, well characterized, abundant salivary glycoproteins constitute a major part of the salivary proteome with biological and physiological functions, *e.g.* lubrication, protection, agglutination, digestion, defense, and antibacterial^{13,22,23,28,35,53,60,71,79}.

It is well recognized that alterations in glycosylation are reflected in a variety of pathologies^{16,18,21,25,27,34,35,47,54,58,59,80}. Therefore, it is crucial to do a comprehensive and robust glycoproteomic profiling. Many strategies have been proposed for isolating glycosylated proteins and for finding their *N*-glycosylation sites in different body fluids, tissues and cells^{18,19,22,23,25,27,28,35,46,51,54,58,64,81,82}. Among them, lectins (carbohydrate-binding proteins) have been used widely to fractionate the glycoproteome. Lectins recognize specific carbohydrate motifs, binding is reversible, multiple glycan-binding sites are available, and glycans can be recovered for future characterization, while quantification and chemical derivatization are avoided^{15–18,28,34,35,48,51,58,83}. In addition, lectin-based methodologies have been used for diagnosing pathological states, investigating mutants of model organisms, identifying protein polymorphisms, blood group typing, bacteria typing, saliva-induced aggregation assays of oral bacteria, histochemistry, and quality control analysis of recombinant glycoprotein pharmaceuticals^{26,39,48,55,79,83–85}. Consequently, we reported here a combined approach based on lectin-affinity chromatography and MudPIT mass spectrometry to generate an in-depth profile of the SM/SL glycoproteome. This strategy allowed the identification of both *N*-linked and *O*-linked glycoproteins^{18,58}, not previously done on different types of saliva whether glycosylated proteins were mapped either using a lectin-blot approach^{28,35} or using other glycochemistries allowing the identification of different *N*-glycosylation sites^{19,22,23}.

Of note, the methodology used here identified 125 proteins not previously found in the SM/SL human saliva proteome. The use of lectins like Con A, RCA-I and UEA-I improved the dynamic range of salivary glycoprotein analysis by overcoming the lack of comprehensive capture typically achieved with the use of a single lectin due to its binding selectivity limitations for specific conformations of different carbohydrate moieties^{15,27,51,54}. Con A identified the largest number of glycoproteins (109 out of 262), in comparison to RCA-I and UEA-I, with an overlapping of 38 proteins among the three lectins. This is explained by the broad selectivity of Con-A since it binds α -linked mannose residues, a common feature in the oligosaccharide core of many asparagine-linked glycans^{15,16,18,44,48,58}. It is important to emphasize that RCA-I and UEA-I have a narrower selectivity, thus allowing for diversity enrichment of the salivary glycoproteome³⁶. UEA-I binds fucosyl (α -1,2) galactosyl (β -1,4) *N*-acetylglucosamine (β -1,6)-R and other α -monofucosyl oligomer residues^{18,28}. FUT5 α -1,3-fucosyltransferase which catalyzes α -1,3 glycosidic linkages is an example of proteins isolated by this lectin, and has not been identified before in saliva^{1,12,19,22,23,28,35,72,86}. RCA-I, which has been used previously to analyze parotid saliva³⁵, identified glycoproteins with a primary structure of β -D galactose and *N*-acetyl-

alpha-D-galactosamine^{35,36,39}. Isoform TGN51 of Trans-Golgi network integral membrane protein 2 (protein destined to membrane trafficking) was unique amongst the proteins identified. The glycoproteome enrichment provided by RCA-I could be attributed to not only galactose specificities *per se*, but to its wide range and more complex binding properties for structures with substituted or unsubstituted *N*-acetyllactosamine moieties⁸³.

As expected, there was also some non-specific binding associated with the use of the lectin chromatography^{15,19,58}. However this was minimal as demonstrated by the attained capture efficiency of more than 90%.

It is well known that glycoproteins are valuable biomarkers due to their biological information and diagnostic value^{16,17,21,27,34,82}. However, to take advantage of medium- and low-abundant proteins, pre-fractionation and enrichment steps are essential before mass spectrometry analysis to enhance detection. The combination of lectin-affinity chromatography coupled to MudPIT significantly improved the detection of medium- and low-abundant proteins not previously identified using other glycochemistries in various types of saliva^{19,22,23,28,35}. There are numerous such examples of these glycoproteins such as isoform 5 of Trans-Golgi network integral membrane protein 2, isoform 2 of granulins and isoform 4 of transmembrane protein 132A. In addition, other proteins like interleukin-1 receptor antagonist protein, monocyte differentiation antigen CD14 and bone-derived growth factor were also identified in plasma⁶¹, demonstrating not only the commonality of salivary proteins with plasma proteins but also supporting the diagnostic value of saliva as an alternative test to blood⁷².

Although their origin is unclear, serum glycosylated proteins previously reported in saliva^{1,2,4,22,72} like serotransferrin, hemoglobin subunit beta, and hemopexin were also identified. Salivary proteins such as lactoferrin, IgA, Zn- α glycoprotein and prolactin-induced proteins were identified in the tears as well^{24,64}. These results are consistent with the similarities found previously by others between the lacrimal and salivary glands¹.

One of the major goals of proteomic studies is to define the complete catalogue of salivary proteins in healthy subjects as a basis for understanding the significance of alterations in the saliva protein related to different pathological conditions of either systemic or local origins^{60,72,78,87}. Lectins have been used to successfully target selective glycan structures (*e.g.* degree of branching) associated with particular tumors such as breast and ovarian cancers^{16,34}. Minor changes in *N*-glycosylation sites of vitronectin have been found in serum of patients with ovarian serous carcinoma⁴⁷. Additionally, increased branching of α -1-acid glycoprotein⁸⁰, and lower degree of galactosylation in IgHG *N*-glycan chains^{70,73,74} have also been associated with rheumatoid arthritis. Moreover, in serum of patients with primary Sjögren's syndrome decreased sialylation and increased exposure of galactose on IgA1 glycans have been observed in comparison to healthy controls⁸⁸. Other glycosylated proteins like cathepsin D, clusterin, and α 2-macroglobulin have been linked to Alzheimer disease⁶¹, while up-regulation of ceruloplasmin has also been found in lung cancer sera, in ovarian cancer, hepatocarcinoma and nasopharyngeal carcinoma¹⁸.

