Exploring the Effects of Pepsin on Salivary Peptidome: Insights from a Proof-of-Concept Proteomic Profiling Study Using MALDI-ToF Mass Spectrometry

Diana-Elena Floria , Andrei Neamțu , Radu Iliescu , Brîndușa Alina Petre * , Batuhan Uzunoglu , Oana-Bogdana Bărboi , Vasile-Liviu Drug

Posted Date: 9 April 2026

doi: 10.20944/preprints202604.0608.v1

Keywords: salivary peptidome; MALDI ToF mass spectrometry; pepsin digestion; gastro-esophageal reflux disease; peptide biomarkers

 doi:10.20944/preprints202604.0608.v1

Article

Exploring the Effects of Pepsin on Salivary Peptidome: Insights from a Proof-of-Concept Proteomic Profiling Study Using MALDI-ToF Mass Spectrometry

Diana-Elena Floria [1,2,3], Andrei Neamțu [1,4], Radu Iliescu [1,4], Brîndușa Alina Petre [4,5,*], Batuhan Uzunoglu [6,7], Oana-Bogdana Bărboi [1,2] and Vasile-Liviu Drug [1,2]

[1] Grigore T. Popa University of Medicine and Pharmacy, Iași, Romania
[2] Institute of Gastroenterology and Hepatology, Saint Spiridon Hospital, Iași, Romania
[3] Centre for Translational Medicine, Semmelweis University, Budapest, Hungary
[4] TRANSCEND Centre, Regional Institute of Oncology, Iași, Romania
[5] Faculty of Chemistry, Alexandru Ioan Cuza University, Iași, Romania
[6] Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY, USA
[7] BIOsem Laboratories, Department of Chemistry and Biochemistry, Clarkson University, Potsdam, NY, USA
* Correspondence: brindusa.petre@uaic.ro

Abstract

Background: Pepsin, a component of gastric refluxate, has been investigated as a potential salivary biomarker for gastro-esophageal reflux disease (GERD), although its diagnostic accuracy remains uncertain. This proof-of-concept study aimed to characterize the effects of pepsin on the human salivary peptidome using matrix-assisted laser desorption/ionization–time of flight (MALDI-ToF) mass spectrometry. **Methods:** Whole saliva samples were collected from ten healthy adult volunteers under fasting conditions and divided into untreated controls and aliquots digested with pepsin at acidic pH. MALDI-ToF MS was used to profile digestion-induced changes in peptide mass patterns. Spectral data were analyzed using multivariate statistical approaches, including principal component analysis (PCA), linear discriminant analysis (LDA), and hierarchical clustering. **Results:** Pepsin digestion increased peptide signal intensity and spectral complexity compared with controls. PCA demonstrated clear separation between native and digested samples along the first principal component. Ten peptide m/z features showed the strongest association with pepsin exposure based on PCA loadings. LDA and hierarchical clustering further supported this distinction, with the top 15 discriminative m/z features showing consistent enrichment in digested samples despite inter-individual variability. **Conclusions:** Pepsin exposure induces reproducible remodeling of the salivary peptidome detectable by MALDI-ToF MS. Although this peptide-level approach cannot resolve the full diversity of salivary proteoforms, the resulting signatures support the feasibility of identifying markers of reflux-associated enzymatic activity and provide a basis for future validation in clinical GERD cohorts.

Keywords: salivary peptidome; MALDI ToF mass spectrometry; pepsin digestion; gastro-esophageal reflux disease; peptide biomarkers

1. Introduction

Gastroesophageal reflux disease (GERD) is one of the most commonly encountered gastrointestinal disorders in clinical practice, with a notable global prevalence estimated to exceed 20% in Western countries [1]. Studies have shown that individuals with reflux disease experience a significant reduction in health-related quality of life compared to the general population, irrespective

 doi:10.20944/preprints202604.0608.v1

of the severity of endoscopic findings [2]. It poses a significant burden on healthcare systems worldwide, requiring substantial resources in order to manage and treat effectively [3]. Currently, ascertaining a GERD diagnosis often relies on performing a series of invasive and expensive procedures, such as upper gastrointestinal endoscopy or ambulatory pH-impedance monitoring [4]. However, these methods are hindered by relatively high costs, limited accessibility in some settings, and the requirement for specialized expertise to accurately interpret the findings [5]. In this context, there is a growing need for reliable non-invasive diagnostic biomarkers that could aid in the diagnosis of this condition.

Saliva, a readily accessible biofluid, has been proposed as a potential source of biomarkers for GERD. To that end, many studies investigated the use of salivary pepsin as a diagnostic test, a digestive enzyme secreted by gastric chief cells and key factor in tissue injury [6]. However, currently there is a lack of consensus regarding optimal collection methods, processing techniques, and timing for measuring salivary pepsin, with varying reports regarding its diagnostic accuracy [7]. While the presence of salivary pepsin has been investigated as a diagnostic biomarker in itself, less focus has been placed on its biological effects on the salivary proteome [8]. Given pepsin's potent proteolytic activity under acidic conditions, reflux events may be associated with a characteristic salivary peptide pattern that persists beyond the transient presence of the enzyme itself.

The rapid advances in proteomic technologies provide an opportunity to examine these effects in more detail, broadening the avenues of non-invasive diagnostics [8,9]. Matrix-Assisted Laser Desorption/Ionization-Time of Flight (MALDI-ToF) mass spectrometry is a robust analytical technique widely used for rapid and precise profiling of peptide and protein mass patterns. Owing to its sensitivity and reproducibility, it represents an appropriate approach for characterizing the salivary peptidome and for exploring potential biomarkers relevant to GERD.

In this proof-of-concept study, we investigated the effects of pepsin exposure under acidic conditions on the human salivary peptidome using MALDI-ToF mass spectrometry. Rather than focusing on the detection of salivary pepsin itself, our approach aimed to characterize the proteolytic peptide signatures generated by pepsin activity. Identifying reproducible peptide patterns may provide insights into reflux-related enzymatic effects and advance the discovery of potential non-invasive diagnostic tools for GERD. To complement the experimental digestion, an *in-silico* simulation of pepsin-mediated cleavage was carried out using MS Digest, a freely available online tool of ProteinProspector program[10], providing a theoretical peptide map to support interpretation of the MALDI-ToF results.

2. Materials and Methods

The study protocol received ethical approval from the Grigore T. Popa University of Medicine and Pharmacy Iași Institutional Review Board (232/13.10.2022). An overview of the study design is presented in Figure 1. Saliva samples were collected and divided into two preparation groups. In the control group, samples were acidified only. In the experimental group, acidified saliva was incubated with pepsin to generate digested peptide products. In silico pepsin digestion was performed to support peptide identification. All samples were analyzed using MALDI-ToF mass spectrometry. Spectral data were interpreted and subjected to comparative analysis and pattern identification to evaluate differences between control and pepsin-digested samples. (Figure created with BioRender.com).

