Published in final edited form as:

J Clin Periodontol. 2012 March ; 39(3): 203–212. doi:10.1111/j.1600-051X.2011.01805.x.

Novel protein identification methods for biomarker discovery via a proteomic analysis of periodontally healthy and diseased gingival crevicular fluid samples

Richard C. Baliban^{*}, Dimitra Sakellari[#], ZuKui Li^{*}, Peter A. DiMaggio[&], Benjamin A. Garcia[&], and Christodoulos A. Floudas^{*}

^{*}Department of Chemical and Biological Engineering, Princeton University, Princeton, USA

[#]Department of Preventive Dentistry, Periodontology and Implant Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece

[&]Department of Molecular Biology, Princeton University, Princeton, USA

Abstract

Aim—To identify possible novel biomarkers in gingival crevicular fluid (GCF) samples from chronic periodontitis (CP) and periodontally healthy individuals by high-throughput proteomic analysis.

Materials and Methods—GCF samples were collected from twelve CP and twelve periodontally healthy subjects. Samples were typically digested, eluted using high-performance liquid chromatography, and fragmented using tandem mass spectrometry (MS/MS). MS/MS were analyzed using PILOT_PROTEIN to identify all unmodified proteins within the samples.

Results—Using the database derived from Homo sapiens taxonomy and all bacterial taxonomies, 432 human (120 new) and 30 bacterial proteins were identified. The human proteins, angiotensinogen, clusterin and thymidine phosphorylase were identified as biomarker candidates based on their high-scoring only in samples from periodontal health. Similarly, neutrophil defensin-1, carbonic anhydrase-1 and elongation factor-1 gamma were associated with CP. Candidate bacterial biomarkers include 33 kDa chaperonin, iron uptake protein A2 and phosphoenolpyruvate carboxylase (health-associated) and ribulose biphosphate carboxylase, a probable succinyl-CoA:3-ketoacid-coenzyme A transferase, or DNA-directed RNA polymerase subunit beta (CP-associated). Most of these human and bacterial proteins have not been previously evaluated as biomarkers of periodontal conditions and require further investigation.

Conclusions—The proposed methods for large-scale comprehensive proteomic analysis may lead to the identification of novel biomarkers of periodontal disease.

Keywords

periodontitis; gingival crevicular fluid; proteomic analysis; tandem mass spectrometry; biomarkers

Corresponding author Christodoulos A. Floudas, Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ, 08544, USA., Tel: (609) 258-4595, Fax: (609) 258-0211, floudas@titan.princeton.edu.

Conflict of interest

The authors declare that there are no conflicts of interest in this study.

Introduction

The search for biomarkers which can act as predictors of periodontal disease at the initiation and progression stage has received considerable interest during the last decade (Champagne et al. 2003, Loos & Toja 2005). The diagnostic potential of Gingival Crevicular Fluid (GCF) has been extensively investigated due to the possibility of non-invasive collection and the complexity of molecules that it contains (Buduneli & Kinane 2011). GCF has been shown to be the transudate of gingival tissue interstitial fluid, but during periodontal disease it is transformed into inflammatory exudate which reflects the composition of serum and includes substances derived from the structural tissues of the periodontium and oral bacteria colonizing the gingival pocket (Delima & Van Dyke 2003).

Several substances (up to 90) including cytokines, proteolytic enzymes, bacterial-derived metabolites, or products of tissue degradation have been investigated as possible indicators or predictors of disease activity, but currently no chairside tests exist that can be reliably applied for accurate diagnosis or prognosis in clinical practice (Champagne et al. 2003, Eley & Cox 2003, Uitto et al. 2003, Lamster & Ahlo 2007). The introduction of large-scale proteomic analysis in host-derived clinical samples such as serum or saliva is an innovative approach that could greatly enhance current knowledge of the proteins involved in health or disease (Loo et al. 2010), but limited data exist in the literature for GCF. Various mass spectrometry techniques have been applied to identify mainly targeted proteins such as the defensins (Dommish et al. 2005, Lundy et al. 2005) or the acid-soluble protein content of GCF (Pisano et al. 2005). Recently, tandem mass spectrometry (MS/MS) techniques have been applied to perform large-scale proteomic analysis of GCF, utilizing gel electrophoresis (Ngo et al. 2010) for protein separation or "shotgun" approaches (Bostanci et al. 2010, Grant et al. 2010). These reports refer to periodontal health and disease (Bostanci et al. 2010), periodontal patients at maintenance phase (Ngo et al. 2010), or investigated changes during the inflammatory process in an experimental gingivitis model (Grant et al. 2010) and have shown an abundance of mainly host-derived proteins in clinical samples. They suggested that studies of GCF are required to determine the composition in periodontal health and disease and identify potential biomarkers.

Using "bottom-up" proteomics, proteins are enzymatically digested, separated based upon their hydrophobicity, vaporized, and then protonated via electrospray ionization. Each peptide is isolated in the mass spectrometer and then fragmented using collision-induced dissociation to generate a MS/MS spectrum that contains the amino acid composition information of that peptide. Using the resulting **b** and **y** ion series in the MS/MS spectrum, the peptide sequence can be derived using either de novo (Frank & Pevzner 2005, DiMaggio & Floudas 2007a,b), database (Eng et al. 1994, Perkins et al. 1999), or hybrid de novo/database methods (Tanner et al. 2005, DiMaggio et al. 2008). To determine the protein sequence from the peptide information, a database must be used to match the peptide annotation from the MS/MS spectrum to a theoretically digested peptide from the list of proteins (Nesvizhskii 2010). Once the list of proteins has been identified for a cellular sample, a search for all post-translational modifications (PTMs) may be performed to identify all PTM types and sites that are present on the proteins (Witze et al. 2007, Baliban et al. 2010).

