Published in final edited form as:

Proteomics Clin Appl. 2010 September ; 4(8-9): 715–725. doi:10.1002/prca.201000018.

Proteomic analysis of endoscopically (ePFT) collected gastroduodenal fluid using in-gel tryptic digestion followed by liquid chromatography-tandem mass spectrometry

Joao A. Paulo, PhD,

Department of Pathology, Children's Hospital Boston, Boston, MA. Proteomics Center at Children's Hospital Boston, Boston, MA. Center for Pancreatic Disease, Division of Gastroenterology, Hepatology and Endoscopy, Brigham and Women's Hospital and Department of Medicine, Harvard Medical School, Boston, MA

Linda S. Lee, MD,

Center for Pancreatic Disease, Division of Gastroenterology, Hepatology and Endoscopy, Brigham and Women's Hospital and Department of Medicine, Harvard Medical School, Boston

Bechien Wu, MD, MPH,

Center for Pancreatic Disease, Division of Gastroenterology, Hepatology and Endoscopy, Brigham and Women's Hospital and Department of Medicine, Harvard Medical School, Boston

Kathryn Repas, BA,

Center for Pancreatic Disease, Division of Gastroenterology, Hepatology and Endoscopy, Brigham and Women's Hospital and Department of Medicine, Harvard Medical School, Boston

Peter A. Banks, MD,

Center for Pancreatic Disease, Division of Gastroenterology, Hepatology and Endoscopy, Brigham and Women's Hospital and Department of Medicine, Harvard Medical School, Boston

Darwin L. Conwell, MD, MS[#], and

Center for Pancreatic Disease, Division of Gastroenterology, Hepatology and Endoscopy, Brigham and Women's Hospital and Department of Medicine, Harvard Medical School, Boston

Hanno Steen, PhD[#]

Departments of Pathology, Children's Hospital Boston and Harvard Medical School, Boston. Proteomics Center at Children's Hospital Boston, Boston

Abstract

Purpose—Proteomic analysis of gastroduodenal fluid offers an alternative strategy to study diseases, such as peptic ulcer disease and gastric cancer. We use in-gel digestion followed by liquid chromatography-tandem mass spectrometry (GeLC-MS/MS) to profile the proteome of gastroduodenal fluid collected during the endoscopic function test (ePFT).

Corresponding authors: Darwin L. Conwell, Center for Pancreatic Disease, Div of Gastroenterology, Hepatology and Endoscopy, 75 Francis St., Boston, MA 02115, dconwell@partners.org, phone: +1-617-732-6747, fax: +1-617-566-0338. Hanno Steen, Ph.D., Children's Hospital Boston, Dept. of Pathology/Enders 1130, 320 Longwood Ave, Boston, MA 02115,

Hanno.Steen@childrens.harvard.edu, phone: +1-617-919-2629, fax +1-617-730-0168.

[#]These authors contributed equally to this work

Conflicts of interests

The authors declare no competing interests.

Author contributions

JP carried out the experiments and drafted the original manuscript. JP, PB, DC and HS conceived of the study, and participated in its design and coordination. DC, LL, and BW collected the specimens and assisted in experimental design. KR coordinated the study. All authors helped to draft the manuscript and approved the final manuscript.

Experimental Design—Gastroduodenal fluid specimens collected during ePFT from six patients with upper abdominal pain were subjected to proteomic analysis. We extracted proteins using three chemical precipitation reagents (acetone, ethanol, and trichloroacetic acid) and analyzed each sample by SDS-PAGE and GeLC-MS/MS for protein identification. Cellular origin and molecular function of the identified proteins were determined via gene ontology analysis.

Results—All three precipitation techniques successfully extracted protein from gastroduodenal fluid, with acetone resulting in excellent resolution and minimal protein degradation compared to the other methods. A total of 134 unique proteins were found in our GeLC-MS/MS analysis of ePFT-collected gastroduodenal fluid samples. Sixty-seven proteins were identified in at least two of the three samples. Gene ontology analysis classified these proteins mainly as being peptidases and localized extracellularly.

Conclusions—ePFT, followed by acetone precipitation, and coupled with LC-MS/MS, can be used to safely collect gastroduodenal fluid from the upper gastrointestinal tract for mass-spectrometry-based proteomic analysis.