 doi:10.20944/preprints202604.0608.v1

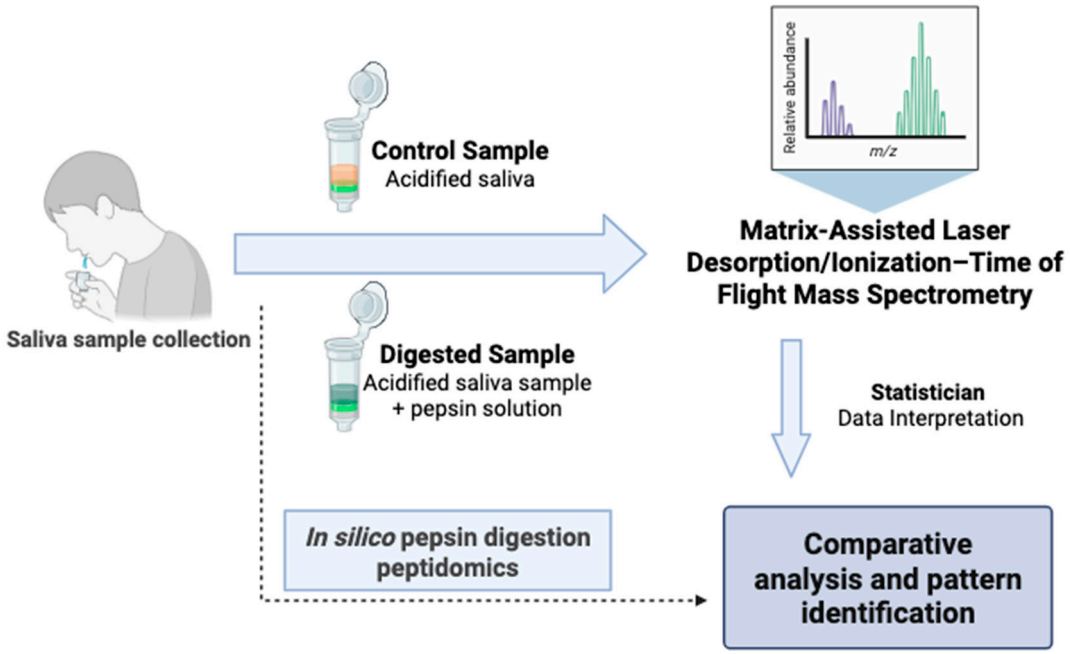

Figure 1. Detailed workflow for salivary peptidomics analysis using MALDI-ToF mass spectrometry in silico digestion and statistical interpretation of mass spectra.

2.1. Saliva Sample Collection

Whole saliva samples were collected from healthy volunteers (n=10) aged 20 to 30 years. Sample collection was performed in the morning following an overnight fast, in the first two hours after waking. To minimize potential contamination, participants were instructed to brush their teeth 30 minutes prior to collection, without consuming any food, beverages, or smoking afterwards. Approximately 2-3 mL of unstimulated saliva was collected in sterile tubes.

2.2. Control Samples

Control samples consisted of acidified whole saliva without the addition of pepsin, processed in parallel with the experimental samples under identical conditions, including incubation, centrifugation, desalination, and mass spectrometry analysis. This allowed for comparison to identify pepsin-specific proteolytic effects. These samples were labeled C1 through C10 to ensure consistent tracking throughout the workflow. For the subsequent digestion experiments, the saliva aliquots exposed to pepsin were labeled D1 through D10, following the same numbering scheme as their corresponding control samples.

2.3. Pepsin Treatment and Acidification

Pepsin solution (Sigma-Aldrich) was combined with saliva samples in a 1:20 ratio (pepsin to saliva). The total protein concentration in unstimulated whole saliva ranged from 2.9 to 3.1 μg/μL, consistent with values previously reported in the literature [11]. Protein concentration was determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific) based on UV–Vis absorbance measurements. The instrument quantified protein levels by measuring absorbance at 280 nm, where aromatic amino acids exhibit characteristic absorption. A sample volume of 1–2 μL was used, and the NanoDrop's short-path optical system allowed direct measurement without cuvettes. Background correction was automatically applied using a reference wavelength (typically 260 nm) to ensure accurate readings for complex biological matrices such as saliva. Acidification of the samples was performed to achieve a pH range of 1-2, ensuring optimal proteolytic activity of pepsin. pH was measured using pH indicator paper, and adjustments were made using hydrochloric acid (0.1 N).

 doi:10.20944/preprints202604.0608.v1

2.4. Digestion Process and Supernatant Collection

The saliva-pepsin mixtures were incubated at 37°C for 30 minutes at 450 rotations per minute (rpm) using a shaking incubator. This process was used to maximize pepsin enzymatic activity. All saliva digestion reactions were performed in technical duplicate to ensure reproducibility of the pepsin-mediated proteolysis workflow. Following digestion, samples were centrifuged at 12,000 rpm for 10 minutes and supernatant was collected.

2.5. Sample Desalination, Concentration, and Purification

Supernatant samples were desalted, concentrated, and purified using ZipTip C18 pipette tips (Millipore), following the manufacturer's protocol[12]. First, the C18 tips were activated by wetting with acetonitrile (100 % ACN). Next, the tips were equilibrated with an aqueous solution containing trifluoroacetic acid (0.1% TFA). The samples were then repeatedly aspirated through the tips, allowing peptides and small proteins to bind to the C18 resin while salts and other impurities were left unbound. After binding, the tips were washed with an acidic solution (0.1% TFA) to remove residual salts and contaminants. Finally, the purified analytes were eluted (50 % ACN in 0.1% TFA), producing a desalted, concentrated solution of analytes ready for analysis by mass spectrometry. The same procedure of purification was applied to the control saliva samples.

2.6. MALDI-ToF Mass Spectrometry

MALDI-ToF MS was performed on a Bruker Ultraflex MALDI ToF/ToF mass spectrometer equipped with a pulsed nitrogen UV laser. The instrument was operated in positive ion mode, and spectra were acquired in reflectron mode. Prior to spotting, peptide samples were mixed with super-DHB (sDHB), a matrix consisting of a 1:1 mixture of 2,5-dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid, commonly used to enhance peptide ionization and improve signal homogeneity in MALDI analyses. Samples were applied onto a 384-spot stainless-steel target plate using the dried-droplet method by depositing 0.8 µL of matrix solution and 1.0 µL of sample solution, mixed directly on the target and allowed to dry at room temperature. The laser attenuation was set to 80%, and spectra were acquired using the following parameters: 20 kV acceleration voltage, 40% grid voltage, 140 ns delay, low-mass gate of 500 Da, and an acquisition mass range of 450–5000 Da. For each final spectrum, 500 laser shots were accumulated. External calibration was performed using the monoisotopic [M+H]+ ions of the following peptide standards: Bradykinin (1–7) (m/z 757.4), Angiotensin II (m/z 1046.5), Angiotensin I (m/z 1296.7), Substance P (m/z 1347.7), Bombesin (m/z 1619.8), Renin Substrate (m/z 1758.9), ACTH (1–17) (m/z 2093.1), ACTH (18–39) (m/z 2465.2), and Somatostatin (m/z 3147.5). Theoretical monoisotopic masses were calculated using the ChemCalc online tool. Spectra were processed using Bruker FlexAnalysis 3.4 software. The MALDI ToF mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository [13] with the dataset identifier PXD076267.

2.7. Statistical Analysis and Data Processing

Data processing and statistical analyses were performed using Python (version 3.9.x). Raw MALDI-ToF mass spectra were exported as CSV files and processed using a custom computational pipeline. Unsupervised Principal Component Analysis (PCA) was conducted to visualize the global variance and identify inherent clusters between the native and digested peptide profiles. To specifically identify the peptide features that best discriminate between the two conditions, a supervised Linear Discriminant Analysis (LDA) was employed. The predictive performance of the LDA model was validated using Leave-One-Out (LOO) cross-validation. The following Python libraries were utilized: (1) Pandas (v1.3.5) and NumPy (v1.21.6) for data manipulation and numerical operations; (2) Scikit-learn (v1.0.2) for PCA, LDA, and cross-validation procedures; (3) SciPy (v1.7.3) for t-test calculations; (4) Matplotlib (v3.5.2) and Seaborn (v0.11.2) for the generation of high-resolution mirror plots, PCA score plots, and heatmaps.

 doi:10.20944/preprints202604.0608.v1

3. Results

To evaluate the proteolytic effects of pepsin on the human salivary proteome, whole saliva samples from ten healthy individuals were subjected to pepsin digestion under acidic conditions and analyzed using MALDI-ToF mass spectrometry. For each participant, one control sample and one digested aliquot were compared. Analysis of the acidified saliva samples provided a stable proteomic baseline. Following exposure to pepsin, distinct pepsin-induced proteomic changes were observed in the salivary samples. To assess the reproducibility of the digestion workflow, all pepsin-treated saliva samples were processed in technical duplicate. The duplicate digestions consistently produced comparable peptide profiles and signal intensities, confirming the robustness of the proteolytic procedure. Representative mass spectra from three duplicate digestion experiments are provided in the Supplementary Material Figure S1 to illustrate the reproducibility of the analytical workflow.

