Published in final edited form as:
J Proteome Res. 2009 December ; 8(12): 5590–5600. doi:10.1021/pr900675w.

A dynamic range compression and three-dimensional peptide fractionation analysis platform expands proteome coverage and the diagnostic potential of whole saliva

Sricharan Bandhakavi[1], **Matthew D. Stone**[1], **Getiria Onsongo**[1], **Susan K. Van Riper**[1], and **Timothy J. Griffin**[1]

[1]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, 6-155 Jackson Hall, 321 Church Street SE., Minneapolis MN 55455, USA

Abstract

Comprehensive identification of proteins in whole human saliva is critical for appreciating its full diagnostic potential. However, this is challenged by the large dynamic range of protein abundance within this fluid. To address this problem, we used an analysis platform that coupled hexapeptide libraries for dynamic range compression (DRC) with three-dimensional (3D) peptide fractionation. This approach identified 2340 proteins in whole saliva and represents the largest saliva proteomic dataset generated using a single analysis platform. Three dimensional peptide fractionation involving sequential steps of preparative IEF, strong cation exchange, and capillary reversed phase liquid chromatography was essential for maximizing gains from DRC. Compared to saliva not treated with hexapeptide libraries, DRC substantially increased identified proteins across physicochemical and functional categories. Approximately 20% of total salivary proteins are also seen in plasma, and proteins in both fluids show comparable functional diversity and disease-linkage. However, for a subset of diseases, saliva has higher apparent diagnostic potential. These results expand the potential for whole saliva in health monitoring/diagnostics and provide a general platform for improving proteomic coverage of complex biological samples.

Keywords

Dynamic range compression; Hexapeptide libraries; 3D-peptide fractionation; Saliva; Shotgun proteomics

Introduction

A major thrust within the field of clinical proteomics is the cataloguing of proteins within different body fluids such as plasma/serum, urine, cerebrospinal fluid, tears, seminal fluid, amniotic fluid, and saliva [1]. Body fluids are in contact either with specific organs and tissues or circulate through the entire organism, into which cellular proteins are either shed or secreted

Corresponding author: Timothy J. Griffin, Ph.D., Associate Professor, Associate Director, Center for Mass Spectrometry and Proteomics, tgriffin@umn.edu, Tel: 612-624-5249; Fax: 612-624-0432.

Two collated scaffold files generated after 2D or 3D peptide fractionation during this study are available for viewing at Tranche (ProteomeCommons.Org). The freely available Scaffold viewer is required for viewing this data and can be downloaded from Proteome Software, Inc. (http://www.proteomesoftware.com/). The data associated with this manuscript may be downloaded from ProteomeCommons.org Tranche, https://proteomecommons.org/tranche/, using the following hash:
AGV8tQVVVCFHWThjxtXBzBcsaiuFv+AqIW7El6n741ZVR20SCdGJvZkKctHm3pR
DqlGmbXesmjBmYawVB6Zn0KEkkw4AAAAAAAACGw==

[2–5]. In a variety of diseases and during therapy, altered signal transduction pathways result in a distinct profile of proteins that are released into these fluids. These protein biomarkers can be invaluable for early diagnosis of a variety of diseases, evaluating treatment regimens, and disease prognosis.

Among human body fluids with health diagnostic/monitoring potential, whole saliva is gaining increasing attention. Whole saliva is produced by the parotid, submandibular and sublingual glands into the oral cavity; it maintains oral health, protects from foreign micro-organisms, and facilitates chewing/swallowing of food [2–5]. Interestingly, molecular constituents in plasma (the current gold standard in biomarker studies, [6,7]) pass into whole saliva resulting in significant overlap between proteins found in both fluids [8]. Thus, in addition to containing biomarkers of oral health, the salivary proteome holds the potential for predicting systemic diseases such as breast cancer [9], diabetes [10], and autoimmune disorders [11,12]. These features, combined with the fact that whole saliva is easily collected in a non-invasive manner, make it an attractive fluid for biomarker discovery.

Although not as severe as in plasma, proteomic characterization of saliva is also complicated by its broad dynamic range of protein abundance. Coomassie staining identifies approximately ten proteins in whole saliva, which account for nearly 98% of total salivary protein [13]. These high abundance proteins obscure the detection of proteins present at much lower concentrations. To cope with this challenge, diverse protein/peptide fractionation and mass spectrometric approaches have been employed for maximizing proteome coverage of this fluid. Combining data from multiple labs using a variety of analysis platforms, greater than 1400 proteins in whole saliva have been identified so far [8]. However, it remains unclear how many additional proteins might exist within saliva and how its expanded proteome catalogue compares to other biological fluids such as plasma.

Combinatorial chemistry derived hexapeptide libraries have been demonstrated to effectively compress dynamic range and increase proteomic coverage of complex biological samples [14–17]. In this approach, complex proteomes are incubated with millions of diverse hexapeptides which effectively act as affinity capture ‘baits’, each having specificity for a single or small number of distinct proteins. Because each hexapeptide has a finite binding capacity for the protein(s) it binds, the most abundant proteins exceed available binding capacity, causing excess unbound protein to ‘flow through,’ resulting in their depletion. Meanwhile, lower abundance proteins are completely bound by the hexapeptide resin and eluted, effectively enriching these proteins in the eluate. The end result of this treatment is a compression of the dynamic range, reducing the relative amount of the highly abundant proteins while increasing the relative amount of low abundance proteins. Despite the demonstrated effectiveness of hexapeptide bead libraries in other contexts, their potential benefits for expanding coverage of the salivary proteome have not yet been explored.

Most studies using hexapeptide libraries have traditionally relied on 1D-or 2D-gel fractionation of samples, followed by in-gel tryptic digestion prior to mass spectrometric analysis [14–17]. Given the higher sensitivity offered by gel-free shotgun proteomics [18,19], there is a need to develop workflows that incorporate DRC with advanced peptide fractionation methodologies, for maximizing proteomic coverage of complex biological samples, such as human saliva.

In this report, we investigated the use of hexapeptide libraries for DRC, integrated with 3D peptide fractionation combining preparative IEF, strong cation exchange (SCX) and microcapillary reversed-phase liquid chromatography (μLC) [20], for increasing proteomic coverage of whole saliva. Our approach generated a high confidence dataset of 2340 salivary proteins, which to our knowledge is the largest dataset of saliva proteins identified using a

single analysis platform. A similar scheme should be generally applicable to expanding proteomic coverage of several other complex biological samples.

Experimental Methods

Saliva collection and processing

Prior to saliva collection volunteers did not eat or drink for 90 minutes, and lightly rinsed their mouth with water. Fresh saliva was collected from 6 healthy volunteers (3 males and 3 females) into beakers chilled on ice. The saliva was pooled and clarified by centrifugation twice at 12,000 g at 4°C for ten minutes each spin. After supplementing with protease inhibitors (Roche Complete EDTA-free), saliva was either boiled in equal volume of 100mM Tris-Cl, pH 6.8, 4% SDS, 10% mercaptoethanol, and 20% glycerol and labeled as 'Untreated Saliva' or processed for treatment with hexapeptide libraries (without boiling).