Keywords

digestive diseases; biomarkers; gastric cancer; gastrointestinal tract; pancreatic function test

1. Introduction

Mass spectrometry-based proteomic techniques can evaluate the protein profile of complex biological samples to understand better the normal physiology and pathogenic mechanisms of human disease. Hundreds or thousands of proteins may be identified for further investigation using recently developed mass spectrometry methodology which has revolutionized protein identification by enabling the comprehensive analysis of complex protein mixtures [1].

We believe that the application of body fluid proteomics to the study of gastroduodenal disease may result the discovery of physiologically- and clinically-relevant biomarkers. Emerging high-resolution and high-sensitivity mass spectrometry technology will enable the detection of low abundant biomarkers [2–4]. Systemic body fluids, such as urine and blood or its derivatives (i.e., plasma or serum), are commonly interrogated in the search for biomarkers, however such *systemic* fluids represent the entire body and likely include proteins not related to the disease of interest [5]. Conversely, proximal body fluids essentially bathe the diseased organ and best represent the proteins in the near-by tissue. Disease-specific markers that are secreted or shed by these tissues into proximal fluids would be at a much higher concentration than in blood or urine. When proximal fluids are analyzed for disease-specific marker investigation, the dynamic range of protein concentrations is decreased, thereby increasing the sensitivity of detection of the likely low abundant, clinically relevant proteins [6]. Gastroduodenal fluid is an excellent clinical specimen for identification of gastrointestinal disease-specific biomarkers by proteomic analysis, as it is a proximal fluid of relatively low complexity, thus facilitating the identification of low-abundant proteins [7–9].

We aim to investigate the proteome of the gastroduodenal fluid that is present in the stomach/duodenum at the time of upper endoscopy of patients with chronic upper abdominal pain using GeLC-MS/MS (SDS-PAGE followed by in-gel digestion and liquid chromatography-tandem mass spectrometry), a powerful approach for proteomic analysis [10, 11]. As the composition of body fluids may be influenced by environmental factors, age, gender, and confounding diseases, may result in substantial subject-to-subject variability. To decrease the impact of this variability on our subsequent proteomic analyses;

we have used a pooled sample of gastroduodenal fluid from six individuals. Proteomic experiments directed toward the study of gastroduodenal disease present a unique opportunity to accelerate the pace of disease-specific marker discovery.

The primary objectives of our current exploratory investigation are as follows: (1) collect gastroduodenal fluid with the ePFT method, (2) compare the proteins extracted from gastroduodenal fluid among three precipitation methods (acetone, ethanol, and trichloroacetic acid), and (3) determine the molecular function and cellular origin of GeLC-MS/MS identified gastroduodenal fluid proteins via gene ontology (GO) analysis.

2. Materials and Methods

Specimens

The protocol for collecting gastroduodenal fluid was approved by the Institutional Review Board (IRB # 2007-P-002480/1) at Brigham and Women's Hospital (BWH). The gastroduodenal fluid samples used for the experiments were from six subjects with chronic abdominal pain with no evidence of chronic gastroduodenal disease based on histologic evaluation of endoscopic biopsy. These patients had undergone an endoscopic pancreas function test (ePFT) for the assessment of chronic pancreatitis.

The collected specimens were regarded as "excess samples" for which IRB approval had been granted. Such specimens were not specifically collected for this study, but their aspiration was an integral part of the ePFT procedure (as described below and in [12]), and would otherwise have been discarded. Although samples used in our experiments were considered "excess," care was taken to maintain the sample on ice immediately upon collection. Particulates were then removed by centrifugation and the resulting supernatants were collected and frozen at -80°C until processing as described in the *Experimental workflow*. The volume of fluid collected ranged from 0.5–1 mL and protein concentrations are listed in Table 1.

Materials

SeeBluePlus2 Pre-Stained standard (LC5925), LDS (lithium dodecyl sulfate) sample buffer (NP0008), NuPAGE 4–12% Bis-Tris polyacrylamide gels (NP0335), SimplyBlue Coomassie stain (LC0665), and MES-SDS (2-(N-morpholino)ethanesulfonic acid-sodium dodecyl sulfate) running buffer (NP002) were from Invitrogen (Carlsbad, CA). Other reagents and solvents were from Sigma-Aldrich and Burdick and Jackson, respectively.