Of importance is the multifunctional role of DMBT1 at oral and systemic levels. In the oral cavity, it plays an innate defense function against streptococci and *Helicobacter pylori*, influenza viruses and HIV. Additionally it interacts with mucin-5B and IgA and has an important role in epithelial and cell differentiation²⁴. Meanwhile inactivation of its gene may lead to the formation of brain and salivary gland tumors, and intrahepatic cholangiocarcinoma²⁴. Fibronectins have been elevated in patients with oral squamous cell carcinoma and α -1-antitrypsin was associated with periodontal disease⁶⁰. Patients with

Sjögren's syndrome also shown altered levels of beta-2-microglobulin, polymeric immunoglobulin receptor, cystatin-C and lactoferrin in parotid and whole salivas ^{26,60,89}.

To date, no universal state-of-the-art technique exists that can profile by itself the entire salivary glycoproteome ^{20,23}. Factors, such as type of saliva, time of collection, type of stimulation used, duration of salivary collection and salivary collection device utilized, should be taken into account when cataloguing the salivary glycoproteome, since each salivary gland requires a particular method of collection and contributes its own molecular signature ^{11,22,60,75}.

While many advances have been made in salivary mass spectrometry analysis ⁹⁰, there are clinical limitations with the use of biomarkers for early diagnosis, due to the lack of high sensitivity and specificity. Therefore, quantifying glycoprotein relative abundances, knowing glycan structure functions, analyzing altered glycosylation patterns and monitoring glycosylation changes during disease onset, progression and cure are highly desired to provide an insight into the specific mechanisms involved in the development of different pathologies ^{20,21,24,34,46,49,60,61,80}. It is believed that further analysis of the glycoproteins identified in this study under various pathophysiological conditions would provide valuable information for a more accurate diagnosis, prognosis and treatment of local and systemic diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Drs. David T. Wong, Joseph A. Loo and Weihong Yan (UCLA) for their assistance with the UCLA glycoproteomics data. Cyrus Salehi for preparing the graph related to the GO analysis and helping with the manuscript, and Nicole B-Gonzalez (Friends of Strong/COB) for her excellent technical assistance. This work was supported in part by NIH grants UO1 DE016267 and P41 RR011823 (JRY). BL was supported by a CFFT computational fellowship BALCH05X5 and MG-B is supported by NIH grant KL2RR024136-04.

References

1. Denny P, Hagen FK, Hardt M, Liao L, Yan W, Arellanno M, Bassilian S, Bedi GS, Boontheung P, Cociorva D, Delahunty CM, Denny T, Dunsmore J, Faull KF, Gilligan J, Gonzalez-Begne M, Halgand F, Hall SC, Han X, Henson B, Hewel J, Hu S, Jeffrey S, Jiang J, Loo JA, Ogorzalek Loo RR, Malamud D, Melvin JE, Miroshnychenko O, Navazesh M, Niles R, Park SK, Prakobphol A, Ramachandran P, Richert M, Robinson S, Sondej M, Souda P, Sullivan MA, Takashima J, Than S, Wang J, Whitelegge JP, Witkowska HE, Wolinsky L, Xie Y, Xu T, Yu W, Ytterberg J, Wong DT, Yates JR 3rd, Fisher SJ. The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. *J Proteome Res.* 2008; 7(5):1994–2006. [PubMed: 18361515]
2. Yan W, Apweiler R, Balgley BM, Boontheung P, Bundy JL, Cargile BJ, Cole S, Fang X, Gonzalez-Begne M, Griffin TJ, Hagen F, Hu S, Wolinsky LE, Lee CS, Malamud D, Melvin JE, Menon R, Mueller M, Qiao R, Rhodus NL, Sevinsky JR, States D, Stephenson JL, Than S, Yates JR, Yu W, Xie H, Xie Y, Omenn GS, Loo JA, Wong DT. Systematic comparison of the human saliva and plasma proteomes. *Proteomics: Clin Appl.* 2009; 3(1):116–34. [PubMed: 19898684]
3. Bennick A. Extraoral Functions of Salivary Proteins. *J Oral Biosci.* 2007; 49(1):24–26.
4. Gonzalez-Begne M, Lu B, Han X, Hagen FK, Hand AR, Melvin JE, Yates JR. Proteomic analysis of human parotid gland exosomes by multidimensional protein identification technology (MudPIT). *J Proteome Res.* 2009; 8(3):1304–14. [PubMed: 19199708]
5. Oppenheim FG, Salih E, Siqueira WL, Zhang W, Helmerhorst EJ. Salivary proteome and its genetic polymorphisms. *Ann NY Acad Sci.* 2007; 1098:22–50. [PubMed: 17303824]