From the ten saliva samples analyzed, we identified three distinct salivary peptide profiles. Each profile combines a set of shared peptide signals with additional, profile-specific features that differ in intensity, even though the MALDI-ToF analytical conditions were kept constant. These reproducible patterns, observed across individuals, indicate inherent variability in the salivary peptide composition (see Supplementary Figure S2). Such variability is expected, as the salivary peptidome is influenced by physiological factors including individual protease activity, oral microbiota, hydration status, and subtle differences in sample handling. Consequently, the presence of distinct peptide profiles among healthy participants is consistent with previously reported inter-individual diversity in the human salivary proteome[14]. Within these three categories, sample C1 displayed a peptide profile highly similar to that of sample C6, while the profile of sample C2 closely resembled that of sample C9. In contrast, samples C3–C5, C7, C8, and C10 formed a larger group characterized by markedly similar peptide patterns. This clustering further supports the presence of inherent inter-individual variability in salivary peptide composition under identical MALDI-ToF analytical conditions.

To exemplify the three peptide profile types identified among the ten saliva samples, we present representative MALDI-ToF spectra corresponding to each category. Figure 2 illustrates the first profile type, using sample C1 as a representative example. Panel A shows the peptide mass fingerprint of the untreated saliva sample (C1), characterized by a series of well-defined signals spanning the 450–5000 m/z range, reflecting the endogenous salivary peptide distribution under acidic conditions. Panel B displays the corresponding spectrum of the same sample following pepsin digestion (D1), where shifts in peak intensities and the appearance of additional lower-molecular-weight ions indicate pepsin-mediated proteolysis. Figure 3 illustrates the second peptide profile type identified among the ten saliva samples, using sample C2 as a representative example. Panel A shows the MALDI-ToF spectrum of the untreated saliva sample C2, characterized by a distinct distribution of endogenous peptide signals that differentiates this profile from the one shown in Figure 2. Panel B presents the corresponding spectrum of the same sample following pepsin digestion (D2). As observed for the first profile type, pepsin exposure results in the appearance of additional low-molecular-weight ions and modifications in peak intensities; however, the fragmentation pattern and relative signal distribution are characteristic of this second profile category.

 doi:10.20944/preprints202604.0608.v1

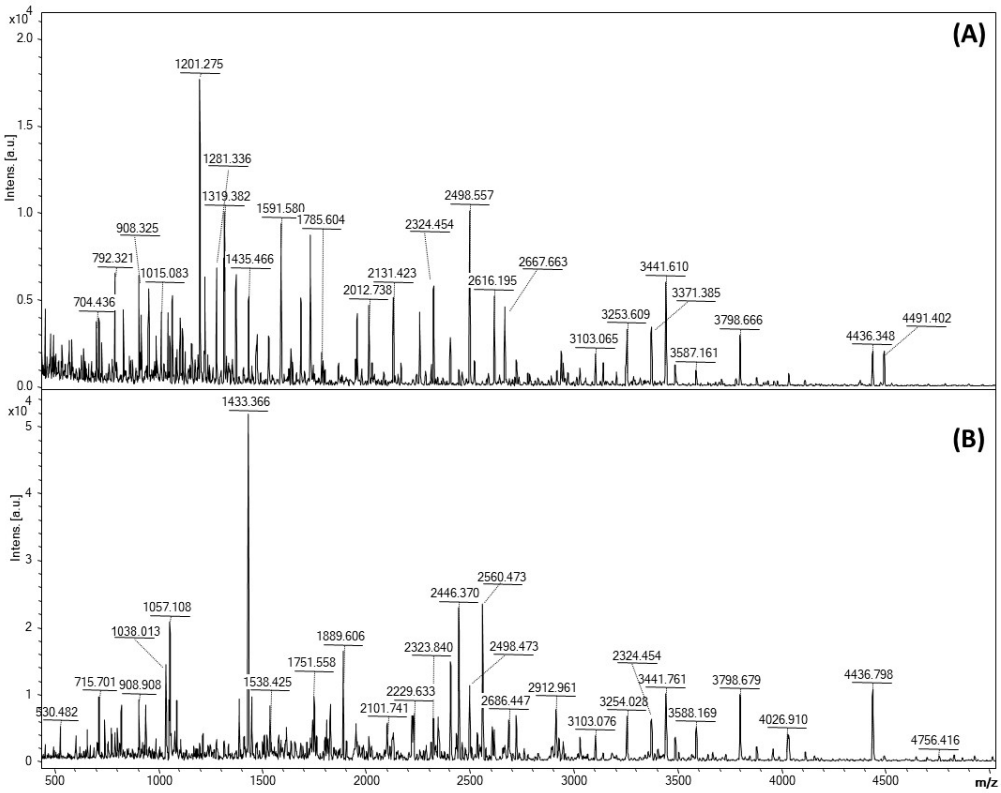

Figure 2. Mass spectrometric peptide mass fingerprints illustrating one of the three salivary peptide profile types identified among the ten saliva samples. (A) MALDI-ToF spectrum of the untreated saliva sample C1 and (B) Corresponding MALDI-ToF spectrum of the same sample following pepsin digestion (D1).

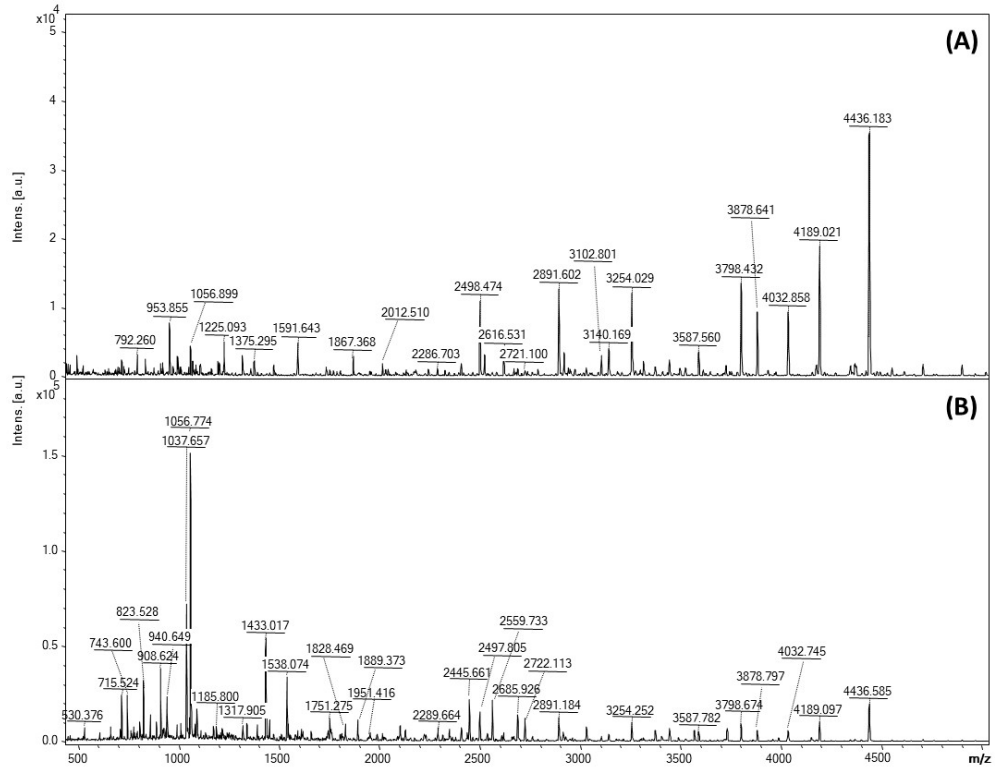

Figure 3. Mass spectrometric peptide mass fingerprints illustrating a second salivary peptide profile type identified among the ten saliva samples. (A) MALDI-ToF spectrum of the untreated saliva sample C2, and (B) Corresponding MALDI-ToF spectrum of the same sample after pepsin digestion (D2).