Treatment of saliva with Proteominer or Library-2 hexapeptide beads

Library-2 was a generous gift from Kate Smith at BioRad (Hercules, CA), while Proteominer was purchased from BioRad (Hercules, CA). Protein concentration was determined before treatment with hexapeptide libraries by the BCA (Pierce, Rockford, IL) method. For initial optimization experiments, 25mg, 50mg, and 75 mg protein samples from fresh saliva were tumbled for 18 hours overnight at room temperature with 100μL Proteominer beads along with protease inhibitors (Roche complete EDTA-free). We chose to incubate saliva for 18 hours with Proteominer beads (instead of the standard recommendation of 2 hours as per the manufacturer), since it was found to recover more protein species in preliminary experiments (as monitored by coomassie staining). Post-incubation, samples were spun in a swinging bucket rotor at 1000g to pellet beads, and the supernatant was labeled as 'Flow through' fraction. Beads were washed thrice in standard PBS (Phosphate buffered saline), and eluted three times by boiling in a modified elution buffer (100mM Tris-Cl, pH 6.8, 4% SDS, 10% mercaptoethanol, and 20% glycerol). Flow through fractions were also boiled in the elution buffer and stored at −80°C until further use. Experiments using library-2 treatment were done by incubating 18.75 mg saliva with 25 μL beads (maintaining the same ratio as incubating 75mg protein and 100 μL Proteominer beads). The lower amount was used because of a limited supply of Library-2. Flow through and eluate fractions were processed identically as described above for Proteominer treatment.

Sample processing for SDS-PAGE, protein staining, and trypsin digestion

Untreated Saliva, or 'Eluates' and 'Flow throughs' obtained after hexapeptide treatment, were precipitated using 4 volumes of ice-cold acetone per volume of protein sample. Precipitation was allowed to proceed overnight at −20°C, and precipitates pelleted by centrifugation. Pellets were dissolved in 50mM Tris, pH 8.0, 5mM TE pH 8.0 containing 0.5% SDS and protein amounts estimated using the BCA protein assay.

For loading protein samples on SDS-PAGE gels, samples containing equal amount of total protein were supplemented with SDS-sample buffer, and after electrophoresis, visualized by coomassie staining, unless indicated otherwise. For trypsin digestion, dissolved protein pellets were diluted 10-fold in 50mM Tris, pH 8.0, 5mM TE pH 8.0 without any SDS to bring total SDS levels to 0.05%. DTT was added to a final concentration of 5mM and trypsin digestion performed overnight at 37°C at a 1:25 (enzyme:substrate) ratio. ProQ and Sypro Ruby (phosphorylation and total protein) staining reagents were purchased from Invitrogen (Invitrogen Corporation, Carlsbad, CA) and manufacturers suggested protocols were followed.

2D- and 3D-peptide fractionation

After tryptic digestion, peptides were purified and concentrated using MCX (mixed mode) cartridges from Waters and resolved using the 3100 OFFGEL instrument (Agilent Technologies, Santa Clara, CA) with a 24-well set up (pH 3–10) as per manufacturer's instructions. Peptide electrofocusing was performed at a maximum current of 50 μA and maximum power of 200 milliwats, for 50 kV-h. After focusing was completed, some fractions were pooled (as indicated in figures) and further purified using 'stage' tips [21] containing 2 plugs of 'Empore' disks (3M, Minneapolis, MN). Samples were them either analyzed directly by mass spectrometry (2D-peptide fractionation) or further fractionated by SCX (strong cation exchange) prior to mass spectrometric analysis (3D peptide fractionation).

Strong cation exchange (SCX) chromatography was performed on a Magic 2002 HPLC system coupled with a Magic Variable Splitter set at position R4 (Michrom BioResources, Inc., Auburn, CA) using a Polysulfoethyl A column, 150 mm length × 1.0 mm ID, 5 μm particles, 200 Å pore size (PolyLC Inc., Columbia, MD). The sample was dissolved in 250 μL of SCX buffer A (20% v/v acetonitrile, 10 mM KH_2PO_4 pH 2.7 with phosphoric acid) and loaded onto the column. Peptides were eluted at a flow rate of 35 μL/min with a gradient of 0–20% buffer B (20% v/v ACN, 10 mM KH_2PO_4 pH 3.0, 500 mM KCl) over 45 minutes followed by a gradient from 20–100% buffer B for 15 minutes. During the chromatography run, absorbance at 215 nm and 280 nm was monitored, fractions were collected at 3-min intervals and each fraction was vacuum centrifuged to dryness.

Capillary LC-MS/MS

Peptides were dissolved in 5 μL of load buffer (98:2:0.01 water: acetonitrile: formic acid). They were loaded directly onto a 13-cm × 100 μm fused silica pulled-tip capillary column packed in-house with Magic C18AQ- 5-μm, 200 Angstrom pore size resin (Michrom Bioresources Inc., Auburn, CA) with load buffer at a flow rate of 1000 nL/min using an Eksigent 1DLC nanoflow system and a MicroAS autosampler. Peptides were eluted using a gradient of 2–40% acetonitrile in 0.1% formic acid over 90 min with a constant flow of 250 nL/min. The column was mounted in a nanospray source directly in line with an LTQ-Orbitrap XL mass spectrometer (ThermoFisher Scientific).

Spray voltage was at 1.75 kV, and the heated capillary was maintained at 160 °C. The orbital trap was set to acquire survey mass spectra (*m*/*z* 360–1800) with a resolution of 60,000 at *m*/*z* 400 with a target value set to 1E6 ions or 500 ms. The 8 most intense ions from the full scan were selected for fragmentation by collision-induced dissociation (normalized collision energy - 35%) in the LTQ ion trap with automatic gain control settings of 5,000 ions or 100 ms concurrent to full-scan acquisition in the orbital trap. For enhanced mass accuracy, the lock mass option was enabled for real-time calibration using the following polysiloxane peaks of m/z 371.1012, 445.1200, and 519.1388. Precursor ion charge state screening was enabled, and all unassigned charge states as well as singly charged species were rejected. We used dynamic exclusion set to a maximum of 500 entries with a maximum retention period of 90 s and mass window of −0.6 to 1.2 amu. Data were acquired using Xcalibur software.

Database searching and Data processing

Data generated from μLC-MS/MS analysis (total of 7,80,285 MS/MS scans) was searched with Sequest V27.0 against a composite database consisting of the NCBI human database V200806 (70,711 entries) and its reversed database. Search parameters used were partial tryptic digestion and variable modification of 15.9949 Da on methionine. Sequest output was organized and peptide probabilities were calculated through Peptide Prophet[22] using Scaffold (Proteome Software Inc., Portland, OR). Peptide identifications were filtered using full tryptic digest specificity, and precursor mass tolerance of <10 ppm employed to generate

false positive rates below 1%. False positive rate at peptide level was determined using the equation, false positive rate % = $[2n_{reverse}/(n_{forward}+n_{reverse})]\times100$, where $n_{forward}$ equals number of peptide/protein sequence matches from the forward database, and $n_{reverse}$ equals number of reversed database matches [23]. False discovery rate at protein and peptide levels were measured using the formula: FDR = 100×[Reverse database matches/(Reverse database matches+Forward database matches)] [24,25]. As part of the Scaffold software, the reported proteins were also subjected to assignment of a protein probability, based on the Protein Prophet [26] program, in order to minimize peptides matching redundantly to proteins and to report the minimal number of unique proteins represented by our data.