6. Schipper RG, Silletti E, Vingerhoeds MH. Saliva as research material: biochemical, physicochemical and practical aspects. *Arch Oral Biol.* 2007; 52(12):1114–35. [PubMed: 17692813]
7. Tabak LA. In defense of the oral cavity: the protective role of the salivary secretions. *Pediatr Dent.* 2006; 28(2):110–7. discussion 92–8. [PubMed: 16708785]
8. Vitorino R, Lobo MJ, Ferrer-Correia AJ, Dubin JR, Tomer KB, Domingues PM, Amado FM. Identification of human whole saliva protein components using proteomics. *Proteomics.* 2004; 4(4): 1109–15. [PubMed: 15048992]
9. Pedersen AM, Bardow A, Jensen SB, Nauntofte B. Saliva and gastrointestinal functions of taste, mastication, swallowing and digestion. *Oral Dis.* 2002; 8(3):117–29. [PubMed: 12108756]
10. Berkovitz, BKB.; Holland, GR.; Moxham, BJ. *Oral Anatomy, Histology & Embriology.* 4. Mosby/ Elsevier; Edinburgh, NY: 2009. *Salivary Glands*; p. 260-77.
11. Hu S, Denny P, Denny P, Xie Y, Loo JA, Wolinsky LE, Li Y, McBride J, Ogorzalek Loo RR, Navazesh M, Wong DT. Differentially expressed protein markers in human submandibular and sublingual secretions. *Int J Oncol.* 2004; 25(5):1423–30. [PubMed: 15492834]
12. Bandhakavi S, Van Riper SK, Tawfik PN, Stone MD, Haddad T, Rhodus NL, Carlis JV, Griffin TJ. Hexapeptide libraries for enhanced protein PTM identification and relative abundance profiling in whole human saliva. *J Proteome Res.* 2011; 10(3):1052–61. [PubMed: 21142092]
13. Helmerhorst EJ, Oppenheim FG. Saliva: a dynamic proteome. *J Dent Res.* 2007; 86(8):680–93. [PubMed: 17652194]
14. Stone MD, Chen X, McGowan T, Bandhakavi S, Cheng B, Rhodus NL, Griffin TJ. Large-scale phosphoproteomics analysis of whole saliva reveals a distinct phosphorylation pattern. *J Proteome Res.* 2011; 10(4):1728–36. [PubMed: 21299198]
15. Butterfield DA, Owen JB. Lectin-affinity chromatography brain glycoproteomics and Alzheimer disease: insights into protein alterations consistent with the pathology and progression of this dementing disorder. *Proteomics: Clin Appl.* 2011; 5(1–2):50–6. [PubMed: 21280237]
16. Calvano CD, Zamboni CG, Jensen ON. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *J Proteomics.* 2008; 71(3):304–17. [PubMed: 18638581]
17. Elschnerbroich S, Kislinger T. Targeted proteomics by selected reaction monitoring mass spectrometry: applications to systems biology and biomarker discovery. *Mol BioSyst.* 2011; 7(2): 292–303. [PubMed: 20976349]
18. Hongsachart P, Huang-Liu R, Sinchaikul S, Pan FM, Phutrakul S, Chuang YM, Yu CJ, Chen ST. Glycoproteomic analysis of WGA-bound glycoprotein biomarkers in sera from patients with lung adenocarcinoma. *Electrophoresis.* 2009; 30(7):1206–20. [PubMed: 19294700]
19. Larsen MR, Jensen SS, Jakobsen LA, Heegaard NH. Exploring the sialome using titanium dioxide chromatography and mass spectrometry. *Mol Cell Proteomics.* 2007; 6(10):1778–87. [PubMed: 17623646]
20. Marino K, Bones J, Kattla JJ, Rudd PM. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol.* 2010; 6(10):713–23. [PubMed: 20852609]
21. Pan S, Chen R, Aebersold R, Brentnall TA. Mass spectrometry based glycoproteomics - from a proteomics perspective. *Mol Cell Proteomics.* 2011; 10(1):R110.003251–1–R110.51–14. [PubMed: 20736408]
22. Ramachandran P, Boontheung P, Pang E, Yan W, Wong DT, Loo JA. Comparison of N-linked Glycoproteins in Human Whole Saliva, Parotid, Submandibular, and Sublingual Glandular Secretions identified using Hydrazide Chemistry and Mass Spectrometry. *Clin Proteomics.* 2008; 4:80–104. [PubMed: 21960768]
23. Ramachandran P, Boontheung P, Xie Y, Sondej M, Wong DT, Loo JA. Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J Proteome Res.* 2006; 5(6):1493–503. [PubMed: 16740002]
24. You J, Fitzgerald A, Cozzi PJ, Zhao Z, Graham P, Russell PJ, Walsh BJ, Willcox M, Zhong L, Wasinger V, Li Y. Post-translation modification of proteins in tears. *Electrophoresis.* 2010; 31(11):1853–61. [PubMed: 20506419]

25. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol.* 2003; 21(6):660–6. [PubMed: 12754519]
26. Carpenter GH, Pankhurst CL, Proctor GB. Lectin binding studies of parotid salivary glycoproteins in Sjogren's syndrome. *Electrophoresis.* 1999; 20(10):2124–32. [PubMed: 10451125]
27. Sun B, Ranish JA, Utleg AG, White JT, Yan X, Lin B, Hood L. Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol Cell Proteomics.* 2007; 6(1):141–9. [PubMed: 17074749]
28. Sondej M, Denny PA, Xie Y, Ramachandran P, Si Y, Takashima J, Shi W, Wong DT, Loo JA, Denny PC. Glycoprofiling of the Human Salivary Proteome. *Clin Proteomics.* 2009; 5(1):52–68. [PubMed: 20161393]
29. Navazesh M. Methods for collecting saliva. *Ann N Y Acad Sci.* 1993; 694:72–7. [PubMed: 8215087]
30. Block P, Brotman S. A method of Submaxillary Saliva collection without cannulation. *N Y State Dent J.* 1962; 28:116–18.
31. Navazesh M, Kumar SK. Measuring salivary flow: challenges and opportunities. *J Am Dent Assoc.* 2008; 139(Suppl):35S–40S. [PubMed: 18460678]
32. Sicher, H.; DuBRUL, EL. *Oral Anatomy.* 6. C.V. Mosby: St. Louis, MO; 1975. The Viscera; p. 192–298.
33. Wolff A, Begleiter A, Moskona D. A novel system of human submandibular/sublingual saliva collection. *J Dent Res.* 1997; 76(11):1782–6. [PubMed: 9372796]
34. Abbott KL, Pierce JM. Lectin-based glycoproteomic techniques for the enrichment and identification of potential biomarkers. *Methods Enzymol.* 2010; 480:461–76. [PubMed: 20816222]
35. Carpenter GH, Proctor GB, Pankhurst CL, Linden RW, Shori DK, Zhang XS. Glycoproteins in human parotid saliva assessed by lectin probes after resolution by sodium dodecyl sulphate-polyacrylamide gel electrophoresis. *Electrophoresis.* 1996; 17(1):91–7. [PubMed: 8907524]
36. Debray H, Decout D, Strecker G, Spik G, Montreuil J. Specificity of twelve lectins towards oligosaccharides and glycopeptides related to N-glycosylproteins. *Eur J Biochem.* 1981; 117(1):41–55. [PubMed: 7262089]
37. Green ED, Brodbeck RM, Baenziger JU. Lectin affinity high-performance liquid chromatography. Interactions of N-glycanase-released oligosaccharides with Ricinus communis agglutinin I and Ricinus communis agglutinin II. *J Biol Chem.* 1987; 262(25):12030–9. [PubMed: 3624246]
38. Haynes PA, Aebersold R. Simultaneous detection and identification of O-GlcNAc-modified glycoproteins using liquid chromatography-tandem mass spectrometry. *Anal Chem.* 2000; 72(21):5402–10. [PubMed: 11080893]
39. Sharon N, Lis H. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology.* 2004; 14(11):53R–62R.
40. Xu T, Venable JD, Kyu Park S, Cociorva D, Lu B, Liao L, Wohlschlegel J, Hewel J, Yates JR 3rd. ProLuCID, a fast and sensitive tandem mass spectra-based protein identification program. *Mol Cell Proteomics.* 2006; 5:S174.
41. Tabb DL, McDonald WH, Yates JR 3rd. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res.* 2002; 1(1):21–6. [PubMed: 12643522]
42. Eng J, McCormack A, Yates JR 3rd. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom.* 1994; 5(11):976–89.
43. Parvin MN, Gerelsaikhan T, Turner RJ. Regions in the cytosolic C-terminus of the secretory Na(+)-K(+)-2Cl(−) cotransporter NKCC1 are required for its homodimerization. *Biochemistry (Mosc).* 2007; 46(33):9630–7.
44. Novotny MV, Mechref Y. New hyphenated methodologies in high-sensitivity glycoprotein analysis. *J Sep Sci.* 2005; 28(15):1956–68. [PubMed: 16276785]
45. Anumula KR. High-sensitivity and high-resolution methods for glycoprotein analysis. *Anal Biochem.* 2000; 283(1):17–26. [PubMed: 10929803]

46. Chen R, Jiang X, Sun D, Han G, Wang F, Ye M, Wang L, Zou H. Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J Proteome Res.* 2009; 8(2):651–61. [PubMed: 19159218]
47. Liu Z, Cao J, He Y, Qiao L, Xu C, Lu H, Yang P. Tandem ¹⁸O stable isotope labeling for quantification of N-glycoproteome. *J Proteome Res.* 2010; 9(1):227–36. [PubMed: 19921957]
48. Madera M, Mechref Y, Novotny MV. Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides. *Anal Chem.* 2005; 77(13):4081–90. [PubMed: 15987113]
49. Rexach JE, Rogers CJ, Yu SH, Tao J, Sun YE, Hsieh-Wilson LC. Quantification of O-glycosylation stoichiometry and dynamics using resolvable mass tags. *Nat Chem Biol.* 2010; 6(9):645–51. [PubMed: 20657584]
50. Soper AS, Aird SD. Elution of tightly bound solutes from concanavalin A Sepharose. Factors affecting the desorption of cottonmouth venom glycoproteins. *J Chromatogr A.* 2007; 1154(1–2):308–18. [PubMed: 17449042]
51. Yang Z, Hancock WS, Chew TR, Bonilla L. A study of glycoproteins in human serum and plasma reference standards (HUPO) using multilectin affinity chromatography coupled with RPLC-MS/MS. *Proteomics.* 2005; 5(13):3353–66. [PubMed: 16052617]
52. Anumula KR, Dhume ST. High resolution and high sensitivity methods for oligosaccharide mapping and characterization by normal phase high performance liquid chromatography following derivatization with highly fluorescent anthranilic acid. *Glycobiology.* 1998; 8(7):685–94. [PubMed: 9621109]
53. Koyama I, Komine S, Yakushijin M, Hokari S, Komoda T. Glycosylated salivary alpha-amylases are capable of maltotriose hydrolysis and glucose formation. *Comp Biochem Physiol Part B: Biochem Mol Biol.* 2000; 126(4):553–60.
54. Cao J, Shen C, Wang H, Shen H, Chen Y, Nie A, Yan G, Lu H, Liu Y, Yang P. Identification of N-glycosylation sites on secreted proteins of human hepatocellular carcinoma cells with a complementary proteomics approach. *J Proteome Res.* 2009; 8(2):662–72. [PubMed: 19196183]
55. Ernst B, Magnani JL. From carbohydrate leads to glycomimetic drugs. *Nat Rev Drug Discovery.* 2009; 8(8):661–77.
56. Washburn MP. Utilization of proteomics datasets generated via multidimensional protein identification technology (MudPIT). *Briefings Funct Genomic Proteomic.* 2004; 3(3):280–6.
57. Yates JR 3rd. Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct.* 2004; 33:297–316. [PubMed: 15139815]
58. Jung K, Cho W, Regnier FE. Glycoproteomics of plasma based on narrow selectivity lectin affinity chromatography. *J Proteome Res.* 2009; 8(2):643–50. [PubMed: 19099503]
59. Merry AH, Merry CL. Glycoscience finally comes of age. *EMBO Rep.* 2005; 6(10):900–3. [PubMed: 16222235]
60. Hu S, Loo JA, Wong DT. Human saliva proteome analysis and disease biomarker discovery. *Expert Rev Proteomics.* 2007; 4(4):531–8. [PubMed: 17705710]
61. Liu T, Qian WJ, Gritsenko MA, Camp DG 2nd, Monroe ME, Moore RJ, Smith RD. Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J Proteome Res.* 2005; 4(6):2070–80. [PubMed: 16335952]
62. Rao RS, Buus OT, Wollenweber B. Evolutionary Pattern of N-Glycosylation Sequon Numbers in Eukaryotic ABC Protein Superfamilies. *Bioinform Biol Insights.* 2010; 4:9–17. [PubMed: 20213012]
63. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology.* 2002; 12(4):43R–56R.
64. Zhou L, Beuerman R, Chew AP, Koh SK, Cafaro T, Urrets-Zavalía E, Urrets-Zavalía J, Li S, Serra H. Quantitative Analysis of N-linked Glycoproteins in Tear Fluid of Climatic Droplet Keratopathy by Glycopeptide Capture and iTRAQ. *J Proteome Res.* 2009; 8:1992–2003. [PubMed: 19714880]
65. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis.* 1998; 19(11):1853–61. [PubMed: 9740045]

66. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta*. 1999; 1473(1):4–8. [PubMed: 10580125]
67. Roth J. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. *Chem Rev*. 2002; 102(2):285–303. [PubMed: 11841244]
68. Gerelsaikhan T, Parvin MN, Turner RJ. Biogenesis and topology of the secretory Na⁺-K⁺-2Cl⁻-cotransporter (NKCC1) studied in intact mammalian cells. *Biochemistry (Mosc)*. 2006; 45(39):12060–7.
69. Morelle W, Canis K, Chirat F, Faïd V, Michalski JC. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics*. 2006; 6(14):3993–4015. [PubMed: 16786490]
70. Axford J, Kieda C, van Dijk W. Meeting report--Jenner 5: glycobiology and medicine. *Glycobiology*. 2001; 11(2):5G–7G.
71. Silletti E, Vitorino RM, Schipper R, Amado FM, Vingerhoeds MH. Identification of salivary proteins at oil-water interfaces stabilized by lysozyme and beta-lactoglobulin. *Arch Oral Biol*. 2010; 55(4):268–78. [PubMed: 20197185]
72. Loo JA, Yan W, Ramachandran P, Wong DT. Comparative human salivary and plasma proteomes. *J Dent Res*. 2010; 89(10):1016–23. [PubMed: 20739693]
73. Stadlmann J, Weber A, Pabst M, Anderle H, Kunert R, Ehrlich HJ, Peter Schwarz H, Altmann F. A close look at human IgG sialylation and subclass distribution after lectin fractionation. *Proteomics*. 2009; 9(17):4143–53. [PubMed: 19688751]
74. Wührer M, Stam JC, van de Geijn FE, Koeleman CA, Verrips CT, Dolhain RJ, Hokke CH, Deelder AM. Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum. *Proteomics*. 2007; 7(22):4070–81. [PubMed: 17994628]
75. Henson BS, Wong DT. Collection, storage, and processing of saliva samples for downstream molecular applications. *Methods Mol Biol*. 2010; 666:21–30. [PubMed: 20717775]
76. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol*. 2003; 21(3):255–61. [PubMed: 12610572]
77. Zhao J, Simeone DM, Heidt D, Anderson MA, Lubman DM. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. *J Proteome Res*. 2006; 5(7):1792–802. [PubMed: 16823988]
78. Scarano E, Fiorita A, Picciotti PM, Passali GC, Calo L, Cabras T, Inzitari R, Fanali C, Messina I, Castagnola M, Paludetti G. Proteomics of saliva: personal experience. *Acta Otorhinolaryngol Ital*. 2010; 30(3):125–30. [PubMed: 20948587]
79. Zhang XS, Proctor GB, Garrett JR, Schulte BA, Shori DK. Use of lectin probes on tissues and sympathetic saliva to study the glycoproteins secreted by rat submandibular glands. *J Histochem Cytochem*. 1994; 42(9):1261–9. [PubMed: 8064133]
80. Alavi A, Axford JS. Sweet and sour: the impact of sugars on disease. *Rheumatology (Oxford)*. 2008; 47(6):760–70. [PubMed: 18375404]
81. Zeng Y, Ramya TN, Dirksen A, Dawson PE, Paulson JC. High-efficiency labeling of sialylated glycoproteins on living cells. *Nat Methods*. 2009; 6(3):207–9. [PubMed: 19234450]
82. Zielinska DF, Gnad F, Wisniewski JR, Mann M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell*. 2010; 141(5):897–907. [PubMed: 20510933]
83. Iskratsch T, Braun A, Paschinger K, Wilson IB. Specificity analysis of lectins and antibodies using remodeled glycoproteins. *Anal Biochem*. 2009; 386(2):133–46. [PubMed: 19123999]
84. Hu S, Wong DT. Lectin Microarray. *Proteomics: Clin Appl*. 2009; 3:148–54. [PubMed: 21132067]
85. Mirth DB, Miller CJ, Kingman A, Bowen WH. Inhibition of saliva-induced aggregation of *Streptococcus mutans* by wheat germ agglutinin. *Caries Res*. 1979; 13(3):121–31. [PubMed: 286641]
86. Bandhakavi S, Stone MD, Onsongo G, Van Riper SK, Griffin TJ. A dynamic range compression and three-dimensional peptide fractionation analysis platform expands proteome coverage and the diagnostic potential of whole saliva. *J Proteome Res*. 2009; 8(12):5590–600. [PubMed: 19813771]