 doi:10.20944/preprints202604.0608.v1

Figure 4 illustrates the third peptide profile type identified among the ten saliva samples, using sample C7 as a representative example. Panel A shows the MALDI-ToF spectrum of the untreated saliva sample C7, characterized by a peptide distribution that is distinct from the patterns observed in Figure 2 and Figure 3. This profile type is defined by a specific arrangement of endogenous peptide signals and relative peak intensities that consistently grouped samples C3–C5, C7, C8, and C10 within the same category. Panel B presents the corresponding spectrum of the same sample following pepsin digestion (D7). As with the other profile types, pepsin exposure results in the appearance of additional low-molecular-weight ions and notable shifts in peak intensities. However, the fragmentation pattern observed for C7/D7 is characteristic of this third profile category and differs from the proteolytic signatures shown in Figure 2 and Figure 3.

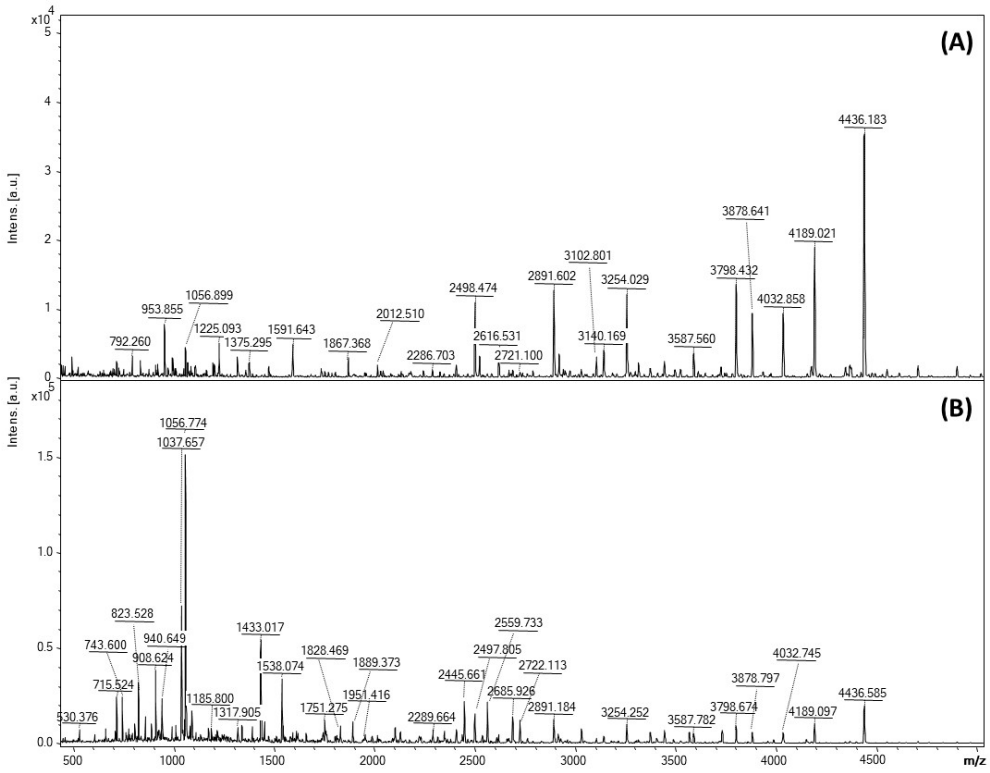

Figure 4. Representative MALDI-ToF peptide mass fingerprints illustrating the third salivary peptide profile type identified among the ten saliva samples. (A) MALDI-ToF spectrum of the untreated saliva sample C7, and (B) Corresponding MALDI-ToF spectrum of the same sample following pepsin digestion (D7).

For the *in silico* component of the study, a dedicated salivary protein database was first constructed from the UniProt Knowledgebase (UniProtKB). A keyword search using the term "salivary" was performed and restricted to the organism Homo sapiens (Human). This query yielded 638 protein entries, which were exported and compiled into a custom FASTA database for subsequent computational analysis.

In silico proteolysis was carried out using the MS-Digest online tool available through the ProteinProspector suite (https://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest). Pepsin (*porcine gastric*) was selected as the proteolytic enzyme to replicate the experimental digestion conditions. Default digestion parameters were applied, including fully specific enzymatic cleavage according to pepsin specificity rules and allowance for up to two missed cleavages. No fixed or variable post-translational modifications were introduced, so the generated peptide list reflected cleavage based solely on primary amino acid sequence.

Application of these parameters to the 638 salivary protein entries resulted in 25009 theoretical peptide fragments. Because MALDI-MS analysis primarily provides peptide mass information without complete sequence coverage for all detected ions, the *in silico* digestion served to generate a

 doi:10.20944/preprints202604.0608.v1

list of putative peptide sequences corresponding to experimentally observed m/z values (assuming a +1 ion charge). Therefore, fragment assignment was based on mass matching and represents predicted candidate sequences rather than unequivocal sequence confirmation in all cases (see Table S1 in supplementary material). This approach enabled a systematic comparison between experimentally detected pepsin-derived fragments and theoretically expected cleavage products within the human salivary proteome.

To further explore the variability observed across the salivary peptide profiles and to quantify the proteolytic effects of pepsin, all MALDI-ToF spectra from the control and digested samples were subjected to comprehensive statistical analysis. After peak alignment and intensity normalization, the dataset was evaluated using multiple multivariate approaches, including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and hierarchical clustering with heatmap visualization. These complementary methods enabled the identification of sample groupings, the discrimination between control and pepsin-digested spectra, and the detection of m/z features contributing most strongly to sample separation.

The PCA plot illustrates the impact of pepsin digestion on the salivary peptide composition Figure 5 [15]. The first principal component (PC1) captures the dominant source of variation in the dataset and clearly separates the digested samples (orange circles) from the native saliva samples (blue circles). This distinct clustering demonstrates that pepsin digestion consistently produces a characteristic peptide pattern that is markedly different from the natural inter-individual variability observed in untreated saliva. The separation along PC1 is driven by a specific subset of peptide ions whose intensities increase substantially following pepsin exposure. These ions correspond to the loadings of the first principal component and represent the features most strongly contributing to the discrimination between the two sample groups.