The aim of the present study was to conduct a comprehensive proteomic analysis of GCF samples from periodontally healthy and chronic periodontitis (CP) subjects utilizing the PILOT_PROTEIN protein identification methods (DiMaggio & Floudas 2007a,b, DiMaggio et al. 2008, Baliban et al., 2010, Baliban et al. 2011) and the recently developed webtool (<http://pumpd.princeton.edu>). This study will show the complete lists of human and bacterial

proteins found in all twenty four samples and will uncover candidates for both human and bacterial biomarkers.

Materials and Methods

Subject sample collection

Twelve periodontally healthy and twelve periodontally diseased and non-previously treated subjects participated in the present study. All participants were patients of the Department of Preventative Dentistry, Periodontology, and Implant Biology, Dental School, Aristotle University, Thessaloniki, Greece or were personnel of the Dental School. Demographic and clinical data for participants are presented in Table 1.

All subjects were systematically healthy, non-smokers, and were not taking medication known to affect periodontal tissues. Subjects reporting antibiotic intake during the previous six months and pregnant or lactating women were excluded from the present study. Subjects were considered as periodontally healthy cases when they displayed bleeding on probing at less than 10% of the examined sites, no probing depth or probing attachment level greater than 3 mm, and no radiographic signs of bone loss. Subjects were considered as chronic periodontitis cases according to the analytical criteria of the American Academy of Periodontology (Armitage 1999). Care was taken to include age-matched individuals across the two groups. All subjects signed an informed consent, and the study was conducted according to the protocol outlined by the Research Committee, Aristotle University of Thessaloniki, Greece, and approved by the Ethical Committee of the School of Dentistry.

Clinical recording and sampling

Clinical recordings were performed by a calibrated examiner (DS) using a manual Williams probe (POW, Hu-Friedy, Chicago, IL). The examiner has reproducible recordings (Pearson's test $r=0.971$) as determined in 10% of her weekly registrations. Parameters assessed included probing depth, recession, and bleeding on probing at six sites of all teeth present in the dentition. For clinical parameters, the statistical analysis of the data was carried out with the statistical package SPSS (14.0 version). Indicators of Descriptive Statistics were used, including mean and standard deviation for each group with the patient as the observational unit. Differences in clinical parameters were sought by applying a Mann-Whitney test, with a significance level of 0.05 (Table 1).

Gingival crevicular fluid samples

Each participant contributed with one pooled GCF sample. For periodontitis cases, the sample was taken from four pre-selected sites which displayed probing depth >6 mm and <8 mm. For periodontally healthy individuals, the samples were taken from the mesiobuccal sites of first molars. GCF samples were obtained as previously described (Sakellari et al. 2008). The samples were immediately placed in Eppendorf tubes containing 100 μ L of 100 mM ammonium bicarbonate, frozen in liquid nitrogen, and stored at -80 °C. GCF samples were collected prior to the clinical measurements and were discarded when visibly contaminated with blood. Samples were lyophilized for 5 hours at -55 °C and 0.03 mbar in an ALPHA 1–4 (Martin Christ, Gefriertrocknungsanlagen GmbH) lyophilizer. Prior to lyophilization, they were vortexed for 20 min and centrifuged for 10 min at 8,161.4 g while all filter strips were discarded.

Sample preparation for mass spectrometry

The lyophilized protein samples were reconstituted in 100 μ L of water. Sample concentrations (μ g/ μ L) were estimated using a Bradford assay. The samples were treated using 10 μ L of 55 mM dithiothreitol in a 51 °C sand bath for one hour and then alkylated in

the dark for 45 minutes using 10 μ L of 168 mM iodoacetimide. Alkylation was quenched using 1 μ L of the 55 mM dithiothreitol. The sample pH was adjusted to 8.0 using ammonium bicarbonate and the proteins were digested using sequencing grade trypsin (Sigma Aldrich) in a 1:20 (w/w) ratio. Digestion was performed using an eight hour incubation time in a sand bath at 37 °C and then quenched using glacial acetic acid in a 1:10 (v/v) ratio. The digested samples were extracted for MS/MS analysis as follows. C18 STAGE tips were activated using 20 μ L of methanol and then washed using 20 μ L of 0.1% acetic acid. The samples were loaded onto the C18 disks, washed with 20 μ L of 0.1% acetic acid, and then eluted with 20 μ L of 75% MeCN/5% acetic acid. The samples were dried down in a Speed Vac to approximately a 5 μ L volume.

Liquid chromatography and mass spectrometry

Nanoflow liquid chromatography tandem mass spectrometry (LC-MS/MS) was performed on a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer (Thermo Electron, San Jose, CA) coupled to an Agilent 1200 Series binary HPLC pump (Agilent Technologies, Palo Alto, CA) and an Eksigent AS2 autosampler (Eksigent Technologies). 0.4–1.2 μ L of protein sample (corresponding to approximately 10 μ g of protein) was diluted to 4.0 μ L in 0.1% acetic acid. 2.0 μ L of the sample volume was loaded on a 75 μ m inner diameter fused silica capillary column constructed with an integrated electrospray tip packed with 15 cm of C18 reversed phase resin (Magic C18, 5 μ m particles, 200 Å pore size; Michrom BioResources, Auburn, CA). Peptides were separated by RP-HPLC using a gradient from 2% to 45% Buffer B (Buffer A, 0.1 M acetic acid; Buffer B, 70% acetonitrile in 0.1 M acetic acid) at a flow rate of 200 nL/min for 110 min. The Orbitrap instrument was operated in data-dependent mode using a resolution of 30,000 to obtain a full MS spectrum followed by seven MS/MS spectra obtained in the ion trap. The MS scans were collected with an automatic gain control target value of 5×10^5 and maximum injection time of 100 ms over a mass range of 300–1650 m/z. MS/MS scans were collected using an automatic gain control value of 4×10^4 and a threshold energy of 35% for collision-induced dissociation.