Experimental workflow

An overview of the procedures used in our investigation is illustrated in Figure 1. Briefly, gastroduodenal fluid was collected via ePFT and immediately placed on ice, particulates were removed by centrifugation ($3,000\times g$ for 15 minutes) at 4°C , and samples were aliquoted on ice and frozen at -80°C for use in the experiments described in detail below. Proteins were precipitated using three methods (see below) from gastroduodenal fluid that was pooled from six patients. SDS-PAGE analysis and gel densitometry was used to determine the relative amount of protein in each sample. Subsequently, GeLC-MS/MS was performed to identify the proteins precipitated in each sample. These identified proteins were compared among the three precipitation techniques. Furthermore, proteins that were present in two or more samples were categorized via gene ontology analysis.

Gastroduodenal fluid collection (ePFT method)

The ePFT procedure is as follows: (A) Pre-procedural assessment, (B) endoscopic procedure and (C) post-procedural assessment/recovery [12].

A. Pre-procedural assessment—Prior to upper endoscopy, all study subjects underwent a history and physical examination including list of allergies, medications, substance use/abuse, and vital signs. Patient pre-intravenous conscious sedation review included airway assessment based on Mallampati airway scale and American Society of Anesthesiologists Physical Status Classification (ASA Class). All study subjects in this protocol had a Mallampati score of B, Class 2 and ASA Class II or better. All patients fasted 12 hours prior to the procedure.

B. Endoscopic procedure—Esophagogastroduodenoscopy (EGD) was performed using a standard (10 mm) upper endoscope for visualization of the esophagus, stomach, and duodenum (2 to 5 minutes). Aspiration of gastroduodenal fluid is performed routinely during an endoscopic pancreatic function test and normally discarded to waste. Here, however, we have collected this fluid immediately prior to the pancreatic function test (ePFT) with the gastroscope. The endoscopic collection was performed in a stepwise manner as follows: The patient was placed in left lateral decubitus position with slight head elevation. The posterior pharynx was sprayed with topical cetacaine spray. A sedation and analgesia bolus was administered. Further sedation doses were given if necessary for patient comfort. After the sedation bolus, a bite-block was placed. Gastroduodenal fluid was aspirated (approximately 1 minute) as completely as possible through the scope into a sterile trap in the suction line for proteomic analysis. A standard secretin stimulated endoscopic pancreatic function test (ePFT) was completed to assess pancreas secretory function [13].

C. Post-procedural assessment/recovery—Study participants were recovered and discharged from the endoscopy unit based on BWH procedural sedation guidelines assessing the level of consciousness, vital signs, oxygen saturation, alertness, gag reflex, degree of nausea, and ability to ambulate.

Gastroduodenal fluid precipitation

We compared three different precipitation reagents and associated methodologies to maximize protein extraction from gastroduodenal fluid. To control for patient-to-patient variability, samples from six individuals (Figure 2) were combined [14, 15]. More specifically, equal volumes of thawed samples from each individual were pooled prior to protein extraction by the three precipitation methods being investigated. To determine the method which maximized protein extraction, gastroduodenal fluid, pooled from six patients, was divided into 200 μ L aliquots to ensure equal protein for each precipitation method (approximately 100 μ g of total protein per aliquot as determined using the BioRad Protein Assay). Protein extraction efficiency was determined based upon SDS-PAGE protein banding patterns and intensities. Gel densitometry was performed using the publically-available ImageJ software [16]. Data points were normalized with respect to the point with the maximum value.

Acetone precipitation—A total of four sample volumes (800 μ L) of ice-cold 100% acetone was added to 200 μ L of gastroduodenal fluid, vortexed briefly, and incubated at -20°C for 3 hours. Subsequently, the samples were centrifuged at $20,000\times g$ at 4°C for 30 minutes. The supernatants were carefully aspirated and the pellets were allowed to air dry at 23°C .

Ethanol precipitation—A total of 0.8 mL of 100% ethanol was added to 200 μ L aliquots of gastroduodenal fluid. The mixtures were vortexed, placed on ice for 30 minutes, and centrifuged ($14,000\times g$ at 4°C for 30 minutes). The supernatants were carefully aspirated and the pellets were allowed to air dry at 23°C .

TCA precipitation—Ice-cold 100% TCA (25 μ L) was added to 200 μ L of gastroduodenal fluid, vortexed and incubated at 4°C for 2 hours. The sample was centrifuged at 20,000 $\times$ g at 4°C for 30 minutes and the supernatant was carefully aspirated. One milliliter of 100% ice-cold acetone was added to the pellets which were briefly vortexed and incubated at -20°C for 1 hour. The sample was centrifuged at 20,000 $\times$ g at 4°C for 30 minutes and the pellet was washed twice with 100% ice-cold acetone. The final pellets were allowed to air dry at 23°C.