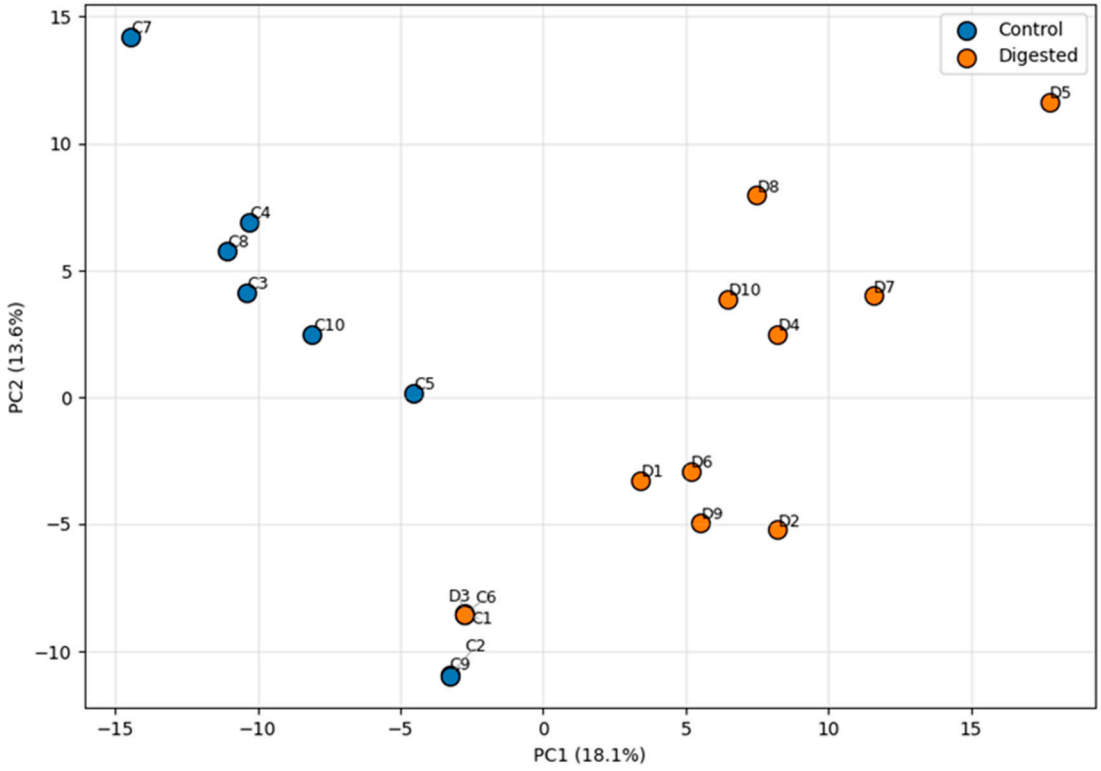

Figure 5. PCA score plot illustrating the separation between control and pepsin-digested saliva samples (variance explained: 31.7%).

In addition, hierarchical clustering of the most discriminative m/z features revealed a distinct grouping of ions that were consistently enriched in the pepsin-digested samples (Table 1). Based on the PCA loadings, the m/z features contributing most strongly to PC1 were ranked in descending

 doi:10.20944/preprints202604.0608.v1

order according to their loading values. The ions with the highest positive influence on PC1 were m/z 530.34, 743.49, 940.50, 1088.82, and 1750.99, followed by m/z 1432.76, 908.46, 2289.27, 1537.94, and 2406.36. Importantly, all top contributors exhibited positive loading values, indicating that their intensities increase systematically in the digested samples and drive the rightward shift of these samples along the PC1 axis. Notably, the peptide masses listed in Table 1 are also found among the theoretical peptides generated in the in silico digestion, further supporting their biological relevance. These reproducible features may represent candidate biomarkers of pepsin-mediated proteolysis, pending confirmation through targeted MS/MS identification and validation in independent sample sets.

Table 1. Top m/z ions ranked by their PC1 loading values, highlighting the peptide features most strongly associated with pepsin-induced spectral changes.

m/z (Ref)	PC1 Loading Value
530.34	+0.1076
743.49	+0.1075
940.5	+0.1074
1088.82	+0.1072
1750.99	+0.1070
1432.76	+0.1062
908.46	+0.1043
2289.27	+0.1036
1537.94	+0.0998
2406.36	+0.0992

Building on the PCA findings and the discriminative m/z features summarized in Table 1, the Linear Discriminant Analysis (LDA) provides a supervised dimensionality-reduction approach designed to identify the peptide signals that best separate the two predefined groups[16]. Unlike PCA, which captures the directions of maximum variance without using class information, LDA explicitly incorporates the sample labels and learns the optimal combination of m/z features that distinguishes the Control from the Digested spectra [17]. The resulting discriminant axis (LD1) positions most native saliva samples on the left and the pepsin-digested samples on the right, reflecting the consistent biochemical shift induced by proteolysis. The central red line represents the decision boundary used for classification. Two samples (C1 and D3) appear near this threshold, indicating intermediate peptide profiles that partially overlap with the opposite group, a finding consistent with the natural inter-individual variability of the salivary peptidome and with the PCA-derived loadings. Model performance was evaluated through Leave-One-Out cross-validation (LOO)[18], yielding an accuracy of 60%, further supporting the conclusion that pepsin digestion introduces systematic and detectable changes in the salivary peptide patterns.

To further extend the multivariate analysis and to visualize sample-level differences across the most relevant peptide markers, a heatmap with hierarchical clustering was generated using the top 15 m/z features that most strongly distinguished the Control and Digested groups Figure 7. These peptides ions correspond to the statistically significant markers identified in the PCA loadings and summarized in Table 1, and their intensity patterns provide a high-density overview of the proteolytic changes induced by pepsin digestion.

 doi:10.20944/preprints202604.0608.v1

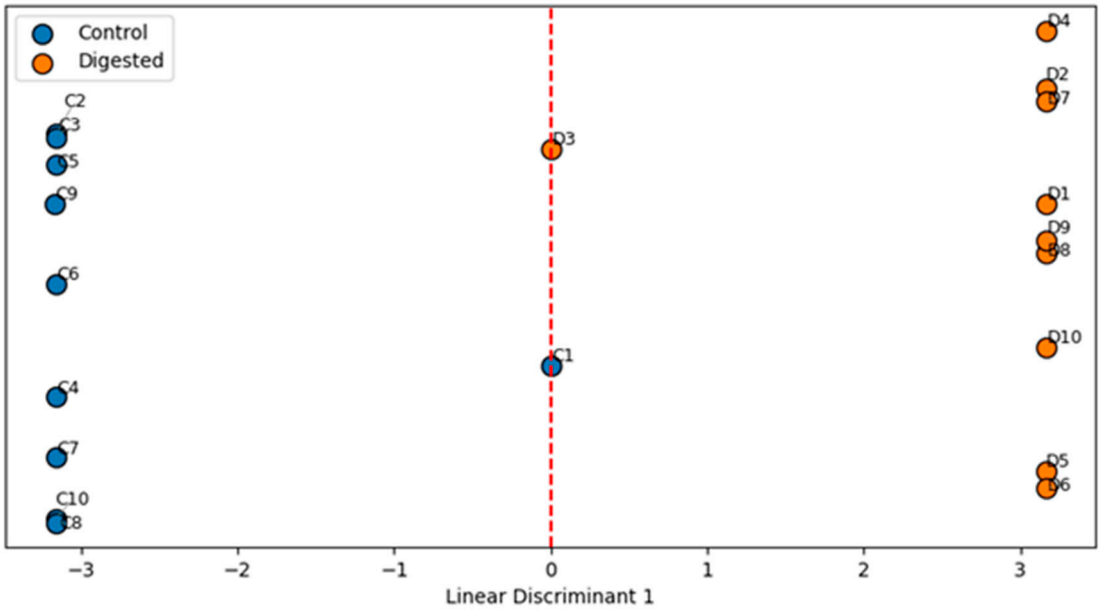

Figure 6. Linear Discriminant Analysis (LDA) of control and pepsin-digested saliva samples, showing group separation along the LD1 axis and the decision boundary used for classification (LOO accuracy: 60%).