87. Ferraccioli G, De Santis M, Peluso G, Inzitari R, Fanali C, Bosello SL, Iavarone F, Castagnola M. Proteomic approaches to Sjogren's syndrome: a clue to interpret the pathophysiology and organ involvement of the disease. *Autoimmun Rev.* 2010; 9(9):622–6. [PubMed: 20462525]
88. Dueymes M, Bendaoud B, Pennec YL, Youinou P. IgA glycosylation abnormalities in the serum of patients with primary Sjogren's syndrome. *Clin Exp Rheumatol.* 1995; 13(2):247–50. [PubMed: 7656472]
89. Ryu OH, Atkinson JC, Hoehn GT, Illei GG, Hart TC. Identification of parotid salivary biomarkers in Sjogren's syndrome by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry and two-dimensional difference gel electrophoresis. *Rheumatology (Oxford).* 2006; 45(9):1077–86. [PubMed: 16522680]
90. Zhang L, Xiao H, Wong DT. Salivary biomarkers for clinical applications. *Mol Diagn Ther.* 2009; 13(4):245–59. [PubMed: 19712004]
91. Lashley KS. Reflex secretion of the human parotid gland. *J Exp Psychol.* 1916; 1(6):461–93.

Synopsis

Submandibular/Sublingual saliva (SM/SL) is composed of a mixed secretion of serous and mucous components making it an attractive fluid for glycoproteomic analysis. Glycosylation is a common posttranslational modification involved in different biological processes and disease development. Con A, RCA-I and UEA-I lectin-affinity chromatographies in conjunction to MudPIT mass spectrometry were used for cataloguing the glycoproteins present in SM/SL secretion. 262 *N*- and *O*-linked proteins were identified. From these, 125 proteins were not previously found in the SM/SL human saliva proteome, demonstrating the power of these analytical techniques for profiling the salivary glycoproteome.

B RCA-I Lectin Chromatography

C UEA-I Lectin Chromatography

Figure 1. Lectin-affinity chromatograms of SM/SL glycoproteins

Enrichment of SM/SL glycoproteins was accomplished using [A] *Canavalia ensiformis* (Con A), [B] *Ricinus communis* (RCA-I) and [C] *Ulex europaeus* (UEA-I) agarose columns. Chromatograms showed the unbound fraction (flow through) of non-glycosylated proteins after using the corresponding equilibration/washing buffers (0.05 M Tris-0.15 M NaCl-0.004 M CaCl₂, pH 7.0 for Con A; and 0.01 M Phosphate-0.15 M NaCl pH 7.2 for RCA-I and UEA-I) for each lectin column. Glycoprotein elution profiles for the moderate-desorbed (peak #1) and strongly-bound or retarded proteins (peak #2) were detected in the elution fraction after using two-different concentrations of competitive eluents with different pHs [Con A: 0.2 M (pH 7.2) and 0.5 M (pH 2.5) methyl α -D-mannopyranoside; RCA-I: 0.1

M (pH 7.2) and 0.25 M (pH 2.5) α -lactose; and UEA-I: 0.05 M (pH 7.2) and 0.5 M (pH 2.5) α -L-fucose buffers].

CON A Chromatography
Total glycoprotein identifications 194

RCA-I Chromatography **UEA-I Chromatography**
Total glycoprotein identifications 97 Total glycoprotein identifications 96

Figure 2. Comparison of SM/SL glycoprotein identifications among the three lectin-affinity chromatography methods

Only 38 (14.5%) proteins overlapped among the three lectins used. 109 (41.6%) proteins were exclusively isolated by Con A chromatography, 41 (15.6%) were separated using RCA-I chromatography and 25 (9.5%) were fractionated by UEA-I chromatography.

Figure 3. Gene Ontology annotation of the lectin-affinity *N*- and *O*-linked glycoproteome and whole saliva proteome. GO annotation by molecular function

Mapping of both proteomes by molecular function indicated that the highest percentage of proteins (48%/54%; glycoproteome/proteome) was associated with binding activities and the lowest (2%/1%; glycoproteome/proteome) with antioxidant activity. **GO annotation by biological process.** Allocation of proteins by biological process showed that the greatest percentage of these proteins was mapped to cellular process (20%/27%; glycoproteome/proteome) and the smallest (5%/3%; glycoproteome/proteome) to multi-organism process. **GO annotation by cellular component.** Distribution of proteins by cellular component demonstrated that the majority were mapped to the extracellular region part (54%/29%; glycoproteome/proteome), while the minority was located in the small ribosomal subunit (1%/3%; glycoproteome/proteome).

Figure 4. Comparisons between human salivary lectin-affinity *N*- and *O*-linked glycoproteome with other salivary proteomes

A. Venn diagram showing the overlap of proteins between the human salivary lectin-affinity *N*- and *O*-linked glycoproteome and the SM and SL human salivary *N*-linked glycoproteome²². 38 (13.5%) proteins were found in both proteomes, whereas 224 (79.4%) were unique to the lectin-affinity *N*- and *O*-linked glycoproteome and 20 (7.1%) were distinctive to the *N*-glycoproteome. **B.** Venn diagram showing the overlap of proteins between the human salivary lectin-affinity *N*- and *O*-linked glycoproteome, human salivary *N*-linked glycoproteomes/Glycoprofile of the human salivary proteome (whole saliva, parotid saliva, submandibular saliva and sublingual saliva)^{22,23,28}, the human salivary sialome in whole saliva¹⁹ and the hexapeptide *N*-linked glycoproteome in whole saliva¹². Only 21 (3.9%) proteins overlapped among the four glycoproteomes, whereas 241 (45.4%) were unique to the human salivary lectin-affinity *N*- and *O*-linked glycoproteome, 79 (14.9%) were distinctive to the *N*-linked glycoproteomes/Glycoprofile of the human salivary proteome in different types of saliva, 24 (4.5%) were exclusive to the human salivary sialome and 166 (31.3%) were only in the hexapeptide *N*-linked