SDS-PAGE analysis of gastroduodenal fluid

In preparation for SDS-PAGE, the proteins from gastroduodenal fluid specimens were isolated by precipitation, as described above. This process was meant to deactivate enzymes and remove salts that interfere with the subsequent fractionation by SDS-PAGE. Pellets were re-dissolved in 20 μ L of 1 $\times$ LDS sample buffer. Pellets were first disrupted by titration, followed by brief sonication to eliminate protein aggregates. For all three experiments, 2 μ L of 0.1 M DTT (dithiothreitol) were added to each sample, which was then incubated at 56°C for 1 hour in a thermomixer (Eppendorf) at 1000rpm. After cooling, samples were alkylated with 2 μ L of 40% acrylamide for 30 minutes at 23°C. The entire sample is then loaded onto the SDS-PAGE gel. SDS-PAGE protein separation was performed at 150 volts in MES buffer for 45 minutes. Gels were rinsed in deionized water for 10 minutes, fixed in 45% methanol/45% water/10% acetic acid for 30 minutes, stained with SimplyBlue Coomassie for 1 hour, and destained overnight in deionized water.

GeLC-MS/MS analysis of gastroduodenal fluid

Three gels lanes of gastroduodenal fluid proteins, which have been extracted by each of the three precipitation methods described above, were subjected to in-gel tryptic digestion followed by reversed-phase liquid chromatography inline with a tandem mass spectrometer (GeLC-MS/MS). In brief, entire gel lanes were divided into 11 sections and proteins in each gel section were digested with trypsin [17, 18]. Peptides extracted from each gel section were fractionated and analyzed by nanoflow reversed-phase UPLC (Eksigent, Dublin, CA) in-line with an LTQ-Fourier transform ion cyclotron mass spectrometer (LTQ-FT Ultra, Thermo Scientific, Waltham, MA). The LC columns (15 cm x 100 μ m ID) were packed in-house (Magic C18, 5 μ m, 100 Å, Michrom BioResources, Auburn, CA). Samples are analyzed with a 60-minute linear gradient (0–35% acetonitrile with 0.2% formic acid) and data are acquired in a data dependent fashion, with 6 MS/MS scans for every full scan spectrum.

Bioinformatics and data analysis

All data generated from the gel sections were searched against the IPI-human database (v3.61) using the Paragon Algorithm [19] integrated into the ProteinPilot search engine (v.3; AB Sciex, Foster City, CA). Search parameters were set as follows: sample type, identification; Cys alkylation, acrylamide; Instrument, Orbitrap/FT (1–3 ppm); special factors, gel-based ID; ID focus, none; database, international protein index (IPI) human (v. 3.61); detection protein threshold, 99.0%; and search effort, thorough ID. Thus, using our stringent criteria, we defined an identified protein with $\geq$ 99% confidence, as determined by the Paragon Algorithm. Gene ontology analysis [20] was performed manually with the UniProt [21] database or using the GoFact online tool [22, 23].

3. Results

Acetone precipitation was superior based on SDS-PAGE analysis

A similar overall protein banding pattern profile for precipitated ePFT-collected gastroduodenal fluid was visualized regardless of the chemical used for the precipitation

(Figure 3A). All three precipitation reagents (acetone, ethanol, and TCA) produced resolved bands of relatively equal intensity. However, slightly less precipitated protein was evident in the ethanol precipitation lane (lane 2). In addition, a relatively large blue smear appeared near the bottom of the TCA extracted protein lane (lane 3). Thus protein degradation is evident when using this acidifying precipitation method. This increased degradation in the presence of TCA can be rationalized by the significant abundance of pepsin which is particularly active below pH 2. Acetone precipitation showed excellent protein extraction with a distinct banding pattern and less degradation, as can be seen in Figure 3A (lane 1).

To acquire quantitative data, we performed gel densitometry analysis using ImageJ [16]. To assess relative protein amounts, we only analyzed the portion of the gel lane migrating above the 7 kDa protein standard band. We reasoned that as the starting sample is equivalent, proteins less than 7 kDa are more likely to be breakdown products and any estimate of protein yield will be confounded by the large blue smear in the TCA lane that migrates below the 7 kDa protein standard band. In a recently published work where we discuss protein degradation in secretin-stimulated pancreatic fluid [24], similar low molecular weight smears were present, which were indicative of protein degradation. Gel densitometry analysis of the proteins bands appearing above 7kDa agrees with visually estimated protein yields as being greater with precipitation in acetone, followed by ethanol and TCA which both yielded approximately 25% less protein (Figure 3B).