Bioinformatic analysis

Proteins were analyzed using Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Inc., Redwood City, CA). Core analysis was performed using direct and indirect relationships. Using Fishers' Exact test, core comparison analysis was used to compare functional diversity of plasma and saliva proteins. Annotation data was downloaded from IPA and loaded into a local installation of MySQL database. Gene symbols, obtained from IPA, were used for Gene Ontology and membership analysis.

Using software previous developed by our group [27], plasma proteins (available at http://www.bioinformatics.med.umich.edu/hupo/ppp, http://www.ebi.ac.uk/pride, and http://www.peptideatlas.org) and saliva proteins were separately analyzed to determine their distribution in SLIM Gene Ontology categories. SQL operators were then used to determine protein membership in different datasets.

Results and Discussion

DRC optimization using whole saliva proteins

Unstimulated, whole saliva was obtained from healthy volunteers, clarified by centrifugation (see Experimental Methods), and used for studies outlined in this report. For DRC of saliva, we initially tested the effect of incubating increasing amounts of saliva protein (25, 50, and 75 mg protein respectively) with 100µl of Proteominer hexapeptide library beads (Figure 1A).

As shown in Figure 1A, Untreated Saliva contained a handful of proteins detectable by commassie staining, which represent the most abundant species within this fluid. Proteominer treatment depleted most of these proteins substantially; simultaneously, it resulted in appearance of several new bands that were previously undetectable by coomassie staining (see Eluate25/50/75 lanes). Most of these bands were roughly comparable in their intensities in Eluate25, −50, and −75. However, bands that correspond to those detectable in untreated saliva (*i.e.*, most abundant salivary proteins), were progressively reduced in Eluate50 and −75 fractions, and increased in their corresponding flow through fractions, respectively (see asterisks in Figure 1A). Given that Eluate75 recovered the same amount of total protein as Eluate50 and −25 (1mg each as same amount of protein binding capacity/beads were used in each experiment), this suggested that loading increased protein amounts provided more efficient binding of lower abundance proteins and depletion of high abundance proteins. Hence, we chose Eluate75 for mass spectrometric analysis.

DRC increased proteins identified in saliva, but sample complexity still limited depth of coverage

Prior to mass spectrometric analysis, 100µg of protein from Untreated Saliva and Eluate75 were trypsinized, purified, and separated into 24 fractions using isoelectric focusing on an OFFGEL instrument. Fractions expected to be 'peptide-sparse' based on previous reports using the OFFGEL instrument [28–30] were pooled (as indicated in Figure 1B, **left panel**), resulting

in a total of 20 fractions for analysis by μLC-MS/MS on an LTQ-Orbitrap mass spectrometer. As shown in Figure 1B (**right panel**), Proteominer treatment increased the number of total unique peptides identified across every fraction. Combining data from all twenty runs, we identified a total of 251 proteins and 2238 peptides in Untreated Saliva versus 693 proteins and 5277 peptides in Proteominer Eluate75 (Figure 1C). False positive rates maintained here and throughout this paper are below 1% at peptide level, and calculated as described in Experimental methods. For additional stringency, we used mass accuracy as a filter [31,32] and all peptide-hits were identified within 10ppm accuracy in the Orbitrap.

Two-dimensional (2D) peptide fractionation (preparative IEF and μLC) revealed significant benefits of DRC on identification of proteins (and peptides) in saliva. However, the total protein identifications were substantially lower than those obtained by more recent exhaustive single analyses detecting >1200 proteins in whole or ductal saliva [33,34], or >1400 total proteins identified in whole saliva by combining results from multiple studies using several different analysis platforms [8]. We reasoned that in spite of its benefits, DRC combined with 2D peptide fractionation was not sufficient for maximal proteomic coverage of this fluid due to saliva's extreme complexity.

3D peptide fractionation enhanced gains in proteome coverage enabled by DRC

To increase proteins identified in saliva post-DRC, we incorporated SCX fractionation as an intermediate step in between OFFGEL preparative IEF and μLC fractionation. A recent report by our group demonstrated that IEF and offline step-elution SCX are orthogonal separation methods, and that their combination can result in substantial improvements in proteomic coverage of cellular lysates [20]. For the present study, rather than offline step-elution SCX we chose offline gradient SCX HPLC of pooled IEF fractions, as it is less manual and should provide better separation of peptide mixtures.

To combine preparative IEF and SCX HPLC fractionation across the entire dataset (Figure 2A), we initially separated 200μg of peptides from Untreated Saliva and Proteominer Eluate75 by OFFGEL preparative IEF. Next, to maintain roughly equal number of peptides per pool, we combined OFFGEL fractions 1 to 6 (most acidic peptides), 7 to 14, and 15 to 24 (basic peptides). Each of these pools was further resolved by SCX fractionation using an automated HPLC gradient (see Experimental Methods). The amount of peptides fractionated by OFFGEL was doubled (200μg compared to 100μg used in 2D-analysis) as we anticipated having 40–50 fractions (twice the number analyzed previously) after SCX fractionation for LC-MS/MS analysis.

When compared with 2D fractionation, incorporation of an intermediate SCX fractionation step, resulted in a ~5-fold increase in total proteins identified in Untreated Saliva, from 251 to 1226 proteins (Figure 2B). Combining data from 2D peptide fractionation, we identified a total of 1236 proteins in Untreated Saliva. There was a ~3-fold increase in proteins identified in Proteominer-treated saliva to 1919 proteins. Combined with 2D analysis, total identifications in Proteominer-treated Saliva were 1950 proteins. Interestingly, 304 proteins in Untreated Saliva were not identified in Proteominer Eluate75. Thus, when all 2D and 3D analyses were combined, we identified a total number of 2254 proteins in whole saliva (Figure 2B and 2C).

By approximation, DRC and 3D peptide fractionation contributed equally to this total number of 2254 proteins. Among these, 932 proteins were detected by 3D peptide fractionation without the need for Proteominer treatment, and were common to both Proteominer Eluate75 and Untreated Saliva. However, an additional 1019 proteins could only be identified after both DRC and 3D peptide fractionation (see Figure 2C). Compared to previous results (Figure 1C), this illustrates the importance of 3D peptide fractionation for maximizing gains from DRC of complex biological samples, such as saliva.

DRC using an alternate hexapeptide library modestly expanded proteomic coverage of whole saliva

Recently, a carboxylated derivative of Proteominer (also known as Library-2) has been developed and shown to increase proteins identified by ~10% [16]. To further expand the salivary proteome, whole saliva was treated with Library-2 beads, and proteins eluted (identically as done for Proteominer treatment) for SDS-PAGE and mass spectrometric analysis. SDS-PAGE indicated that Library-2 and Proteominer depleted abundant proteins from saliva similarly (compare flow through fractions from each treatment; Figure 3A), but had distinct enrichment patterns of lower abundance proteins (compare eluate fractions from each treatment). As shown in Figure 3A, many more proteins were enriched to the point of being detectable by coomassie staining in eluate from Proteominer versus Library-2.

Because Untreated Saliva and Proteominer Eluate75 were extensively analyzed (>60 μLC-MS/MS runs were performed for each sample when combining both 2D and 3D analysis), we reasoned that additional proteins identified after Library-2 treatment would be modest. We expected that a majority of proteins unique to Library-2 should be identified after 2D peptide fractionation, and hence elected to not perform an extensive 3D peptide fractionation for this sample.