glycoproteome. **C. Venn diagram showing the overlap of glycoproteins between the lectin-affinity *N*-and *O*-linked glycoproteome and the SM/SL human saliva proteome**¹. 137 (13.1%) proteins overlapped between the two proteomes, while 125 (12%) were unique to the lectin-affinity glycoproteome and 780 (74.9%) were exclusive to the SM/SL human saliva proteome. **D. Venn diagram showing the overlap of glycosylated proteins between the lectin-affinity *N*-and *O*-linked glycoproteome and the whole saliva proteome**⁷². 133 (5.5%) proteins overlapped between the two proteomes, while 129 (5.3%) were only found in the lectin-affinity *N*-and *O*-linked glycoproteome and 2,157 (89.2%) were exclusive to the whole saliva proteome.

Figure 5. Western blot analyses of SM/SL glycoproteins isolated from Unbound (UF) and Eluted (EF) fractions

Panel A: Isoform 1 of *N*-acetylated- α -linked acidic dipeptidase-like protein was detected using goat polyclonal antibody to *NAALADL1* (E-18) (Santa Cruz Biotechnology, Inc., Santa Cruz, CA). The observed molecular weight for this protein was around ~85 kDa. **Panel B:** Alpha-(1,3)-fucosyltransferase was detected with rabbit polyclonal antibody against *FUT5* (Abgent, Inc., San Diego, CA). This antibody recognized a protein of ~43 kDa. **Panel C:** Isoform 2 of Solute carrier family 12 member 2 (*NKCC1*) protein was detected using 2ndMet rabbit antibody against amino acids 209–285 of rat *NKCC1* fusion protein. The molecular weight of this protein was ~160 kDa. **Panel D:** Isoform TGN46 of Trans-Golgi network integral membrane protein 2 was identified with rabbit polyclonal antibody to *TGN46* (Abcam, Inc., Cambridge, MA). The expected molecular weight for this glycoprotein was ~57 kDa.

Table #1

Human Salivary Projects compared.

Project Name	Author and Affiliation	Biological Source	Collection Method	Processing Method	MS/MS Analysis	Number of Identifications
Human salivary Lectin-affinity <i>N</i> - and <i>O</i> -linked glycoproteome	Gonzalez et al. (TSRI/UR) ^A	Combined stimulated- SM/SL saliva.	SM/SL gland collector ^a	Lectin-affinity chromatography.	MudPIT and LC-ESI-MS/MS Nano-2D-LC-MS-MS Tandem mass spectra (646,516)	5,963 peptides. 262 Glycoproteins
Human salivary <i>N</i> -linked glycoproteome	Ramachandran et al. ^b (UCLA) ^B	Unstimulated/Resting-whole saliva.	Draining method ^{29,31 c,d}	Hydrazide chemistry ^e and PNGase F release method. In-solution-IEF (Zoom-IEF fractionation). 1D or 2D-PAGE analysis.	LC-ESI-MS/MS Nano-LC-MS-MS	84 <i>N</i> -glycosylated peptides. 45 <i>N</i> -glycoproteins. 44 New sites of <i>N</i> -glycosylation.
Human Salivome	Larsen et al. ^f (SDU) ^C	Unstimulated/Resting- whole saliva. Human plasma from healthy controls v/s. bladder cancer patients.	Draining method	High-affinity Titanium dioxide in combination with phosphatase treatment toward Sialic acid residues. ¹⁸ O and ¹⁶ O Labeling for tryptic peptides. <i>N</i> -glycosidase F-digestion.	MALDI-MS Nanoscale LC- MS/MS	45 salivary SA-containing glycoproteins. 97 SA-glycosylation sites in saliva. 29 New glycosylation sites. 100 plasma SA-containing proteins. 192 SA-glycosylation sites.
Human salivary <i>N</i> -linked glycoproteome	Ramachandran et al. ^g (UCLA)	Unstimulated/resting- whole saliva. Stimulated- parotid saliva (PA). Individual Stimulated- SM and SL salivas.	Draining method Parotid gland collector ^h SM/SL gland collector ⁱ	Modified hydrazide chemistry ^j and PNGase F release method. Solution-IEF fractionation (Zoom-IEF fractionation).	1D and 2D-LC-MS/MS equipped with nano-electrospray	156 <i>N</i> -glycosylated peptides representing 77 unique <i>N</i> -glycoproteins identified in the salivary fluids. 122/62 <i>N</i> -glycopeptides and glycoproteins in WS. 62/34 <i>N</i> -glycopeptides and glycoproteins in PA. 80/44 <i>N</i> -glycopeptides and glycoproteins in SM. 98/53 <i>N</i> -glycopeptides and glycoproteins in SL.
Glycoproteome of the Human Salivary Proteome	Sondej et al. ^k (UCLA)	Unstimulated/resting-whole saliva.	Draining method	2D-PAGE analysis (Gels stained with Sypro Ruby stain or Pro-Q Emerald 488 stain. Biotinylated-Lectin blotting (Spots visualized with avidin D-alkaline phosphatase. Lectin spots ranked as having low, medium, high and no reactivity to the lectins used. Spots from the Pro-Q Emerald stained- gels and lectin blots	LC-MS/MS equipped with nano electrospray performed on Sypro Ruby stained spots.	166 2D-gel spots identified (most abundant salivary proteins). From these spots only 158 spots reacted with the lectins used. 50 new proteins were identified from the Sypro Ruby stained spots.