Gastroduodenal fluid proteins were readily identified using GeLC-MS/MS analysis

A total of 134 distinct, non-redundant proteins were found in our GeLC-MS/MS analysis of ePFT-collected gastroduodenal fluid when combining proteins identified from each precipitation method. Using our stringent identification criteria, 87, 71, and 78 proteins were identified in samples precipitated with acetone, ethanol and TCA, respectively. As shown in Figure 4, of these proteins, 67 distinct proteins were identified in at least two of the three sample methods, and a total of 35 distinct proteins were found in all three samples (Table 2). Supplemental Tables 1-3 list additional proteins that were identified by GeLC-MS/MS but only appeared in a single sample (i.e. acetone, ethanol, and TCA precipitation, respectively), and in neither of the other two methods. As expected, in this admixture, gastric, duodenal, pancreatic, and biliary fluid proteins were identified. Consistently high scoring proteins ($\geq 99\%$ confidence) that were identified by all three precipitation methods included gastric proteins, such as mucin-5, gastricsin, pepsinogen, $\alpha 1$ -anti-chymotrypsin (SERPINA3), and $\alpha 1$ -antitrypsin (SERPINA1) and the duodenal protein, enteroprotease (enterokinase).

Proteins identified in gastroduodenal fluid exhibit the expected subcellular localization and molecular function

Gene ontology (GO) analysis was performed manually with the UniProt [21] database and using GoFact server [22, 23] for the 67 proteins that were present in at least 2 of the 3 samples, which were extracted by different precipitation reagents. Many of the identified proteins were classified as being of extracellular origin (Figure 5A). Similarly, when examining the molecular functions of these proteins (Figure 5B), most of the proteins were involved in protein binding or had peptidase activity.

4. Discussion

We have shown that mass spectrometric analysis of ePFT-collected fluid is a suitable option to investigate the gastroduodenal proteome. We have successfully (1) collected gastroduodenal fluid during a routine upper endoscopy, (2) discovered acetone to be the superior protein extraction reagent among the precipitation methods used, and (3) determined via gene ontology analysis the primary molecular function and subcellular origin

of GeLC-MS/MS identified gastroduodenal fluid proteins to be enzymes and extracellular, respectively. To our knowledge, this is the largest gastroduodenal fluid proteome described to date. Elucidation of differences in the proteins that are present in the gastroduodenal fluid of healthy and diseased states with a mass spectrometry-based proteomic approach as we outline herein, may provide a better understanding of the molecular mechanisms leading to the onset and progression of the disease.

Recent reviews have summarized studies investigating biomarkers associated with gastric cancer and gastric ulcers [25–28]. Of note, a study investigating gastric cancer identified 60 protein peaks that were up-regulated and 46 that were down-regulated in gastric fluid using surface-enhanced laser desorption/ionization time of flight (SELDI-TOF), however their approach did not provide any protein identifications [29]. While several studies identified disease-specific protein markers in tissue [30] and serum [31], to date, there have been only two liquid chromatography tandem mass spectrometry-based LC-MS/MS proteomic studies directly examining gastric fluid [32, 33]. These studies identified 10 and 32 unique proteins using two-dimensional gel electrophoresis and either MALDI or nanoelectrospray ionization mass spectrometry for the identification of individual protein spots. Both studies determined that the up-regulation of α 1-antitrypsin in gastric fluid was a potential biomarker of gastric cancer. Similarly, we have also identified α 1-antitrypsin using all of our precipitation methods. In total, our analysis has revealed 134 total proteins, which currently represents the largest proteomics data set for gastroduodenal fluid. We envision that our methodology and the insights from our work can be expanded further using large cohorts, which may result in the determination of statistically significant differences in the gastroduodenal fluid proteomes of diseased and non-diseased patients. Such data, potentially coupled with innovative quantitative proteomics techniques, can be used to generate a panel of several protein biomarkers [34–38] that can enable the rapid, practical, sensitive, and accurate diagnosis of gastroduodenal disease.