As in previous experiments, 100μg of peptides from Library-2 eluate was fractionated by OFFGEL and analyzed by μLC-MS/MS (Figure 1). In contrast to results from Proteominer, which consistently increased unique peptides across every OFFGEL fraction analyzed, a more variable effect was seen with Library-2 eluates (Figure 3B). In some fractions, total unique identifications were comparable to or lower than the number seen in Untreated Saliva (1& 2, 6, 17& 18, and 19, for example). In other fractions however, Library-2 eluate yielded even more peptides than in Proteominer eluate (3, 4, 5, and most dramatically in 23& 24, for example). After combining results from all fractions, Library-2 eluate analysis yielded a total of 832 proteins and ~4000 unique peptide identifications.

Combining results from both Proteominer and Library-2 analysis, we identified a total of 2072 proteins (Figure 3C, left panel). Since both samples were obtained after DRC of saliva, we refer to this pool as the post-DRC dataset/sample through the rest of this report. Library-2 added a relatively modest 122 proteins not seen in Proteominer Eluate75 samples. Thus, the commercially available Protoeminer hexapeptide library effectively enriches the vast majority of proteins that are present in whole saliva. In spite of these gains, 268 proteins seen in Untreated Saliva remained unidentified in the post-DRC dataset (Figure 3C, **right panel**). Similar results have been seen for previous analyses using hexapeptide libraries as mentioned in a recent review [14]. In line with these findings, our results suggest that not all proteins within saliva are sufficiently enriched by the existing hexapeptide diversity in the libraries used.

Adding unique proteins from Untreated Saliva to the post-DRC dataset, yielded the final number of 2340 salivary proteins identified in this study (Figure 3C, **right panel**). Out of these, 1395 proteins (60% of total) were identified at 2-peptides or more, and 945 proteins (40% of total) were identified based on single peptides. Among the 945 singe peptide hits, 163 proteins were identified by multiple MS/MS scans, and single MS/MS scans resulted in the 782 remaining identifications.

As mentioned previously, each individual analysis (2D fractionation or 3D fractionation and with or without DRC) was filtered at or below 1% false positive rate at the peptide level and merged to result in the total 2340 protein identifications. As an additional criterion for data quality, mass filtering [31,32] was used to ensure that we only included those identifications resulting from within 10 ppm mass accuracy of the precursor ions. In the final dataset of 2340

proteins, false discovery rate (FDR) was estimated to be 0.44% at peptide level, and 2.7% at protein level. Since most of our protein identifications (>95%) were from our 3D-analysis platform (either Untreated Saliva or Proteominer Eluate75), we also estimated false discovery rates in the 3D-dataset alone. From a total of 27,049 peptides and 2243 proteins identified by 3D analysis, estimated false discovery rate (FDR) was 0.49% at peptide level and 2.4% at the protein level. Taken together, these results indicate that our dataset represents a high confidence catalogue of salivary proteins.

The relevant analyses that yielded these 2340 proteins, their corresponding peptides, and all other relevant mass spectrometric information are shown in Supplementary Table 1. A listing of salivary proteins and their protein descriptions are included in Supplementary Table 2. Finally, the entire dataset of both single and multiple peptide identifications is viewable/downloadable as two Scaffold .sfd files for 2D and 3D analyses from the Tranche system at ProteomeCommons.org. The freely available Scaffold viewer required for viewing these results may be downloaded from Proteome Software Inc (http://www.proteomesoftware.com/).

DRC increases proteomic coverage within whole saliva across physicochemical and Gene Ontological categories

Although DRC introduced substantial gains, we wanted to determine if the hexapeptide library based approach introduced any potential biases in our analyses, i.e., if specific classes or types of proteins might have been enriched/identified post-DRC. Towards these goals, we estimated the molecular weight (MW) and isoelectric point (pI) of all proteins identified post-DRC and compared their distributions with those seen for Untreated Saliva. Both parameters were estimated using the ‘Compute pI/MW’ tool on EXPASY server (http://www.expasy.ch/tools/). As shown in Figure 4A, increased protein identifications were seen in post-DRC saliva across the entire range of predicted protein molecular weights and isoelectric points, indicating absence of any significant physicochemical bias introduced by hexapeptide libraries.

Untreated Saliva contained a majority of proteins with acidic isoelectric points (Figure 4A). Also, a significant number of proteins were below 40 kDa (647 proteins; ~52% of total) or 60 kDa (928 proteins; ~75% of total). In line with the increased coverage across both categories, post-DRC Saliva was enriched for acidic proteins (1465 proteins below isoelectric point of 7; ~73% of total), or those with small to medium molecular weights (1479 proteins below 60 kDa; ~74% of total). The median molecular weight and pI of proteins identified in Untreated Saliva was 39.7 kDa and 6.1 versus 43 kDa and 6.1 in post-DRC saliva, respectively, indicating very slight physicochemical differences between the two samples. The mean molecular weight and isoelectric point in the pooled dataset of total salivary proteins were 42.8 kDa and 6.1, respectively. Interestingly, even after the identification of several new proteins, the physicochemical profile of our expanded salivary dataset is in line with a recent comparative analysis using a smaller number of consensus saliva proteins [8]. Thus saliva contains a preponderance of acidic pI and smaller molecular weight proteins.

The untreated and post-DRC saliva proteins were further compared in terms of their slim Gene Ontology (GO) categories (Cellular component, Biological processes and Molecular function) to gain a better understanding of saliva protein functions and independently test for any potential bias/selectivity within post-DRC saliva. A large number of proteins in all samples (~50%) grouped into either ‘Other’ (a large assortment of functions with few proteins per functional categories) or ‘Unknown’ classes/categories. Since these do not significantly add to our understanding of saliva, only significantly represented and well defined classes are shown in Figure 4B. A survey of these categories clearly demonstrated that post-DRC saliva was enriched across essentially all of them. The notable exception was ‘plasma membrane’

proteins (within GO cellular component) wherein analysis of Untreated Saliva yielded 84 proteins versus 76 proteins identified post-DRC (Figure 4B). Although this category was not the most extensively populated, it might suggest that hydrophobic proteins (such as those associated with plasma membrane) might not be enriched to the same extent as soluble proteins using Proteominer and Library-2. In spite of the slightly lower total numbers, post-DRC analyses identified 22 new plasma membrane proteins not seen in Untreated Saliva, resulting in a total of 106 salivary proteins in this category.

Consistent with saliva being a 'secreted' fluid, the major GO cellular component categories in saliva are cytoplasmic, organellar and extracellular proteins. With regard to GO biological processes, the largest set of proteins grouped into protein metabolic categories. In GO categories of molecular function (Figure 4B, **bottom panel**), saliva contained a preponderance of proteins involved in catalytic and protein binding functions. An interesting line of future investigation would be to determine the panoply of catalytic activities and protein-protein interactions that contribute to the biological functions of saliva.

Comparison to previous studies of Saliva

A compilation of previous results from multiple laboratories, generated a 'consensus' listing of 1444 proteins within whole saliva, and 1939 proteins from whole and ductal saliva [8]. To determine the efficiency of our current approach, we compared our total 2340 salivary protein identifications against the consensus collection of 1444 previous whole saliva proteins identified at 2-peptides or greater.

482 proteins from this collection of previously identified salivary proteins were not identified in our current study. Adding these to our total salivary identifications indicates the existence of greater than 2800 unique proteins in saliva. Thus, the proteomic complexity is substantially more than previously thought.