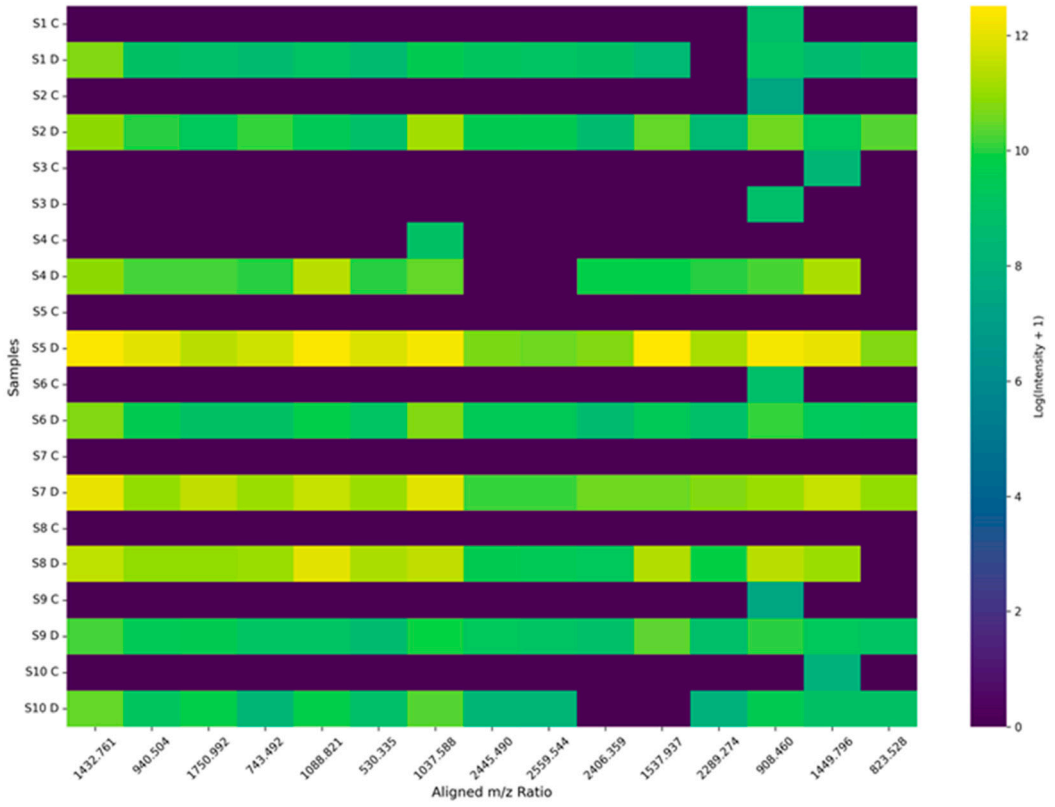

Figure 7. Heatmap with hierarchical clustering of the top 15 discriminative m/z features, illustrating distinct intensity patterns between control and pepsin-digested saliva samples and highlighting sample-level variability across the dataset.

The heatmap reveals two clearly separated clusters, with control samples displaying consistently lower intensities for these m/z values, while digested samples exhibit marked increases, in agreement with the PCA and LDA results. Notably, Sample 3 shows an atypical pattern: although internally variable, it lacks several of the discriminative m/z features present in the other samples. This explains

 doi:10.20944/preprints202604.0608.v1

its divergent position in the clustering and reflects natural inter-individual variability rather than a computational artifact. Overall, the heatmap reinforces the biological relevance of the peptide markers enriched in the digested samples and supports their potential as candidate indicators of pepsin-mediated proteolysis, pending MS/MS validation.

4. Discussion

GERD is one of the most prevalent gastroenterological disorders encountered in clinical practice, and there is a pressing need for non-invasive diagnostic tools to improve diagnostic accuracy. Saliva has emerged as an attractive and easily accessible diagnostic medium, capable of reflecting both systemic and local pathological processes [19,20]. Salivary pepsin has been investigated as a potential biomarker for GERD; however, reported sensitivity and specificity vary widely across studies, likely due to substantial methodological heterogeneity [21,22]. These limitations suggest that measuring pepsin concentration alone may not fully capture the complexity of reflux-related pathophysiology.

Pepsin is not merely a passive marker of reflux but an active proteolytic enzyme that remains biologically functional under acidic conditions and has been implicated in mucosal injury and inflammation in reflux disease [23,24]. Salivary proteins are more susceptible to proteolytic degradation than serum proteins [25,26], making the salivary peptidome particularly responsive to pepsin-mediated cleavage. Disease- or exposure-related alterations can reshape the proteome by generating distinct peptide patterns that reflect functional proteome states, rendering peptide-level analysis especially informative in dynamic biofluids such as saliva [27].

Recent advancements in proteomic technologies have expanded the potential of saliva as a diagnostic medium for a wide range of conditions, including head-and-neck and oral cancer, diabetes, gastric cancer, Sjögren's syndrome, primary sclerosing cholangitis, and COVID-19 infection [8,28–30]. As a proof-of-concept, this study employed MALDI-ToF mass spectrometry to characterize pepsin-induced alterations in the salivary peptidome of healthy volunteers, with the aim of identifying digestion-associated peptide patterns that may serve as indicators of reflux-related enzymatic activity. MALDI-ToF enables rapid and accurate analysis of complex biological samples by determining the mass-to-charge (m/z) ratios of ionized peptides based on their time-of-flight, making it a promising tool for medical diagnostics and biomarker discovery [31,32].

Previous studies have shown that individuals with GERD exhibit lower oral pH values, which may influence the oral environment [33–35]. Proteomic analyses of the acquired enamel pellicle in GERD patients have revealed profound alterations compared to controls, both in individuals with and without erosive tooth wear [36,37]. Moreover, clinical studies of saliva in GERD have demonstrated reflux-associated remodeling of the salivary proteome, particularly affecting proteins involved in acid resistance and oral tissue protection [38]. These findings collectively support the rationale for exploring pepsin-induced peptide signatures as potential biomarkers of reflux activity.

Our results demonstrate that pepsin induces marked and reproducible alterations in the salivary peptidome, reflected by both quantitative and qualitative changes across multiple analytical levels. MALDI-ToF profiling revealed a significant increase in the number of detectable peptide signals following enzymatic digestion, together with the appearance of novel m/z peaks absent in untreated saliva. These findings were further supported by multivariate statistical analyses. PCA showed a clear separation between native and digested samples along PC1, driven by a subset of peptide ions with strongly positive loading values, indicating systematic increases in their intensities after pepsin exposure. The most influential of these features are summarized in Table 1 and overlap with peptides predicted through *in silico* digestion, reinforcing their biological plausibility.

LDA provided complementary evidence by identifying the optimal combination of discriminative m/z features that maximized group separation. Despite natural inter-individual variability, most samples were correctly classified, and the model achieved a 60% LOO accuracy. Samples positioned near the decision boundary (e.g., C1 and D3) displayed intermediate peptide profiles, consistent with the variability observed in the PCA and heatmap analyses. The hierarchical clustering heatmap of the top 15 discriminative m/z features further confirmed the presence of two

 doi:10.20944/preprints202604.0608.v1

distinct sample clusters, with digestion-specific peptides showing consistently elevated intensities in the digested group. One sample (Sample 3) exhibited an atypical pattern due to the absence of several discriminative m/z values, reflecting internal variability rather than analytical error.

Together, these converging lines of evidence highlight robust pepsin-induced remodeling of the salivary peptidome. The reproducible detection of digestion-specific peptides across multiple individuals, combined with their statistical relevance and alignment with in silico predictions, supports their potential as candidate biomarkers of reflux-associated pepsin activity. As a proof-of-concept, this study demonstrates the feasibility of using salivary peptidomics to detect enzymatic signatures associated with gastric reflux.

To advance this line of research, we will perform high-resolution ESI-MS/MS analysis of the salivary samples to determine the precise amino-acid sequences and protein origins of the candidate biomarker peptides. This step is essential for confirming their biological relevance and for establishing mechanistic links between pepsin activity and specific proteolytic cleavage events. In parallel, we will expand the study to include clinical cohorts, particularly patients with confirmed GERD, as well as individuals with oral or upper gastrointestinal cancers and patients undergoing chemotherapy, in whom reflux is a frequent adverse effect. Such cohorts will allow us to evaluate whether the digestion-associated peptide signatures identified here are detectable in real-world pathological conditions and whether they correlate with disease severity, symptom burden, or treatment-related reflux.