One of the major limitations of body fluid proteomics is susceptibility to protein degradation during sample processing. In efforts to maximize protein integrity, we have ensured that our ePFT-collected sample is immediately placed on ice in chilled tubes following collection, promptly centrifuged at 4°C to remove particulates, and immediately frozen at –80°C for future analysis. Prior to desalting and protease inactivation via the precipitation methods that we investigated, frozen samples are thawed on ice. Even with these precautions, protein degradation is evidenced by the low molecular weight blue smear near the bottom of the TCA precipitated gastroduodenal fluid lane of the Coomassie-stained gel in Figure 3A (lane 3). Future experiments may exploit peptidomics-based techniques to examine further the degradation of gastroduodenal fluid proteins at the peptide level [39, 40]. In addition, peptide labeling isobaric tag labeling tags (e.g., TMT [41] or iTRAQ [42]) can be performed in a multiplexed manner to compare the relative amounts of proteins extracted by the different precipitation methods.

Under acidifying precipitation conditions, protein degradation of gastroduodenal fluid is expected. As a result of acid secretion from parietal cells, gastric fluid has a very acidic pH (pH 1–2) [43]. Therefore, it is expected that the acidification of gastroduodenal fluid by TCA was unable to rapidly and completely denature gastroduodenal proteases, and thus prevent protein degradation. In fact, the conversion of inactive pepsinogen to the protease pepsin is typically activated by the acidic pH, thus, acidic TCA-based protein precipitation is counterproductive in the case of gastroduodenal fluid, as it may (re-)activate gastric proteases. As gastroduodenal fluid protein degradation may occur during the incubation period of the precipitation process, the addition of protease inhibitors and/or neutralization of pH with base early into the procedure may be advantageous. This addition may be particularly useful if an acidifying precipitation reagent (e.g., TCA, formic acid,

trifluoroacetic acid) is used. However, we have shown that little protein degradation is evident when using acetone precipitation, which may eliminate the need for the additional step of protease inhibitor supplementation.

We acknowledge that our gastroduodenal fluid sample is an admixture of several upper gastrointestinal fluids, many with similar gene ontologies as in Figure 5. The presence of pancreatic and biliary proteins in gastroduodenal fluid is expected due to basal secretions emanating from the papilla. However, due to the location of collection, the restriction of food intake by the patient for 12 hours, and lack of pancreatic (secretin) and biliary (cholecystokinin, CCK) stimulation, the majority of the proteins are not pancreaticobiliary in origin. In fact, a similar proteomic analysis of ePFT-collected secretin-stimulated pancreatic fluid [44] has revealed that the proteins in pancreatic fluid differ substantially from that of gastroduodenal fluid both by SDS-PAGE protein banding pattern profile and protein identity as determined by mass spectrometry. Also, unlike as noticed with gastroduodenal fluid, there is minimal degradation of secretin-stimulated, ePFT-collected pancreatic fluid upon incubation with TCA [24], which provides evidence that gastroduodenal fluid contamination is not prevalent in pancreatic fluid collected with our ePFT-based approach.

We have recently investigated and optimized the extraction of proteins from secretin-stimulated, ePFT-collected pancreatic fluid for GeLC-MS/MS analysis [24]. In contrast to what we discovered for gastroduodenal fluid, the highest amount of pancreatic fluid protein was extracted using TCA, both in the number of sharp, distinct bands and band intensity. In addition, lower molecular weight band smears were minimized when TCA is used compared to the other protein extraction techniques. In the case of pancreatic fluid (generally pH 8–8.5), the decrease in pH resulting from TCA addition successfully precipitates proteins and inhibits the activity of pancreatic proteases. This is in contrast to gastroduodenal fluid, in which gastric enzymes, such as gastrin and pepsin, are active at very acidic pH. Consequently, as mentioned above, the addition of TCA to gastroduodenal fluid promotes proteolysis, as was demonstrated in Figure 3. Although both are body fluids of the upper gastrointestinal tract and can be collected using the ePFT method, our data emphasize that pancreatic and gastroduodenal fluids require different sample preparation techniques for GeLC-MS/MS analysis.