Mass spectrometry data analysis

All MS/MS spectra were processed using the on-line version of the PILOT_PROTEIN protein identification method and its webtool (Baliban et al. 2011). A hybrid de novo/ database MS/MS search (DiMaggio et al. 2008) with protein identification (Baliban et al. 2011) and an untargeted PTM search was utilized (Baliban et al. 2010). Search tolerances included a value of 0.1 Da for the precursor ion and 0.5 Da for the fragment ion. Searches were performed using a maximum of 2 missed cleavages and a static cysteine modification of 57 Da due to the iodoacetimide treatment. The de novo sequences (DiMaggio & Floudas 2007a,b) were searched against two databases. The first database is derived from the swissprot database and uses a combination of Homo sapiens and all bacterial taxonomies. The second database is a subset of the first and uses only those bacterial taxonomies identified to be associated with periodontal health or periodontal disease (Socransky et al. 1998). These bacterial genera include Porphyromonas, Bacteroides, Treponema, Fusobacterium, Prevotella, Campylobacter, Eubacterium, Streptococcus, Capnocytophaga, Eikenella, Actinomyces, Selenomonas, and Aggregatibacter along with the bacterial family Actinobacteria.

Results

The complete list of peptide and protein identifications for all twenty four samples obtained from PILOT_PROTEIN is given as Supplementary Information. No PTMs passing the threshold scoring criterion were found using the PILOT_PTM untargeted search algorithm.

Human proteins

Using the database derived from *Homo sapiens* taxonomy and all bacterial taxonomies, a total of 432 human proteins were identified (Table 2). The complete list of human proteins identified is presented in Supplementary Table 1. The 432 human proteins are separated into three groups, depending on whether they were found in both the CP and healthy samples, only in the healthy samples, or only in the CP samples. 230 of the human proteins were found to be in both the healthy and CP samples while 123 human proteins were only found in the CP samples and 79 were only in the healthy samples. Table 3 shows the list of all proteins that are identified in only the healthy samples and only in the CP samples at least three times. Next to each protein is listed the average amount of peptides and average amount of spectra (MS/MS) that were annotated across all samples. It is important to note that the identification of the proteins was solely based on the identification of the corresponding peptides. The spectral counts for each protein are indicative of how many pieces of MS/MS data belong to a particular protein, but were not used to determine the relative abundance of the proteins.

Several of the proteins in Supplementary Table 1 are identified using only one peptide. To demonstrate the effect of a stricter protein selection criterion (Carr et al. 2010), Supplementary Table 2 shows the human protein list identified when a minimum of two peptides must be annotated for each protein. When comparing all proteins identified by a minimum of two peptides from Supplementary Table 2 to those from Table 3 of Bostanci et al. (2010), a total of 120 new proteins are found and are listed in boldface in Supplementary Table 2. The authors note that only two publications were found comparing periodontally healthy and periodontally diseased GCF samples (i.e., Bostanci et al. and Grant et al.). The comparison to Grant et al. was not performed because the diseased proteins were determined using a 21 day gingivitis model and therefore are not categorically equivalent to the chronic diseased proteins in our study.

When using this stricter protein selection criterion, it is interesting to note that all proteins in Table 4 except glucosamine-6-phosphate deaminase and argininosuccinate lyase remain the same while many of the proteins listed in Table 3 are reduced. The suggested biomarkers for bacterial proteins remain the same while the suggested biomarkers for human proteins now become angiotensinogen, thymidine phosphorylase, and clusterin for the periodontally healthy samples and carbonic anhydrase 1 for the chronic periodontitis samples.

Bacterial proteins

Using the database without restriction on bacterial taxonomy, a total of 30 bacterial proteins were found (Table 4). 17 bacterial proteins were found within the healthy samples, and 13 bacterial proteins were found within the CP samples (Table 2). There were no proteins identified that existed within both the healthy and CP samples. This is very beneficial for biomarker analysis since it provides a clear distinction between the results of the healthy patients and those of the CP patients. For the healthy samples, there are 7 proteins that appeared at least five times throughout the twelve samples. A total of 9 proteins appeared at least five times for the CP samples. The total amount of human or bacterial proteins present in at least five or six healthy or CP samples is shown in Table 6. Note that the threshold number of samples listed in Tables 3 – 6 is intended to provide a more thorough visualization of a smaller list of potential biomarker proteins. The comprehensive list of human (Supplementary Table 1) and bacterial (Tables 4 and 5) proteins are provided and the reader can impose stricter conditions for biomarker analysis if desired.

Though the bacterial proteins in Table 4 present a clear distinction between the healthy and CP samples, none of the proteins is part of a genus that has been previously associated with

periodontal health or disease (Socransky et al. 1998). Thus, a secondary analysis using a targeted bacterial database was used to analyze the experimental data. Though the resulting human protein list did not change, the bacterial protein list was altered (Table 5). A total of 20 proteins were identified, with 3 belonging to the healthy samples, 15 to the CP samples, and 2 to both the healthy and CP samples.

Gene ontology

The gene ontology for all identified human proteins (Supplementary Table 1) from both the healthy and CP groups is shown in Figure 1. All relevant biological processes, cellular components, and molecular functions were extracted from the UniProt database for each protein. The total number of proteins found for each ontological feature is shown in Supplementary Table 3 for the healthy groups and Supplementary Table 4 for the CP groups. The healthy samples recorded a total of 357 biological processes, 67 cellular components, and 127 molecular functions. The CP samples had 348 biological processes, 69 cellular components, and 136 molecular functions. For each major category, all ontological features that had a minimum number of protein hits are plotted in Figure 1. For biological processes, this minimum was set to 10 proteins while for cellular components and molecular functions, the minimum was set to 5 proteins.

Discussion

Human and bacterial protein identification

Among the larger number of proteins found in both healthy and CP samples are serum albumin, serotransferrin, and α -2-macroglobulin, reflecting the serum origin of GCF as also previously identified (Curtis et al 1990). A variety of apolipoproteins were also detected in both healthy and CP samples. Of particular interest is the fact that apolipoprotein-J (clusterin) is reported for the first time in GCF in the present study and was only detected in periodontally healthy sites. The presence of apolipoprotein-B was recently confirmed in GCF (Sakiyama et al. 2010) and apolipoprotein L was also detected in healthy sites by Bostanci et al. (2010).