By integrating controlled experimental data with clinical validation, future studies will determine the diagnostic robustness, specificity, and translational potential of these salivary peptide markers. Ultimately, this research may contribute to the development of rapid, non-invasive, saliva-based diagnostic tools capable of complementing or reducing reliance on invasive procedures such as pH-monitoring or endoscopy.

5. Conclusions

This proof-of-concept study demonstrates that pepsin exposure under acidic conditions induces significant and reproducible alterations in the salivary peptidome, generating distinct peptide signatures that differentiate digested from native saliva. These findings provide foundational evidence supporting the feasibility of saliva-based, non-invasive diagnostics for reflux-associated enzymatic activity. However, several limitations of this proof-of-concept study should be considered. In particular, the human salivary proteome exhibits substantial proteome complexity, consisting of multiple proteoforms generated through proteolytic processing, post-translational modifications, and sequence variation. While MALDI-ToF peptide profiling enables rapid detection of digestion-associated peptide patterns, this approach does not resolve intact protein species and therefore cannot fully capture the diversity of proteoforms present in saliva. Consequently, the peptide signals reported here represent inferred fragments of the salivary proteome, and their identities remain putative until confirmed by tandem MS analysis. In addition, the relatively small sample size and the use of experimentally digested saliva from healthy volunteers limit the generalizability of the findings. Future studies will focus on structural identification of the candidate peptides through ESI-MS/MS and on validating these biomarkers in clinical cohorts, including patients with GERD and individuals with reflux-related symptoms secondary to oncologic therapies.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: Representative MALDI-ToF mass spectra obtained from three saliva samples digested in technical duplicate. (A) and (B) correspond to digested C3; (C) and (D) correspond to digested C4 and (E) and (F) correspond to digested C7; Figure S2: MALDI-ToF mass spectra of control saliva samples collected from the 10 individuals included in the study and Table S1: Comparison of MALDI-TOF MS–detected peptides with in silico pepsin digestion predictions. Peptides below 800 Da were not generated in silico and are marked as "*Not generated in silico*".

 doi:10.20944/preprints202604.0608.v1

Author Contributions: Conceptualization, D.E.F., R.I. and V.L.D.; methodology, B.A.P., A.N.; software, A.N. and B.U.; validation, B.A.P., B. U and A. N.; investigation, D.E.F., A. N., R.I and B.A.P.; resources, D.E.F., O.B.B., and B.A. P.; data curation, D.E.F., B.A.P., A.N. and B.U.; writing—original draft preparation, D.E.F. and O.B.B.; writing—review and editing, B.A.P., R.I., A.N,; and V.L.D.; supervision, V.L.D. and B.A.P.; project administration, V.L.D.. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. B.A.P. gratefully acknowledges the support received, in 2025, through the Fulbright Visiting Scholars Program, awarded by the Romania–USA Fulbright Commission, for her fellowship at Clarkson University, Potsdam, NY, USA.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by Grigore T. Popa University of Medicine and Pharmacy Iași Institutional Review Board (232/13.10.2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The MALDI ToF mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD076267.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GERD	Gastroesophageal reflux disease
MALDI-ToF	Matrix-Assisted Laser Desorption/Ionization-Time of Flight
m/z	Mass-to-charge ratio
rpm	Rotations per minute

References

1. El-Serag, H.B.; Sweet, S.; Winchester, C.C.; Dent, J. Update on the Epidemiology of Gastro-Oesophageal Reflux Disease: A Systematic Review. *Gut* **2014**, *63*, 871–880, doi:10.1136/gutjnl-2012-304269.
2. Wiklund, I. Review of the Quality of Life and Burden of Illness in Gastroesophageal Reflux Disease. *Dig. Dis.* **2004**, *22*, 108–114, doi:10.1159/000080308.
3. Gawron, A.J.; French, D.D.; Pandolfino, J.E.; Howden, C.W. Economic Evaluations of Gastroesophageal Reflux Disease Medical Management. *Pharmacoeconomics* **2014**, *32*, 745–758, doi:10.1007/s40273-014-0164-8.
4. Katz, P.O.; Dunbar, K.B.; Schnoll-Sussman, F.H.; Greer, K.B.; Yadlapati, R.; Spechler, S.J. ACG Clinical Guideline for the Diagnosis and Management of Gastroesophageal Reflux Disease. *Am. J. Gastroenterol.* **2022**, *117*, 27–56, doi:10.14309/ajg.0000000000001538.
5. Maresova, P.; Rezny, L.; Hruska, J.; Klimova, B.; Swanstrom, L.L.; Kuca, K. Diagnosis and Treatment of Patients with Gastroesophageal Reflux Disease – a Systematic Review of Cost-Effectiveness and Economic Burden. *BMC Health Serv. Res.* **2024**, *24*, 1351, doi:10.1186/s12913-024-11781-8.
6. Samuels, T.L.; Johnston, N. Pepsin in Gastroesophageal and Extraesophageal Reflux: Molecular Pathophysiology and Diagnostic Utility. *Curr. Opin. Otolaryngol. Head Neck Surg.* **2020**, *28*, 401–409, doi:10.1097/MOO.0000000000000664.
7. Han, F.; Li, X.; Song, Z.; Xie, J.; Wang, N.; Yao, J. The Association between Salivary Pepsin and Gastroesophageal Reflux Disease: A Meta-Analysis. *Neurogastroenterol. Motil.* **2024**, 1–10, doi:10.1111/nmo.14905.
8. Zhang, A.; Sun, H.; Wang, P.; Wang, X. Salivary Proteomics in Biomedical Research. *Clin. Chim. Acta.* **2013**, *415*, 261–265, doi:10.1016/j.cca.2012.11.001.
9. Zambonin, C.; Aresta, A. MALDI-TOF/MS Analysis of Non-Invasive Human Urine and Saliva Samples for the Identification of New Cancer Biomarkers. *Molecules* **2022**, *27*, doi:10.3390/molecules27061925.
10. MS-Digest, Prospector v.5.22.1, UCSF Mass Spectrometry Facility, University of California, San Francisco. Available at: Https://Prospector.Ucsf.Edu/Prospector/Cgi-Bin/Msform.Cgi?Form=msdigest.