In summary, we have defined a methodology for the proteomic analysis of endoscopically collected gastroduodenal fluid, which is amenable to more extensive studies. The data obtained can promote a better understanding of the proteins regulating the pathophysiology of gastric and duodenal mucosal disease at the macromolecular level. In addition, similar analyses can enable the comparison of the gastroduodenal fluid proteomes of healthy individuals and patients with gastroduodenal disease for the discovery of clinically-relevant, disease-specific markers. It is important for future studies to elucidate the gastroduodenal proteome of healthy individuals, which will serve as a baseline for the study of diseased conditions. Such analyses are possible with safe collection using the ePFT collection method [12]. Mass spectrometry-based proteomics of gastroduodenal fluid collected during ePFT can enhance our knowledge of the onset of gastroduodenal disease by generating distinctive protein fingerprints for various disease states. In essence, the ePFT technique coupled with GeLC-MS/MS provides a valuable foundation upon which future proteomic investigations of gastroduodenal fluid can be developed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funds were provided by the Harvard Digestive Diseases Center (NIH 5 P30 DK034854-24) and the NIH/NIDDK NRSA Fellowship (NIH NIDDK 1 F32 DK085835-01A1). In addition, we would like to thank the Burrill family for their generous support through the Burrill Research Grant. We would like to thank members of the Steen Lab at Children's Hospital Boston and Harvard Medical School, in particular Saima Ahmed, Robert Everley, and John FK Sauld for their technical assistance and critical reading of the manuscript.

Abbreviations

ePFT	endoscopic pancreatic function test
GeLC-MS/MS	in-gel tryptic digestion followed by liquid chromatography-tandem mass spectrometry
LTQ-FTICR	linear trap quadrupole- Fourier transform ion cyclotron resonance mass spectrometry
TCA	trichloroacetic acid

References

1. Thongboonkerd, V. Proteomics of human body fluids : principles, methods, and applications. Humana Press; Totowa, N.J.: 2007.
2. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng. 2009; 11:49–79. [PubMed: 19400705]
3. Pan S, Aebersold R, Chen R, Rush J, et al. Mass spectrometry based targeted protein quantification: methods and applications. J Proteome Res. 2009; 8:787–797. [PubMed: 19105742]
4. Second TP, Blethrow JD, Schwartz JC, Merrihew GE, et al. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. Anal Chem. 2009; 81:7757–7765. [PubMed: 19689114]
5. Issaq HJ, Xiao Z, Veenstra TD. Serum and plasma proteomics. Chemical reviews. 2007; 107:3601–3620. [PubMed: 17636887]
6. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol. 2006; 24:971–983. [PubMed: 16900146]
7. Anderson NL, Polanski M, Pieper R, Gatlin T, et al. The human plasma proteome: a nonredundant list developed by combination of four separate sources. Mol Cell Proteomics. 2004; 3:311–326. [PubMed: 14718574]
8. Gronborg M, Bunkenborg J, Kristiansen TZ, Jensen ON, et al. Comprehensive proteomic analysis of human pancreatic juice. Journal of proteome research. 2004; 3:1042–1055. [PubMed: 15473694]
9. Muthusamy B, Hanumanthu G, Suresh S, Rekha B, et al. Plasma Proteome Database as a resource for proteomics research. Proteomics. 2005; 5:3531–3536. [PubMed: 16041672]
10. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc. 2006; 1:2856–2860. [PubMed: 17406544]
11. Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Analytical chemistry. 1996; 68:850–858. [PubMed: 8779443]
12. Wu B, Conwell DL. The endoscopic pancreatic function test. The American journal of gastroenterology. 2009; 104:2381–2383. [PubMed: 19806083]
13. Conwell DL, Zuccaro G Jr, Vargo JJ, Trolli PA, et al. An endoscopic pancreatic function test with synthetic porcine secretin for the evaluation of chronic abdominal pain and suspected chronic pancreatitis. Gastrointestinal endoscopy. 2003; 57:37–40. [PubMed: 12518128]
14. Diz AP, Truebano M, Skibinski DO. The consequences of sample pooling in proteomics: an empirical study. Electrophoresis. 2009; 30:2967–2975. [PubMed: 19676090]
15. Karp NA, Lilley KS. Investigating sample pooling strategies for DIGE experiments to address biological variability. Proteomics. 2009; 9:388–397. [PubMed: 19105178]