Important components of the immune system such as the immunoglobulins and complement were also widely detected in both groups. Thirty-one proteins related to IgA, all four IgG subclasses, and a group of twelve components, subcomponents, and factors of the complement system were frequently identified. The transudatory nature of GCF is also reflected by the four different hemoglobins.

Type I and type II keratins have been identified in both healthy and CP samples in agreement with earlier studies (McLaughlin et al. 1996, Bostanci et al. 2010, Grant et al. 2010) and support the high turnover and rate of differentiation of oral epithelia which would result in an expanded number of cytoskeletal proteins in GCF. The presence of certain keratins (notably Keratin 9) as a result of contamination during the experimental procedure should be acknowledged (Moll et al. 2008).

Other intracellular proteins related to the cytoskeleton and its functions have also been identified. Findings from the present study confirm and expand previous ones (Ngo et al. 2010), since eight actin and actin-related proteins, as well as gelsolin, profilin-1, and cofilin-1 have been identified. In the present study these proteins have been frequently identified in healthy samples, suggesting that they are components of desquamative cells from the gingival sulcus. The actin-binding proteins plastin-1 and plastin-2, the latter being indicative of the presence of leucocytes, have been identified in investigated samples. Other structural elements of cells, such as alpha and beta tubulins, myosin-9 (a protein known to affect cell shape and cytokinesis and a ligand for actin), and components of the extracellular

matrix such as desmoplakin, fibronectin, vitronectin, and vimentin were frequently found in both groups.

Eleven histone compounds were identified in agreement with previous recent studies applying proteomic analysis approaches (Ngo et al. 2010, Bostanci et al. 2010). These proteins have been shown to possess antibacterial properties by being components of "Neutrophil Extracellular Traps", a defense mechanism recently described in the GCF of periodontitis patients, which could assist in bacterial clearance of the gingival crevice (Vitkov et al. 2009). Identification of several histones suggests the presence of neutrophils in both clinical situations.

Cytokines, with the notable exception of Interleukin-1 receptor antagonist, were not detected in agreement with previous studies applying proteomic approaches (Bostanci et al. 2010, Grant et al. 2010, Ngo et al. 2010). Possible explanations for this finding have been reported in the literature and include the possibility that previously applied methodologies (e.g., immunoassays) may not discriminate between the protein or peptide fragments and that these substances have low, sub-attomolar concentrations in GCF (Puklo et al. 2008, Grant et al. 2010).

Matrix metalloproteinases MMP-8 and MMP-9 and neutrophil gelatinase-associated lipocalin have been frequently identified in both groups. Although MMP-8 and MMP-9 were detected almost twice more frequently in CP samples, they were also detected in periodontally healthy samples, in contrast to the findings of Bostanci et al. (2010). Leucocyte elastase and cathepsins G and D which are associated with neutrophils have also been previously investigated as candidate biomarkers (Uitto et al. 2003, Budunelli & Kinane 2011). These substances were detected in both healthy and CP samples, although cathepsin G appeared four times more frequently in samples from CP subjects.

A substantial number of antimicrobial proteins and peptides (AMPs), representing all functional classes (Gorr 2009, Gorr & Abdolhosseini 2011) were identified. The cationic peptide neutrophil defensin-1 (HNP-1) was detected in half of the CP samples, suggesting an abundance of these cells in the periodontal pocket. In contrast, Bostanci et al. (2010) identified neutrophil-defensins (1–3) only in healthy sites, supporting the idea that the progression of periodontal disease could be correlated with a decrease in the defensins. Note that their samples originated from aggressive periodontitis cases (known to possibly involve neutrophil dysfunction) and not chronic periodontitis as in the present study (Nussbaum & Shapira 2011). Cathelicidin (LL-37) has also been identified in both healthy and CP samples, but important members of this group of AMPs such as the β -defensins and the histatins (1 and 3) were not detected. The antimicrobial peptides lactotransferrin, serotransferrin, calgranulin-A and -B (proteins S100-A8 and -A9, respectively) which act as metal-ion chelators, members of the protease inhibitors cystatins (A, B, S, and SN), as well as lysozyme-C and myeloperoxidase have also been identified regardless of periodontal status.

Members of the Annexin family (1, 2, 3, and 8) which are related to the inflammatory process were also identified with a tendency to be rather associated with periodontal health. Findings from the studies by Bostanci et al (2010) and Grant et al. (2010) also suggest a possible role of these proteins in maintaining periodontal health which require further documentation. Summarizing, all functional classes of antimicrobial proteins are frequently found in both healthy and CP samples, which was expected due to the continuous presence of bacterial challenge in the gingival crevice in both clinical situations.

Although the bacterial proteins identified were distinct between samples from diseased and non- diseased subjects and thus offer an advantage for biomarker analysis, they belonged to

species not associated with the periodontal environment (Table 4). By applying a more targeted approach (Table 5) proteins deriving from putative periodontal pathogens such as Treponema, Fusobacterium, and Campylobacter species were identified albeit in a small number of samples.

Biomarker analysis

Through a comprehensive proteomic analysis, and in contrast to the traditional more "targeted" approaches previously applied in GCF studies, biomarkers or set of biomarkers were sought which could identify periodontal disease presence or absence. The list of all human biomarkers with a minimum sample number of 5 and 6 is detailed as part of Table 3 and summarized in Table 6. Candidate human proteins biomarkers for periodontal health include angiotensinogen (i.e., the uncleaved component of the rennin-angiotensin system; 9 samples), clusterin (i.e., an apolipoprotein related to removal of cellular debris and apoptosis; 5 samples) and thymidine phosphorylase (i.e., an enzyme that may have a role in maintaining the integrity of the blood vessels, and which demonstrates growth promoting activity on endothelial cells; 6 samples). Candidate biomarkers for CP include neutrophil derived defensin-1 (6 samples), carbonic anhydrase-1 (an enzyme far less investigated in oral conditions than carbonic-anhydrase-6 found in saliva; 7 samples) and elongation factor 1-gamma (i.e., a protein known to facilitate the events of translational elongation and thus suggest an increased metabolic activity in the periodontal pocket; 5 samples).