Although we cannot unambiguously explain why we did not identify some proteins within the consensus listing of saliva proteins, we offer a few potential reasons. These include possible differences in saliva samples themselves, sample collection or processing, and potential losses due to insufficient enrichment by hexapeptide beads. Further, even after DRC and 3D-fractionation, individual fractions might be sufficiently complex resulting in peptide under-sampling from single μLC-MS/MS runs (as done for this report). Thus, simply repeating mass spectrometric analyses might be expected to capture more of the missing proteins. Inspite of these 'misses', 1454 proteins not present in this consensus dataset were identified by our approach.

Comparison of whole saliva proteins to those in plasma revealed comparable functional diversity and highlights the diagnostic potential of saliva

Plasma is the current gold-standard fluid for diagnostic/prognostic investigations. Compared to plasma, saliva is dilute (protein concentration of ~1mg/ml versus >50mg/ml for plasma), and appears to contain fewer total proteins. The 'core' plasma dataset contains 3020 proteins identified at 2-peptides or more [35]. Given our expansion of the saliva proteome and its diagnostic prospects, we sought to compare this expanded salivary dataset against this proteomic listing of plasma proteins.

A recent report that compared the consensus of 1444 whole saliva proteins with plasma found ~33% overlap between the two proteomes [8]. In contrast, only 463 proteins, ~20% of 2340 total whole saliva proteins presented in this report, are also seen in plasma. We interpret this result to indicate that our approach has identified several new proteins that are possibly unique

to saliva, or are below the current detection limits in plasma. Identified salivary proteins that are unique to saliva (i.e., not seen in plasma) are presented in Supplementary Table S3.

Next, we sought to compare the proteomes of plasma and saliva for their functional diversity. We utilized Ingenuity Pathway Analysis (IPA) and the previously released list of 3020 plasma proteins [35] for comparison against our total salivary protein dataset of 2340 proteins. Saliva and plasma contained a diverse spectrum of molecular/cellular functions (Figure 5A). Both fluids also contained functional categories that were unique to each. For example, proteins within gene expression, cellular response to therapeutics, cell cycle, amino acid metabolism, vitamin and mineral metabolism, and energy production were unique to plasma; Free radical scavenging, post-translational modification, and protein folding functions appeared unique to whole saliva. Interestingly, several of the functions that were common to both fluids were more highly represented in saliva than in plasma.

Given the role of the observed protein functions in human health and disease, we compared the two fluids using the IPA 'disease function' analysis to gauge their disease diagnostic potential. As shown in Figure 5B, both saliva and plasma contained a significant number of proteins that are implicated in a range of human diseases/health conditions. When compared with saliva, plasma appeared to contain more proteins involved in neurological, cardiovascular, genetic, skeletal & muscular, metabolic and endocrine system diseases/disorders. In contrast, saliva was more enriched for proteins implicated in cancers, gastrointestinal, inflammatory, infectious, and respiratory diseases. Contrary to expectations, saliva contained more proteins that are involved in hematological disease than plasma. A potential basis for this observation is the possibility that blood-borne diseases are linked with several proteins that reside outside of blood, and plasma is not as enriched as saliva for these 'extra-hematological' molecules. Taken together these findings illustrate a much broader diagnostic potential of human whole saliva than previously anticipated.

Recent analyses have suggested that the presence of plasma proteins in saliva makes it suitable for diagnosing/monitoring systemic health conditions [8]. Given that plasma is in contact with all tissues within the human body, potential leaking of its constituents undoubtedly contributes to the 'systemic' diagnostic potential of saliva. However, the results shown in Figure 5B (especially for cancers, gastrointestinal, inflammatory, infectious and respiratory diseases where saliva has a higher number of implicated proteins than plasma) indicate strongly that the systemic diagnostic potential of saliva cannot be explained by the presence of known plasma proteins alone. Since only ~20% of total salivary proteins are seen in plasma, proteins unique to saliva must also contribute significantly to its 'system-wide' diagnostic potential. Our observations suggest that at the very least, the combined use of saliva and plasma will increase effectiveness for diagnostic and prognostic investigations for a variety of health conditions.

Different human diseases are characterized by dysregulated extra-, inter-, and/or intracellular signal transduction networks. Our analysis has expanded the salivary proteome to contain several new proteins that regulate these events, and provides a molecular framework within which expanded health monitoring/diagnostic studies using this fluid can be conducted in the future.

Conclusion

In this report, we have developed an analysis platform integrating DRC with 3D peptide fractionation for expanding proteomic coverage of human whole saliva. Two different hexapeptide libraries, Proteominer and its carboxylated derivative, Library-2 were used in our analysis and led to the generation of a high confidence dataset of 2340 proteins within whole saliva. DRC and 3D peptide fractionation cooperatively maximized the depth of protein

identification, demonstrating that the complexity of the whole saliva proteome necessitates integrated methods for its comprehensive analysis. DRC increased proteins identified without any significant biases towards specific physiochemical or biological functional categories. Several new proteins were identified that enhance the applicability of using saliva as a diagnostic/prognostic indicator of human health. Additionally, this report provides an analysis platform capable of expanding proteomic inventories in complex biological samples using shotgun proteomics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded in part by NIH grant 1R01DE017734. We are thankful to Dr. Weihong Yan (University of California, Los Angeles) for providing us with a list of consensus salivary proteins and plasma proteins for comparison with our datasets. Kate Smith at BioRad laboratories (Hercules, CA) graciously shared library-2 during the course of our studies. We also thank members of the Griffin laboratory, Leann Higgins, and Todd Markowski of the Proteomics Core Facility for helpful comments, advice, and SCX fractionations. Minnesota Supercomputing Institute at the University of Minnesota members, John Chilton, Mark Nelson, and Pratik Jagtap provided computational support.

References

1. Li SJ, et al. Sys-BodyFluid: a systematical database for human body fluid proteome research. Nucleic Acids Res 2009;37(Database issue):D907–D912. [PubMed: 18978022]
2. Zhang H, et al. Mass spectrometric detection of tissue proteins in plasma. Mol Cell Proteomics 2007;6 (1):64–71. [PubMed: 17030953]
3. Defabianis P, Re F. [The role of saliva in maintaining oral health]. Minerva Stomatol 2003;52(6):301–308. [PubMed: 12874534]
4. Humphrey SP, Williamson RT. A review of saliva: normal composition, flow, and function. J Prosthet Dent 2001;85(2):162–169. [PubMed: 11208206]
5. Turner RJ, Sugiya H. Understanding salivary fluid and protein secretion. Oral Dis 2002;8(1):3–11. [PubMed: 11936453]
6. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002;1(11):845–867. [PubMed: 12488461]
7. Anderson NL, et al. The human plasma proteome: a nonredundant list developed by combination of four separate sources. Mol Cell Proteomics 2004;3(4):311–326. [PubMed: 14718574]
8. Wieihong Yan RA, Balgley BrianM, Boontheung Pinmanee, et al. Systematic comparison of the human saliva and plasma proteomes. Proteomics-Clinical Applications 2009;3(1):116–134. [PubMed: 19898684]
9. Streckfus CF, et al. Breast cancer related proteins are present in saliva and are modulated secondary to ductal carcinoma in situ of the breast. Cancer Invest 2008;26(2):159–167. [PubMed: 18259946]
10. Rao PV, et al. Proteomic identification of salivary biomarkers of type-2 diabetes. J Proteome Res 2009;8(1):239–245. [PubMed: 19118452]
11. Hu S, et al. Salivary proteomic and genomic biomarkers for primary Sjogren's syndrome. Arthritis Rheum 2007;56(11):3588–3600. [PubMed: 17968930]
12. Rujner J, et al. Serum and salivary antigliadin antibodies and serum IgA anti-endomysium antibodies as a screening test for coeliac disease. Acta Paediatr 1996;85(7):814–817. [PubMed: 8819547]
13. Messana I, et al. Facts and artifacts in proteomics of body fluids. What proteomics of saliva is telling us? J Sep Sci 2008;31(11):1948–1963. [PubMed: 18491358]
14. Boschetti E, Righetti PG. The art of observing rare protein species in proteomes with peptide ligand libraries. Proteomics 2009;9(6):1492–1510. [PubMed: 19235170]