 doi:10.20944/preprints202604.0608.v1

11. Humphrey, S.P.; Williamson, R.T. A Review of Saliva: Normal Composition, Flow, and Function. *J. Prosthet. Dent.* **2001**, *85*, 162–169, doi:10.1067/mpr.2001.113778.
12. Millipore Corporation. ZipTip® Pipette Tips: Protocol for Peptide Purification and Desalting. Millipore; 2010. Available from: Http://Www.Millipore.Com/Catalogue/Module/C5737.
13. Perez-Riverol, Y.; Bandla, C.; Kundu, D.J.; Kamatchinathan, S.; Bai, J.; Hewapathirana, S.; John, N.S.; Prakash, A.; Walzer, M.; Wang, S.; et al. The PRIDE Database at 20 Years: 2025 Update. *Nucleic Acids Res.* **2025**, *53*, D543–D553, doi:10.1093/nar/gkae1011.
14. Dongiovanni, P.; Meroni, M.; Aiello, G.; D'Amato, A.; Cenzato, N.; Casati, S.; Damiani, G.; Fenoglio, C.; Galimberti, D.; Grossi, E.; et al. Salivary Proteomic Profile of Young Healthy Subjects. *Front. Mol. Biosci.* **2023**, *10*, 1327233, doi:10.3389/fmolb.2023.1327233.
15. Shao, C.; Tian, Y.; Dong, Z.; Gao, J.; Gao, Y.; Jia, X.; Guo, G.; Wen, X.; Jiang, C.; Zhang, X. The Use of Principal Component Analysis in MALDI-TOF MS: A Powerful Tool for Establishing a Mini-Optimized Proteomic Profile. *Am. J. Biomed. Sci.* **2012**, *4*, 85–101, doi:10.5099/aj120100085.
16. Balakrishnama, S.; Ganapathiraju, A. Linear Discriminant Analysis-a Brief Tutorial. In *Institute for Signal and Information Processing*; 1998; pp. 1–8.
17. Nasteski, V. An Overview of the Supervised Machine Learning Methods. *Horizons. b* **2017**, *4*, 56.
18. Gronau, Q.F.; Wagenmakers, E.-J. Limitations of Bayesian Leave-One-Out Cross-Validation for Model Selection. *Comput. brain Behav.* **2019**, *2*, 1–11, doi:10.1007/s42113-018-0011-7.
19. Surdu, A.; Foia, L.G.; Luchian, I.; Trifan, D.; Tatarciuc, M.S.; Scutariu, M.M.; Ciupilan, C.; Budala, D.G. Saliva as a Diagnostic Tool for Systemic Diseases-A Narrative Review. *Medicina (Kaunas).* **2025**, *61*, doi:10.3390/medicina61020243.
20. Miller, C.S.; Foley, J.D.; Bailey, A.L.; Campell, C.L.; Humphries, R.L.; Christodoulides, N.; Floriano, P.N.; Simmons, G.; Bhagwandin, B.; Jacobson, J.W.; et al. Current Developments in Salivary Diagnostics. *Biomark. Med.* **2010**, *4*, 171–189, doi:10.2217/bmm.09.68.
21. Jing, W.; Luo, W.; Lou, L. Diagnostic Utility of Salivary Pepsin in Laryngopharyngeal Reflux: A Systematic Review and Meta-Analysis. *Braz. J. Otorhinolaryngol.* **2023**, *89*, 339–347, doi:https://doi.org/10.1016/j.bjorl.2022.10.050.
22. Guo, Z.; Wu, H.; Jiang, J.; Zhang, C. Pepsin in Saliva as a Diagnostic Marker for Gastroesophageal Reflux Disease: A Meta-Analysis. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **2018**, *24*, 9509–9516, doi:10.12659/MSM.913978.
23. Tan, J.-J.; Dai, Y.-F.; Wang, F.; Lv, Z.-H.; Huang, L.-J.; Peng, L.-Y.; Li, X.-P. Pepsin-Mediated Inflammation in Laryngopharyngeal Reflux via the ROS/NLRP3/IL-1β Signaling Pathway. *Cytokine* **2024**, *178*, 156568, doi:https://doi.org/10.1016/j.cyto.2024.156568.
24. Stanforth, K.J.; Wilcox, M.D.; Chater, P.I.; Brownlee, I.A.; Zakhour, M.I.; Banecki, K.M.R.M.; Pearson, J.P. Pepsin Properties, Structure, and Its Accurate Measurement: A Narrative Review. *Ann. Esophagus; Vol 5 (September 25, 2022) Ann. Esophagus* **2021**.
25. Helmerhorst, E.J.; Oppenheim, F.G. Saliva: A Dynamic Proteome. *J. Dent. Res.* **2007**, *86*, 680–693, doi:10.1177/154405910708600802.
26. Schulz, B.L.; Cooper-White, J.; Punyadeera, C.K. Saliva Proteome Research: Current Status and Future Outlook. *Crit. Rev. Biotechnol.* **2013**, *33*, 246–259, doi:10.3109/07388551.2012.687361.
27. Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, *537*, 347–355, doi:10.1038/nature19949.
28. Ceccherini, E.; Michelucci, E.; Signore, G.; Coco, B.; Zari, M.; Bellini, M.; Brunetto, M.R.; Cecchettini, A.; Rocchiccioli, S. The Clinical Utility of the Saliva Proteome in Rare Diseases: A Pilot Study for Biomarker Discovery in Primary Sclerosing Cholangitis. *J. Clin. Med.* 2024, *13*, 544.
29. Lazari, L.C.; Zerbinati, R.M.; Rosa-Fernandes, L.; Santiago, V.F.; Rosa, K.F.; Angeli, C.B.; Schwab, G.; Palmieri, M.; Sarmento, D.J.S.; Marinho, C.R.F.; et al. MALDI-TOF Mass Spectrometry of Saliva Samples as a Prognostic Tool for COVID-19. *J. Oral Microbiol.* **2022**, *14*, 2043651, doi:10.1080/20002297.2022.2043651.
30. Xiao, H.; Zhang, Y.; Kim, Y.; Kim, S.; Kim, J.J.; Kim, K.M.; Yoshizawa, J.; Fan, L.-Y.; Cao, C.-X.; Wong, D.T.W. Differential Proteomic Analysis of Human Saliva Using Tandem Mass Tags Quantification for Gastric Cancer Detection. *Sci. Rep.* **2016**, *6*, 22165, doi:10.1038/srep22165.

 doi:10.20944/preprints202604.0608.v1

31. Singhal, N.; Kumar, M.; Kanaujia, P.K.; Virdi, J.S. MALDI-TOF Mass Spectrometry: An Emerging Technology for Microbial Identification and Diagnosis. *Front. Microbiol.* **2015**, *6*, 791, doi:10.3389/fmicb.2015.00791.
32. Clark, A.E.; Kaleta, E.J.; Arora, A.; Wolk, D.M. Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: A Fundamental Shift in the Routine Practice of Clinical Microbiology. *Clin. Microbiol. Rev.* **2013**, *26*, 547–603, doi:10.1128/CMR.00072-12.
33. Sujatha, S.; Jalihal, U.; Devi, Y.; Rakesh, N.; Chauhan, P.; Sharma, S. Oral PH in Gastroesophageal Reflux Disease. *Indian J. Gastroenterol. Off. J. Indian Soc. Gastroenterol.* **2016**, *35*, 186–189, doi:10.1007/s12664-016-0659-7.
34. Bechir, F.; Pacurar, M.; Tohati, A.; Bataga, S.M. Comparative Study of Salivary PH, Buffer Capacity, and Flow in Patients with and without Gastroesophageal Reflux Disease. *Int. J. Environ. Res. Public Health* **2021**, *19*, doi:10.3390/ijerph19010201.
35. Preetha, A.; Sujatha, D.; Patil, B.A.; Hegde, S. Oral Manifestations in Gastroesophageal Reflux Disease. *Gen. Dent.* **2015**, *63*, e27-31.
36. Alharthi, R.; Mutahar, M.; Bartlett, D.; Jafari, J.; Moazzez, R. Salivary Proteins in Human Acquired Enamel Pellicle (AEP) on Eroded and Uneroded Teeth in Patients with Gastro-Oesophageal Reflux Disease (GORD). *Dent. J.* 2024, *12*, 235.
37. Martini, T.; Rios, D.; Cassiano, L.P.S.; Silva, C.M. de S.; Taira, E.A.; Ventura, T.M.S.; Pereira, H.A.B.S.; Magalhães, A.C.; Carvalho, T.S.; Baumann, T.; et al. Proteomics of Acquired Pellicle in Gastroesophageal Reflux Disease Patients with or without Erosive Tooth Wear. *J. Dent.* **2019**, *81*, 64–69, doi:10.1016/j.jdent.2018.12.007.
38. Martini, T.; Câmara, J.V.F.; Dionizio, A.; Ventura, T.M.O.; Cassiano, L. de P.S.; de Souza e Silva, C.M.; Taira, E.A.; Araujo, T.T.; Santos, L.A.; Ferrari, C.R.; et al. Proteomic Analysis of Stimulated Saliva in Gastroesophageal Reflux Disease Patients with and without Erosive Tooth Wear: Observational Study. *J. Dent.* **2023**, *139*, 104724, doi:https://doi.org/10.1016/j.jdent.2023.104724.