Project Name	Author and Affiliation	Biological Source	Collection Method	Processing Method	MS/MS Analysis	Number of Identifications
Human Saliva Proteome	Denny et al. ¹ (NIDCR-Research supported Consortium) ²	Stimulated- Parotid Saliva. Combined Stimulated- SM/SL Saliva (TSRI/UR). Individual-stimulated SM and SL salivas (UCLA/USC)	Parotid gland collector <i>g</i> (TSRI/UR; UCLA/USC; UCSF). SM/SL gland collector <i>a</i> (TSRI/UR; UCSF). SM/SL gland collector <i>i</i> (UCLA/USC)	<p>were matched to the Sypro Ruby identified stained spots.</p> <p>TSRI/UR: Both salivas were pre-fractionated using:</p> <ul style="list-style-type: none"> Size fractionation (Tangential flow ultrafiltration) or size exclusion chromatography on Sephadex G100/G200 or Sepharose CL-6B. Anion-exchange chromatography (DE52, DEAE Sepharose CL-6B, or DEAE Ceramic HyperD® F). Reversed-phase high-pressure LC (C₄ resin). Immobilized metal chelate affinity chromatography. Hydrophobic charge interaction chromatography (MEP Hypercel). 1D and 2D SDS-PAGE. ZOOM-IEF. <p>UCLA/USC:</p> <ul style="list-style-type: none"> 2D SDS-PAGE Reversed-phase high-pressure LC (C₄ resin). ZOOM-IEF. <p>UCSF:</p> <ul style="list-style-type: none"> 2D SDS-PAGE 	MudPIT-LC-ESI-MS/MS	In total the three research groups identified 11,592 distinct peptides sequences matching 2,153 distinct proteins. After data integration: A total of 1166 non-redundant protein identifications were obtained, with 914 proteins found in parotid saliva and 917 found in SM/SL saliva.

Project Name	Author and Affiliation	Biological Source	Collection Method	Processing Method	MS/MS Analysis	Number of Identifications
Whole saliva Proteome	Loo et al. ^m (UCLA)	Yan et al. Unstimulated/Resting- whole saliva.	Draining method	<p>Compendium of Yan et al. ⁿ and Bandhakavi et al. ^o studies.</p> <p>Yan et al. 4 Research groups contributed:</p> <p><i>University of Minnesota (UMN)</i>;</p> <ul style="list-style-type: none"> • IEF • PolySULFOETHYL strong cation exchange. <p><i>Research Triangle Institute (RTI)</i>;</p> <ul style="list-style-type: none"> • IPG-IEF-strips • Oasis HLB SPE resin. <p><i>Calibrant Biosystems/University of Maryland (CB/UM)</i>;</p> <ul style="list-style-type: none"> • Transient capillary isotachopheresis/CZE (CITP/CZE). <p><i>University of California-Los Angeles (UCLA)</i>;</p> <ul style="list-style-type: none"> • Prefractionation using ultracentrifugation filters • Zoom IEF 	<p>LC-QqTOF MS equipped with nanoelectrospray interface.</p> <p>µLC-MS/MS</p> <p>LC-MS/MS Nano-2D-LC</p> <p>Nano-RP-LC-MS/MS</p> <p>LC-MS/MS equipped with nanoelectrospray interface.</p>	<p>Yan et al. 11,893 peptides and 1,444 proteins.</p>

Project Name	Author and Affiliation	Biological Source	Collection Method	Processing Method	MS/MS Analysis	Number of Identifications
Hexapeptide N-linked glycoproteome	Bandhakavi et al. <i>P</i> (UMN) ^E	Bandhakavi et al. Unstimulated/Resting- whole saliva.	Draining method	<p>Bandhakavi et al. <i>University of Minnesota</i> (UMN);</p> <ul style="list-style-type: none"> Dynamic range compression using Proteominer or Library-2 hexapeptide beads. 2D and 3D-peptide fractionation (combination of OFFGEL-IEF, strong cation exchange and microcapillary reverse liquid chromatography). 	μLC-MS/MS	<p>Bandhakavi et al. 9,010 unique peptides 2,340 proteins 497 new WS protein identifications.</p> <p>Loo et al. Integration of both studies rendered a total of 26,430 peptides and 2,290 proteins identified.</p>
				<ul style="list-style-type: none"> Dynamic range compression using Proteominer or Library-2 hexapeptide beads. No Dynamic range compression Hydrazide Chemistry ^j Strong cation exchange-HPLC 	μLC-MS/MS	268 N-glycosites 193 N-glycoproteins

J Proteome Res. Author manuscript; available in PMC 2012 November 1.

^ATSRI/UC: The Scripps Research Institute/University of Rochester. SM/SL= Submandibular/Sublingual saliva. MudPIT= multidimensional protein identification technology. LC-ESI-MS/MS=Liquid chromatography electrospray ionization tandem mass spectrometry.

^aSM/SL collector ³⁰.

^BUCLA: University of California, Los Angeles. IEF= isoelectric focusing. 1D or 2D= one or two dimension. PNGase F= Peptide N-Glycosidase F. WS= whole saliva.

^{b, g}Ramachandran's studies ^{22,23};

^{c,d}Collection method ^{29,31};

^ehydrazide method ²⁵;

^hParotid collector ⁹¹;

- i*_{SM/SL} collector 33;
- j*_{modified hydrazide method} 27;
- k*_{Sondej's study} 28;
- m*_{Loo's study} 72;
- n*_{Yan's study} 2;
- o*_{Bandhakavi's study} 86.
- C*_{SDU}: University of Southern Denmark. SA= Sialic acid. MALDI= matrix-assisted laser desorption/ionization mass spectrometry.
- f*_{Larsen's study} 19.
- D*_{National Institute of Dental and Craniofacial Research (NIDCR)-Consortium: Integrated by three research-supported groups: 1) TSRI/UR, 2) University of California-Los Angeles/University of Southern California (UCLA/USC) and 3) University of California-San Francisco (UCSF).}
- I*_{Denny's study} 1.
- E*_{UMN}: University of Minnesota.
- P*_{Bandhakavi's study} 12.