15. Guerrier L, Righetti PG, Boschetti E. Reduction of dynamic protein concentration range of biological extracts for the discovery of low-abundance proteins by means of hexapeptide ligand library. Nat Protoc 2008;3(5):883–890. [PubMed: 18451796]

16. Roux-Dalvai F, et al. Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry. Mol Cell Proteomics 2008;7 (11):2254–2269. [PubMed: 18614565]

17. Thulasiraman V, et al. Reduction of the concentration difference of proteins in biological liquids using a library of combinatorial ligands. Electrophoresis 2005;26(18):3561–3571. [PubMed: 16167368]

18. Gygi SP, et al. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci U S A 2000;97(17):9390–9395. [PubMed: 10920198]

19. Hubner NC, Ren S, Mann M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. Proteomics 2008;8(23–24):4862–4872. [PubMed: 19003865]

20. Xie H, et al. Proteomics analysis of cells in whole saliva from oral cancer patients via value-added three-dimensional peptide fractionation and tandem mass spectrometry. Mol Cell Proteomics 2008;7 (3):486–498. [PubMed: 18045803]

21. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser desorption/ ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem 2003;75 (3):663–670. [PubMed: 12585499]

22. Keller A, et al. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 2002;74(20):5383–5392. [PubMed: 12403597]

23. Peng J, et al. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2003;2(1):43–50. [PubMed: 12643542]

24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc 1995;57:289–300.

25. Reidegeld KA, et al. An easy-to-use Decoy Database Builder software tool, implementing different decoy strategies for false discovery rate calculation in automated MS/MS protein identifications. Proteomics 2008;8(6):1129–1137. [PubMed: 18338823]

26. Nesvizhskii AI, et al. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 2003;75(17):4646–4658. [PubMed: 14632076]

27. Onsongo, GaX; H; Griffin, TJ.; Carlis, J. Generating GO Slim Using Relational Database Management Systems to Support Proteomics Analysis. 21st IEEE International Symposium on Computer-Based Medical Systems, 2008; Finland. 2008.

28. Chenau J, et al. Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling. Proteome Sci 2008;6:9. [PubMed: 18302743]

29. de Godoy LM, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 2008;455(7217):1251–1254. [PubMed: 18820680]

30. Horth P, et al. Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. Mol Cell Proteomics 2006;5(10):1968–1974. [PubMed: 16849286]

31. Perry RH, Cooks RG, Noll RJ. Orbitrap mass spectrometry: instrumentation, ion motion and applications. Mass Spectrom Rev 2008;27(6):661–699. [PubMed: 18683895]

32. Yates JR, et al. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. Anal Chem 2006;78(2):493–500. [PubMed: 16408932]

33. Denny P, et al. 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]

34. Guo T, et al. Characterization of the human salivary proteome by capillary isoelectric focusing/ nanoreversed-phase liquid chromatography coupled with ESI-tandem MS. J Proteome Res 2006;5 (6):1469–1478. [PubMed: 16739998]

35. Omenn GS, et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 2005;5(13):3226–3245. [PubMed: 16104056]

36. Messana I, et al. Trafficking and postsecretory events responsible for the formation of secreted human salivary peptides: a proteomics approach. Mol Cell Proteomics 2008;7(5):911–926. [PubMed: 18187409]

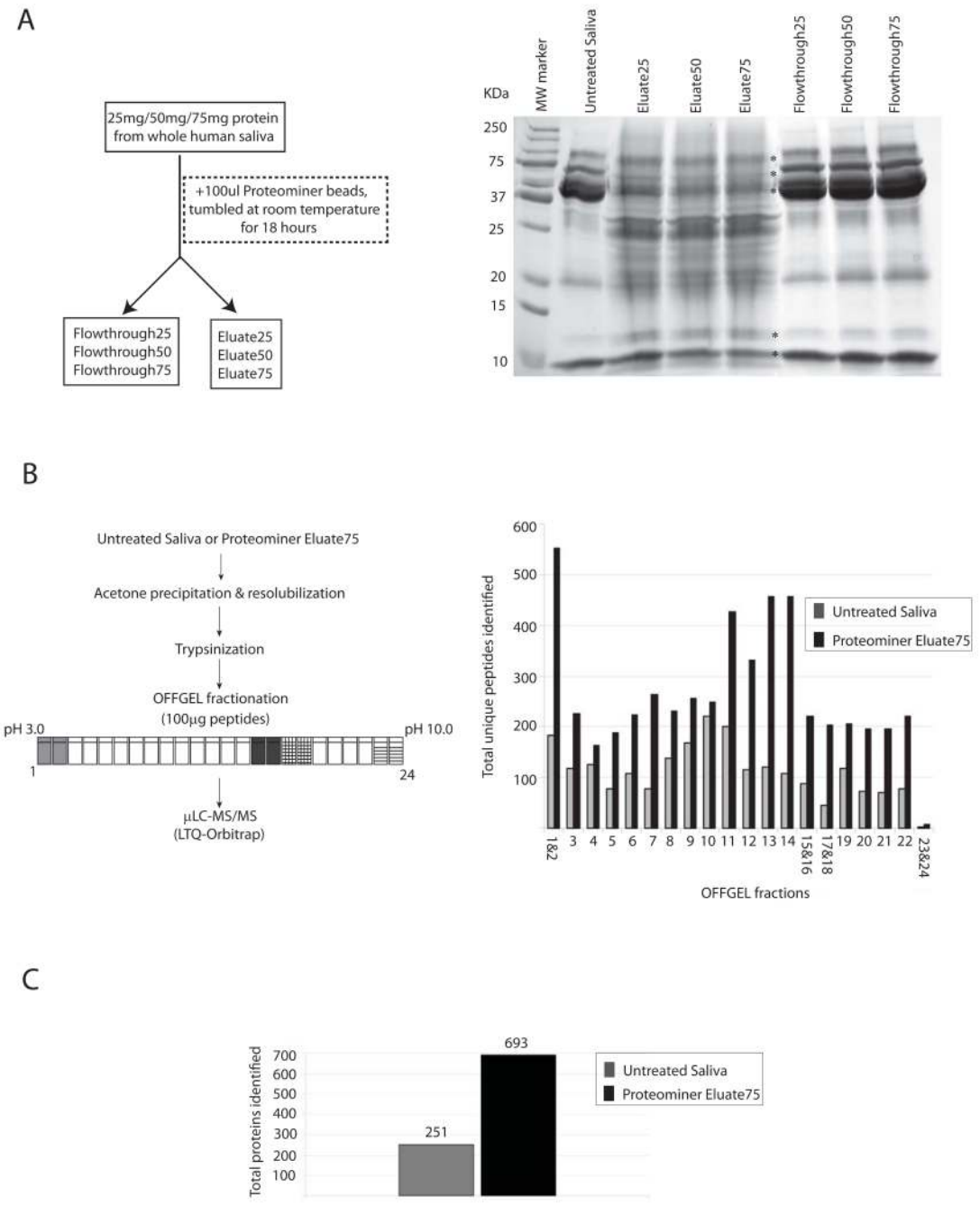

Figure 1. Optimization of proteominer treatment with whole saliva and results from 2D-peptide fractionation with/without dynamic range compression of saliva