Candidate bacterial biomarkers are chosen using the comprehensive bacterial database. Bacterial biomarkers are displayed in Table 4 with a summary provided in Table 6. No bacterial biomarkers that appeared in at least 5 samples were identified using the targeted bacterial database (Table 5). The top health-associated candidate proteins would be the 33 kDa chaperonin (9 samples), iron uptake protein A2 (7 samples), and phosphoenolpyruvate carboxylase (7 samples). The top CP-associated candidates would be ribulose biphosphate carboxylase (12 samples), a probable succinyl-CoA:3-ketoacid-coenzyme A transferase (10 samples), or DNA-directed RNA polymerase subunit beta (8 samples). Most of the above mentioned human and bacterial proteins have not been previously evaluated as biomarkers of periodontal conditions and therefore require further investigation.

In conclusion, a comprehensive proteomic analysis of twenty four GCF samples from periodontally healthy and CP subjects was performed using liquid chromatography and tandem mass spectrometry. The LC-MS/MS samples were analyzed using the high-throughput modified protein identification algorithm PILOT_PROTEIN to determine the comprehensive list of unmodified proteins in both groups. From the list of human proteins, three proteins from each group were identified as biomarker candidates based on their frequent appearance in only one of either the healthy or CP sample sets. Two distinct sets of bacterial proteins were identified based on the use of a database containing (i) all bacterial taxonomies or (ii) a targeted subset of bacterial taxonomies. The bacterial proteins identified in both data sets differed, and no crossover of proteins was observed. The potential for bacterial biomarkers is very high as very few of the proteins were found in both the diseased and healthy samples.

Clinical Relevance

Scientific rationale for study

Although gingival crevicular fluid has been extensively investigated, no biomarkers for diagnosis or prognosis of periodontal disease have been currently identified.

Principal findings

A total of 432 human (120 new) and 30 bacterial proteins were identified in GCF from periodontally healthy and chronic periodontitis subjects and novel candidate biomarkers were identified in both clinical situations.

Practical implications

High-throughput proteomic analysis of gingival crevicular fluid may lead to the discovery of novel biomarkers for periodontal disease which, after proper validation, could reach clinical praxis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding Statement

C.A.F. and B.A.G. acknowledge financial support from the National Science Foundation (CBET-0941143). C.A.F. acknowledges financial support from the National Institute of Health (R01LM009338). Although the research described in the article has been funded in part by the U.S. Environmental Protection Agency's STAR program through grant (R832721-010), it has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. B.A.G. acknowledges support from Princeton University and the American Society for Mass Spectrometry Research award.

References

- Armitage GC. Development of a classification system for periodontal diseases and conditions. *Annals of Periodontology*. 1999; 4:1–6. [PubMed: 10863370]
- Baliban RC, DiMaggio PA, Plazas-Mayorca MD, Young NL, Garcia BJ, Floudas CA. A novel approach for untargeted post-translational modification identification using integer linear optimization and tandem mass spectrometry. *Molecular & Cellular Proteomics*. 2010; 9:764–779. [PubMed: 20103568]
- Baliban RC, DiMaggio PA, Plazas-Mayorca MD, Garcia BJ, Floudas CA. Modified protein identification using high-resolution tandem mass spectrometry and integer linear optimization. *Molecular & Cellular Proteomics*. 2011 submitted.
- Bostanci N, Heywood W, Mills K, Parkar M, Nibali L, Donos N. Application of label-free absolute quantitative proteomics in human gingival crevicular fluid by LC/MS E (gingival exudatome). *Journal of Proteome Research*. 2010; 9:2191–2199. [PubMed: 20205380]
- Buduneli N, Kinane DF. Host-derived diagnostic markers related to soft tissue destruction and bone degradation in periodontitis. *Journal of Clinical Periodontology*. 2011; 38:85–105. [PubMed: 21323706]
- Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A. The Need for Guidelines in Publication of Peptide and Protein Identification Data. *Molecular & Cellular Proteomics*. 2004; 3:531–533. [PubMed: 15075378]
- Champagne CM, Buchanan W, Reddy MS, Preisser JS, Beck JD, Offenbacher S. Potential for gingival crevice fluid measures as predictors of risk for periodontal diseases. *Periodontology 2000*. 2003; 31:167–180. [PubMed: 12657001]
- Curtis MA, Sterne JAC, Price SJ, Grittiths GS, Coulthurst SK, Wilton JMA, Johnson NW. The protein composition of gingival crevicular fluid sampled from male adolescents with no destructive periodontitis: Baseline data of a longitudinal study. *Journal of Periodontal Research*. 1990; 25:6–16. [PubMed: 2137172]
- Delima AJ, Van Dyke TE. Origin and function of the cellular components in gingival crevice fluid. *Periodontology 2000*. 2003; 31:55–76. [PubMed: 12656996]
- DiMaggio PA Jr, Floudas CA. De Novo Peptide Identification via Tandem Mass Spectrometry and Integer Linear Optimization. *Analytical Chemistry*. 2007a; 79:1433–1446. [PubMed: 17297942]