(a) Increasing amounts of clarified whole saliva protein (25, 50 and 75 mg) were incubated with 100 µL of Proteominer beads overnight in the presence of protease inhibitors. Bound protein (eluate) and Flow through fractions from each treatment were collected. Untreated Saliva, flow through, and eluate fractions from samples were separated by SDS-PAGE and visualized by coomassie staining. **(b)** 100 µg peptides from Untreated Saliva, or Proteominer Eluate75 were trypsinized and fractionated by preparative IEF (OFFGEL) based on their isoelectric points (pH 3 to 10), and processed for µLC-MS/MS analysis. Total unique peptides

identified in each OFFGEL fraction(s) are indicated for Untreated Saliva and Proteominer Eluate75. **(c)** Total proteins identified in Untreated Saliva versus Proteominer Eluate75.

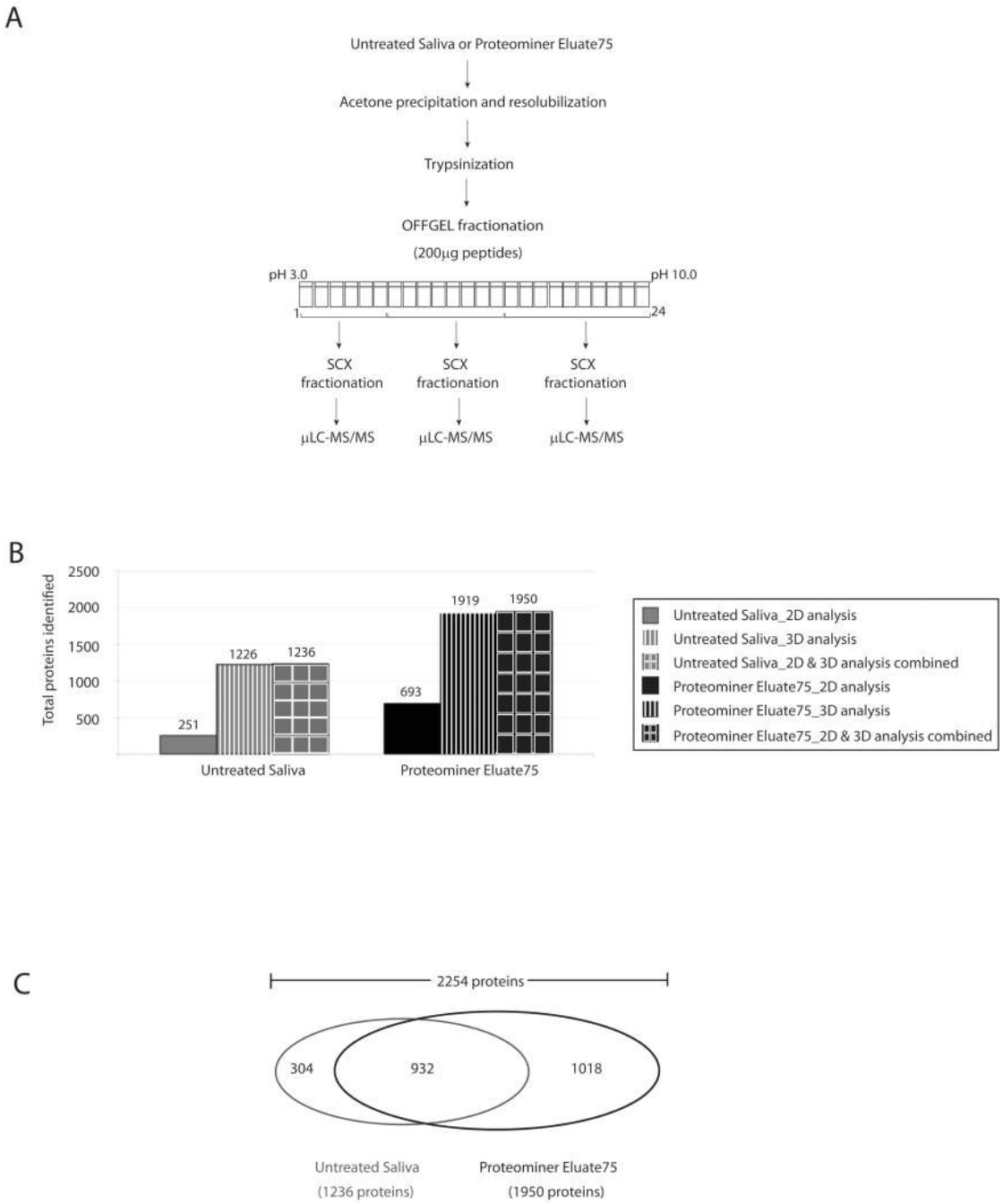

Figure 2. 3D-peptide fractionation increased total proteins identified in Untreated and Proteominer-treated whole saliva

(a) Scheme for 3D-peptide fractionation. **(b)** Total proteins identified in Untreated Saliva and Proteominer Eluate75 by 2D- versus 3D-peptide fractionation. Results from combining both analyses are also shown. **(c)** Venn diagram illustrating total number of proteins those are specific to either Proteominer treatment or Untreated Saliva and those identified in both samples.

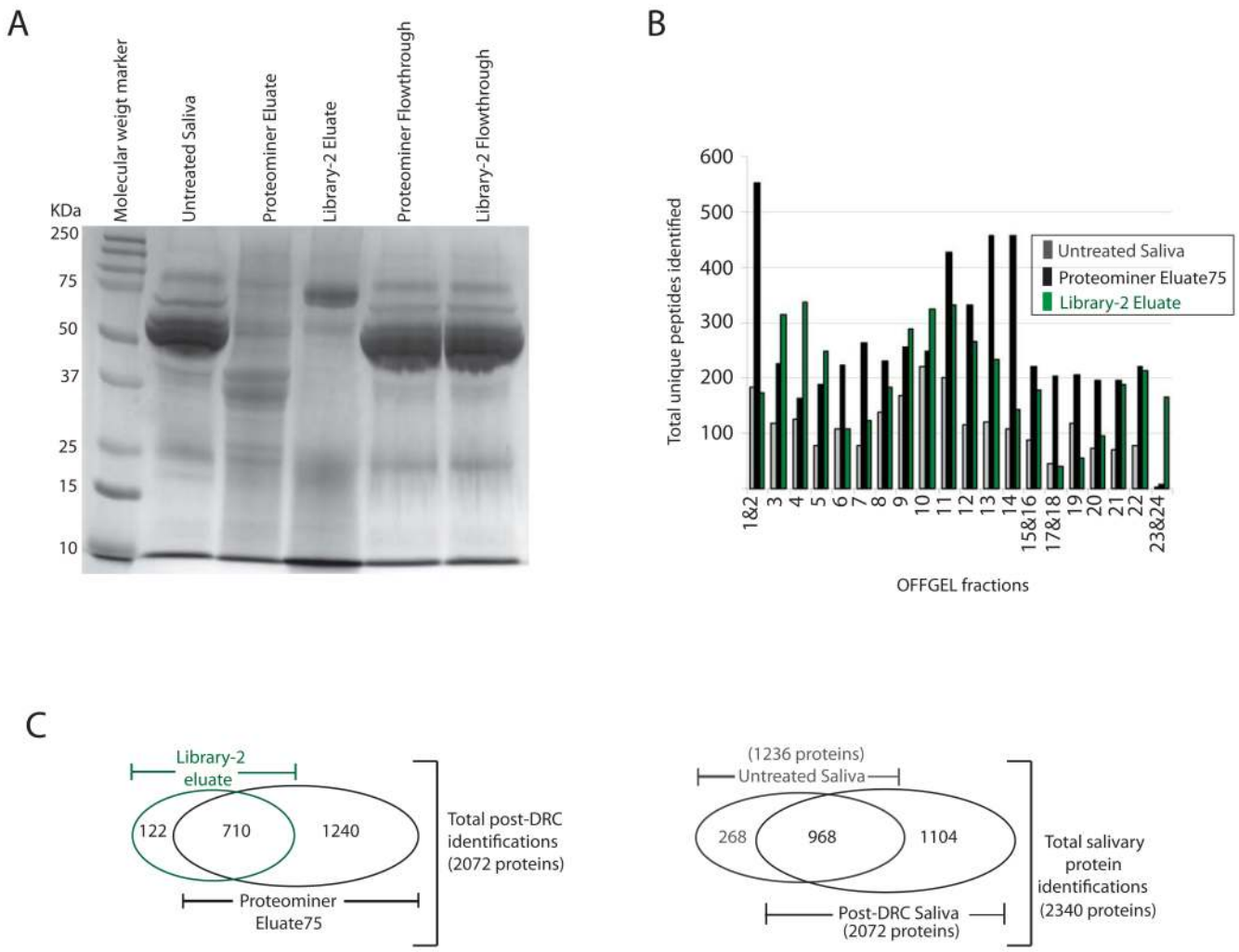

Figure 3. Additional identifications obtained after Library-2 treatment of whole saliva further expanded our total salivary dataset

(a) Library-2 eluates and flow through fractions were compared with those obtained after Proteominer treatment under identical conditions by SDS-PAGE analysis and coomassie staining. **(b)** Total unique peptides identified in each OFFGEL fraction/s after Library-2 treatment of saliva versus Proteominer treatment or Untreated Saliva. **(c) Left panel**, Venn diagram illustrating total proteins obtained from DRC of saliva and the individual contributions of Library-2 versus Proteominer hexapeptide libraries. **Right panel**, Venn diagram illustrating total proteins identified in this study and the relative contribution of DRC to these identifications.

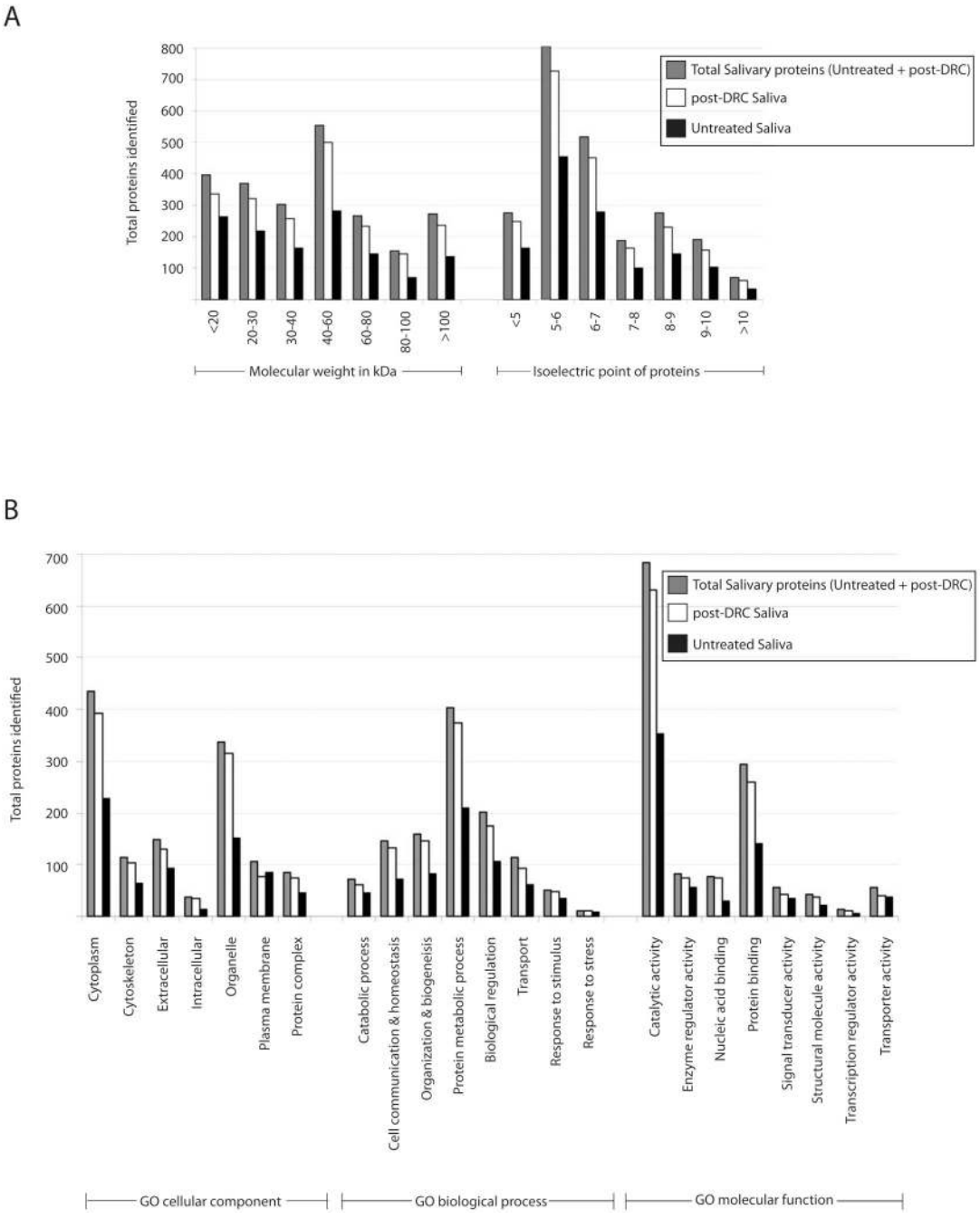

Figure 4. DRC increased proteins across physicochemical and Gene Ontological categories without apparent selectivity

Proteins identified in Untreated Saliva, post-DRC saliva and total salivary identifications were grouped into individual categories as described in Results and Discussion section of the manuscript.

Figure 5. Comparison of total salivary protein identifications against those found in plasma for determining functional diversity and diagnostic potential of both fluids

Ingenuity pathway analysis (IPA) was done as described in Experimental Methods.