- DiMaggio PA Jr, Floudas CA. A Mixed-Integer Optimization Framework for De Novo Peptide Identification. *AIChE Journal*. 2007b; 53:160–173. [PubMed: 19412358]
- DiMaggio PA Jr, Floudas CA, Lu B, Yates JR III. A Hybrid Method for Peptide Identification Using Integer Linear Optimization, Local Database Search, and Quadrupole Time-of-Flight or Orbitrap Tandem Mass Spectrometry. *Journal of Proteome Research*. 2008; 7:1584–1593. [PubMed: 18324765]
- Dommish H, Acil Y, Dunsche A, Winter J, Jepsen S. Differential gene expression of human b-defensins (hBD-1-2-3) in inflammatory gingival diseases. *Oral Microbiology and Immunology*. 2005; 20:186–190. [PubMed: 15836521]
- Eley BM, Cox SW. Proteolytic and hydrolytic enzymes from putative periodontal pathogens: characterization, molecular genetics, effects on host defenses and tissues and detection in gingival crevice fluid. *Periodontology 2000*. 2003; 31:105–124. [PubMed: 12656998]
- Eng JK, McCormack AL, Yates JR III. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *Journal of the American Society for Mass Spectrometry*. 1994; 5:976–989.
- Frank A, Pevzner P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Analytical Chemistry*. 2005; 77:964–973. [PubMed: 15858974]
- Gorr S-U. Antimicrobial peptides of the oral cavity. *Periodontology 2000*. 2000; 51:152–180. [PubMed: 19878474]
- Gorr S-U, Abdolhosseini M. Antimicrobial peptides and periodontal disease. *Journal of Clinical Periodontology*. 2011; 38:126–141. [PubMed: 21323710]
- Grant MM, Creese AJ, Barr G, Ling MR, Scott AE, Matthews JB, Griffiths HR, Cooper HJ, Chapple ILC. Proteomic Analysis of a Noninvasive Human Model of Acute Inflammation and Its Resolution: The Twenty-one Day Gingivitis Model. *Journal of Proteome Research*. 2010; 9:4732–4744. [PubMed: 20662485]
- Lamster IB, Ahlo JK. Analysis of gingival crevicular fluid as applied to the diagnosis of oral and systemic diseases. *Annals of the New York Academy of Sciences*. 2007; 1098:216–229. [PubMed: 17435131]
- Loo JA, Yan W, Ramachandran P, Wong DT. Comparative Human Salivary and Plasma Proteomes. *Journal of Dental Research*. 2010; 89:1016–1023. [PubMed: 20739693]
- Loos BG, Tjoa S. Host-derived diagnosis markers for periodontitis: do they exist in gingival crevice fluid? *Periodontology 2000*. 2005; 39:53–72. [PubMed: 16135064]
- Lundy FT, Orr DF, Show C, Lamey P-J, Linden GJ. Detection of individual human neutrophil a-defensins (human neutrophil peptides 1, 2 and 3) in unfractionated gingival crevicular fluid: A MALDI-MS approach. *Molecular Immunology*. 2005; 42:575–579. [PubMed: 15607816]
- McLaughlin WS, Kirkham J, Kowolik MJ, Robinson C. Human gingival crevicular fluid keratin at healthy, chronic gingivitis and chronic adult periodontitis sites. *Journal of Clinical Periodontology*. 1996; 23:331–335. [PubMed: 8739164]
- Moll R, Divo M, Langbein L. The human keratins:biology and pathology. *Histochemistry and Cell Biology*. 2008; 129:705–733.
- Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods*. 2007; 4:787–797. [PubMed: 17901868]
- Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics*. 2010; 73:2092–2123. [PubMed: 20816881]
- Ngo LH, Veith PD, Chen Y-Y, Chen D, Cardy IB, Reynolds EC. Mass Spectrometric Analyses of Peptides and Proteins in Human Gingival Crevicular Fluid. *Journal of Proteome Research*. 2010; 9:1683–1693. [PubMed: 20020772]
- Nussbaum G, Shapira L. How has neutrophil research improved our understanding of periodontal pathogenesis? *Journal of Clinical Periodontology*. 2011; 38 Suppl. 11:49–59. [PubMed: 21323704]
- Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching databases using mass spectrometry data. *Electrophoresis*. 1999; 20:3551–3567. [PubMed: 10612281]

- Pisano E, Cabras T, Montaldo C, Piras V, Inzitari R, Olmi C, Castagnola M, Messana I. Peptides of human gingival crevicular fluid determined by HPLC-ESI-MS. *European Journal of Dental Education*. 2005; 113:462–468.
- Puklo M, Guentsch A, Hiemstra PS, Eick S, Potempa J. Analysis of neutrophil-derived antimicrobial peptides in gingival crevicular fluid suggests importance of cathelicidin LL-37 in the innate immune response against periodontogenic bacteria. *Oral Microbiology and Immunology*. 2008; 23:328–335. [PubMed: 18582333]
- Sakellari D, Menti S, Konstantinidis A. Free soluble receptor activator of nuclear factor- κ B ligand in gingival crevicular fluid correlates with distinct pathogens in periodontitis patients. *Journal of Clinical Periodontology*. 2008; 35:938–943. [PubMed: 18988315]
- Sakiyama Y, Kato R, Inoue S, Suzuki K, Itabe H, Yamamoto M. Detection of oxidized low-density lipoproteins in gingival crevicular fluid from dental patients. *Journal of Periodontal Research*. 2010; 45:216–222. [PubMed: 19778326]
- Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL. Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology*. 1998; 25:134–144. [PubMed: 9495612]
- Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V. InsPecT: Identification of Posttranslationally Modified Peptides from Tandem Mass Spectra. *Analytical Chemistry*. 2005; 77:4626–4639. [PubMed: 16013882]
- Uitto VJ, Overall CM, McCulloch C. Proteolytic host cell enzymes in gingival crevice fluid. *Periodontology 2000*. 2003; 31:77–104. [PubMed: 12656997]
- Vitkov L, Klappacher M, Hannig M, Krautgartner D. Extracellular neutrophil traps in periodontitis. *Journal of Periodontal Research*. 2009; 44:664–672. [PubMed: 19453857]
- Witze ES, Old WM, Resing KA, Ahn NG. Mapping protein post-translational modifications with mass spectrometry. *Nature Methods*. 4:798–806. [PubMed: 17901869]

Figure 1.

Gene ontology for the healthy (a) and the diseased (b) human protein samples. All ontological features that has a minimum number of protein hits (biological process: 10 proteins; molecular function: 5 proteins; cellular component: 5 proteins) are reported in the chart.

Table 1

Demographic and clinical characteristics of the subject sample.

<i>Diagnosis</i>	<i>Total</i>	<i>Male</i>	<i>Female</i>	<i>Age range</i>	<i>Mean age ± sd</i>	<i>Sampled sites (Mean ± sd)</i>				
						Probing depth	Recession	Bleeding on probing	Probing depth	Bleeding on probing
Periodontally healthy	12	7	5	42–62	49.75 ± 5.75	1.71 ± 0.24	0.06 ± 0.1	0.08 ± 0.06	1.7 ± 0.49	0.08 ± 0.16
Periodontitis	12	8	4	45–77	56.0 ± 10.9	4.13 ± 0.77	1.47 ± 0.64	0.72 ± 0.16	6.5 ± 0.32	0.8 ± 0.26
										1

No differences were observed between groups concerning mean age (Mann-Whitney test $p > 0.05$). Statistically significant differences in clinical parameters among groups are indicated by bold lettering (Mann-Whitney test $p < 0.05$).

Table 2

Number of human and bacterial proteins found.

	Healthy Only	Chronic Periodontitis Only	Healthy and Chronic Periodontitis	Total
Human	79	123	230	432
Bacterial (Complete)	17	13	-	30
Bacterial (Targeted)	3	15	2	20

Table 3

Human proteins found only in at least 3 healthy samples or only in at least 3 diseased samples.

Accession	Protein	No.	Pep. Count	Spec. Count
<i>Periodontally Healthy</i>				
P01019	Angiotensinogen	9	1.7	4.5
P19971	Thymidine phosphorylase	6	1.9	4.0
P10909	Clusterin	5	2.0	3.0
P14923	Junction plakoglobin	4	1.4	1.9
P22392	Nucleoside diphosphate kinase B	3	1.7	2.7
Q04695	Keratin, type I cytoskeletal 17	5	1.2	2.5
P52566	Rho GDP-dissociation inhibitor 2	4	1.0	2.1
P52565	Rho GDP-dissociation inhibitor 1	3	1.0	2.3
P35749	Myosin-11	3	1.0	2.7
P30101	Protein disulfide-isomerase A3	3	1.0	1.0
Q96FQ6	Protein S100-A16	3	1.0	1.5
<i>Chronic Periodontitis</i>				
P00915	Carbonic anhydrase 1	7	1.9	2.3
P59665	Neutrophil defensin 1	6	1.1	2.0
P26641	Elongation factor 1-gamma	5	1.0	1.4
P22531	Small proline-rich protein 2E	3	1.7	3.0
Q04760	Lactoylglutathione lyase	4	1.0	1.2
Q8TC07	TBC1 domain family member 15	4	1.0	2.0
Q71U36	Tubulin alpha-1A chain	4	1.0	2.8
P63241	Eukaryotic translation initiation factor 5A-1	4	1.0	1.1
P02814	Submaxillary gland androgen-regulated protein 3B	4	1.0	2.8
Q96EN8	Molybdenum cofactor sulfurase	4	1.0	1.7
P09972	Fructose-bisphosphate aldolase C	3	1.0	1.1
P07357	Complement component C8 alpha chain	3	1.0	3.0
Q14CN4	Keratin, type II cytoskeletal 72	3	1.1	2.6
O75367	Core histone macro-H2A.1	3	1.0	1.9
Q9UL46	Proteasome activator complex subunit 2	3	1.0	1.3
Q9NUT2	ATP-binding cassette sub-family B member 8, mitochondrial	4	1.0	1.1
P55854	Small ubiquitin-related modifier 3	3	1.0	1.7
P14678	Small nuclear ribonucleoprotein-associated proteins B and B'	3	1.0	2.9
A0RQ16	Serine hydroxymethyltransferase	3	1.0	2.8
P01814	Ig heavy chain V-II region OU	3	1.0	1.0
Q96DA0	Uncharacterized protein UNQ773/PRO1567	3	1.0	2.8
Q9NPG4	Protocadherin-12	3	1.0	1.6
O83110	Chaperone protein clpB	3	1.0	1.0
Q96TA1	Niban-like protein 1	3	1.0	1.4
Q9UHV7	Mediator of RNA polymerase II transcription subunit 13	3	1.0	1.3
Q86UK0	ATP-binding cassette sub-family A member 12	3	1.0	2.8

Accession	Protein	No.	Pep. Count	Spec. Count
Q8N398	von Willebrand factor A domain-containing protein 5B2	3	1.0	1.6
Q8NDL9	Cytosolic carboxypeptidase-like protein 5	3	1.0	1.4
A0RNI3	Protein translocase subunit secA	3	1.0	2.5

For each protein, the UniProt accession number is given along with the total number (No.) of samples that contained the protein, the average peptide count (Pep. Count), and the average MS/MS count (Spec. Count).

Table 4

Bacterial protein list for the twelve healthy samples and twelve diseased samples using the comprehensive bacterial database.

Accession	UniProt ID	Protein Name	No.	Pep. No.	Spec. No.
<i>Periodontally Healthy</i>					
Q8CXP5	HSLO_OCEIH	33 kDa chaperonin	9	6.8	9.5
Q55835	FUTA2_SYNY3	Iron uptake protein A2	7	5.0	6.0
Q7U4M0	CAPP_SYNPX	Phosphoenolpyruvate carboxylase	7	4.8	14.0
Q6D065	MUTL_ERWCT	DNA mismatch repair protein mutL	6	5.8	5.8
A7Z4X5	MUTS_BACA2	DNA mismatch repair protein mutS	6	6.0	10.2
Q6F054	Y6523_BACAN	Uncharacterized protein pXO2-25/BXB0023/GBAA_pXO2_0023	6	4.7	7.5
Q9CLV1	Y1101_PASMU	Uncharacterized protein PMI101	6	5.8	9.8
A7MWQ3	UPPP_VIBHB	Undecaprenyl-diphosphatase	4	4.0	7.5
P75749	YBGP_ECOLI	Uncharacterized fimbrial chaperone ybgP	4	3.8	5.5
P23325	ARPA_ECOLI	Ankyrin repeat protein A	4	4.5	9.5
P75396	Y386_MYCPN	Uncharacterized protein MG268 homolog	4	4.0	6.0
Q057X7	DNAJ_BUCCC	Chaperone protein dnaJ	4	3.0	7.0
Q0ANP7	EFG_MARMM	Elongation factor G	3	2.3	5.3
Q046A0	RPOA_LACGA	DNA-directed RNA polymerase subunit alpha	3	2.3	6.5
A713U0	RL7_CAMHC	50S ribosomal protein L7/L12	2	1.5	3.0
Q4QP46	NAGB_HAEI8	Glucosamine-6-phosphate deaminase	1	1.0	3.0
A5GDA7	ARLY_GEOUR	Argininosuccinate lyase	1	1.0	2.0
<i>Chronic Periodontitis</i>					
Q2RRP5	RBL2_RHORT	Ribulose biphosphate carboxylase	12	18.4	46.1
P42316	SCOB_BACSU	Probable succinyl-CoA:3-ketoacid-coenzyme A transferase sub	10	7.2	18.8
Q9AIU3	RPOB_ANAPH	DNA-directed RNA polymerase subunit beta	8	8.8	19.3
A0RHL7	RS16_BACAH	30S ribosomal protein S16	7	8.0	10.3
Q59906	G3P_STREQ	Glyceraldehyde-3-phosphate dehydrogenase	7	5.1	9.6
Q1RFM3	BETA_ECOUT	Choline dehydrogenase	7	5.8	15.1
O32755	G3P_LACDE	Glyceraldehyde-3-phosphate dehydrogenase	6	6.0	18.1
Q223B1	UBID_RHOFD	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	6	4.0	4.0
O07401	HEMH_MYCAV	Ferrochelatase	5	4.0	7.0
P29433	RNPA_BUCAP	Ribonuclease P protein component	3	2.0	3.7

Accession	UniProt ID	Protein Name	No.	Pep. No.	Spec. No.
Q21V38	KDSA_RHOFD	2-dehydro-3-deoxyphosphooctonate aldolase	3	2.7	3.0
O83217	EFTU_TREPA	Elongation factor Tu	3	2.7	5.3
A5CQS6	PYRH_CLAM3	Uridylate kinase	2	1.5	3.0

For each protein, the total number (No.) of samples containing that protein is listed along with the average number of assigned peptides (Pep. No.) and the average number of assigned spectra (Spec. No.).

Table 5

Bacterial protein list for the twelve healthy samples and twelve diseased samples using the targeted bacterial database.

Accession	UniProt ID	Protein Name	No.	Pep. No.	Spec. No.
<i>Periodontally healthy & Chronic Periodontitis</i>					
Q8RGH3	ALF1_FUSNN	Fructose-bisphosphate aldolase class 1	2, 1	1.0, 1.0	2.0, 2.0
A6L1J5	PSD_BACV8	Phosphatidylserine decarboxylase proenzyme	3, 1	1.0, 1.0	1.0, 1.0
<i>Periodontally Healthy</i>					
Q8RHQ8	CLPB_FUSNN	Chaperone protein clpB	1	1.0	1.0
Q8RFG9	GPMA_FUSNN	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	1	1.0	2.0
A8FMX4	MNMC_CAMJ8	tRNA 5-methylaminomethyl-2-thiouridine biosynthesis bifunctional protein mmnC	1	1.0	1.0
<i>Chronic Periodontitis</i>					
A0RQ16	CNPD_CAMFF	Serine hydroxymethyltransferase	3	1.0	3.0
O83110	GLYA_CAMFF	Chaperone protein clpB	3	1.0	1.0
A0RNI3	CLPB_TREPA	Protein translocase subunit secA	3	1.0	3.0
Q64MV4	SECA_CAMFF	Phosphoenolpyruvate carboxykinase [ATP]	2	1.0	3.0
Q3K3V7	PPCK_BACFR	50S ribosomal protein L5	2	1.0	2.0
Q8RDR4	RL5_STRA1	GMP synthase [glutamine-hydrolyzing]	2	1.0	2.0
A0RNB3	GUAA_FUSNN	2,3'-cyclic-nucleotide 2'-phosphodiesterase	1	1.0	3.0
Q73P71	CNPD_CAMFF	Phosphonates import ATP-binding protein phnC	1	1.0	2.0
Q7MVG9	PHNC_TREDE	Uncharacterized RNA methyltransferase PG_1095	1	1.0	2.0
Q8AAW1	ARAA_BACTN	L-arabinose isomerase	1	1.0	2.0
P61348	GATB_TREDE	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	1	1.0	1.0
Q73JP8	SELD_TREDE	Selenide, water dikinase	1	1.0	1.0
P66378	RS12_STRP8	30S ribosomal protein S12	1	1.0	1.0
Q64T66	Y2564_BACFR	UPF0082 protein BF2564	1	1.0	2.0
P54355	ENTM_BACFR	Fragilysin	1	1.0	2.0

For each protein, the total number (No.) of samples containing that protein is listed along with the average number of assigned peptides (Pep. No.) and the average number of assigned spectra (Spec. No.). For proteins found in both healthy and diseased samples, the left hand number refers to the healthy samples while the right hand number refers to the diseased samples.

Table 6

Potential number of unmodified biomarker proteins for the periodontally healthy and chronic periodontitis samples.

	Number of Samples	
	5	6
Periodontally healthy (human)	4	2
Chronic periodontitis (human)	3	2
Periodontally healthy (bacteria)	7	7
Chronic periodontitis (bacteria)	9	8

The number of proteins listed in each column are found in at least as many samples as indicated in the column header. The bacterial proteins are derived from the comprehensive database.