16. Girish V, Vijayalakshmi A. Affordable image analysis using NIH Image/ImageJ. *Indian J Cancer*. 2004; 41:47. [PubMed: 15105580]
17. Neubauer G, Mann M. Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations. *Anal Chem*. 1999; 71:235–242. [PubMed: 9921130]
18. Steen H, Kuster B, Fernandez M, Pandey A, Mann M. Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal Chem*. 2001; 73:1440–1448. [PubMed: 11321292]
19. Shilov IV, Seymour SL, Patel AA, Loboda A, et al. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics*. 2007; 6:1638–1655. [PubMed: 17533153]
20. Ashburner M, Ball CA, Blake JA, Botstein D, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25:25–29. [PubMed: 10802651]
21. Bairoch A, Apweiler R, Wu CH, Barker WC, et al. The Universal Protein Resource (UniProt). *Nucl Acids Res*. 2005; 33:D154–159. [PubMed: 15608167]
22. Li D, Li JQ, Ouyang SG, Wu SF, et al. An integrated strategy for functional analysis in large-scale proteomic research by gene ontology. *Progress in Biochemistry and Biophysics*. 2005; 32:1026–1029.
23. Dong L, Jianqi L, Shuguang O, Songfeng W, et al. An integrated strategy for functional analysis in large scale proteomic research by gene ontology. *Molecular & Cellular Proteomics*. 2005; 4:S34–S34.
24. Paulo JA, Lee LS, Wu B, Repas K, et al. Optimized Sample Preparation of Endoscopic (ePFT) Collected Pancreatic Fluid for SDS-PAGE Analysis. *Electrophoresis*. 2010 In Press.
25. Li JJ, Qi RZ, Ng GKH, Xie D. Proteomics in gastric cancer research: Benefits and challenges. *Proteomics Clinical Applications*. 2009; 3:185–196.
26. Lin Y, Dynan WS, Lee JR, Zhu ZH, Schade RR. The Current State of Proteomics in GI Oncology. *Digestive Diseases and Sciences*. 2009; 54:431–457. [PubMed: 19104933]
27. Ryu JW, Kim HJ, Lee YS, Myong NH, et al. The proteomics approach to find biomarkers in gastric cancer. *J Korean Med Sci*. 2003; 18:505–509. [PubMed: 12923326]
28. Tan S, Liang C, Yeoh KG, So J, et al. Gastrointestinal fluids proteomics. *Proteomics Clinical Applications*. 2007; 1:820–833. [PubMed: 21136736]
29. Kon OL, Yip TT, Ho MF, Chan WH, et al. The distinctive gastric fluid proteome in gastric cancer reveals a multi-biomarker diagnostic profile. *BMC Med Genomics*. 2008; 1:54. [PubMed: 18950519]
30. Jia SQ, Niu ZJ, Zhang LH, Zhong XY, et al. Identification of prognosis-related proteins in advanced gastric cancer by mass spectrometry-based comparative proteomics. *J Cancer Res Clin Oncol*. 2009; 135:403–411. [PubMed: 18830628]
31. Lam KWK, Lo SCL. Discovery of diagnostic serum biomarkers of gastric cancer using proteomics. *Proteomics Clinical Applications*. 2008; 2:219–228. [PubMed: 21136826]
32. Lee K, Kye M, Jang JS, Lee OJ, et al. Proteomic analysis revealed a strong association of a high level of alpha1-antitrypsin in gastric juice with gastric cancer. *Proteomics*. 2004; 4:3343–3352. [PubMed: 15378737]
33. Hsu PI, Chen CH, Hsieh CS, Chang WC, et al. Alpha1-antitrypsin precursor in gastric juice is a novel biomarker for gastric cancer and ulcer. *Clin Cancer Res*. 2007; 13:876–883. [PubMed: 17289880]
34. Kolberg JA, Jorgensen T, Gerwien RW, Hamren S, et al. Development of a type 2 diabetes risk model from a panel of serum biomarkers from the Inter99 cohort. *Diabetes Care*. 2009; 32:1207–1212. [PubMed: 19564473]
35. Yilmaz Y, Ulukaya E, Dolar E. The quest for liver fibrosis biomarkers: promises from the enhanced liver fibrosis panel and beyond. *Hepatology*. 2009; 49:1056–1057. [PubMed: 19133647]
36. Timotheadou E, Skarlos DV, Samantas E, Papadopoulos S, et al. Evaluation of the prognostic role of a panel of biomarkers in stage IB-IIIa non-small cell lung cancer patients. *Anticancer Res*. 2007; 27:4481–4489. [PubMed: 18214064]

37. Simonsen AH, McGuire J, Podust VN, Hagnelius NO, et al. A novel panel of cerebrospinal fluid biomarkers for the differential diagnosis of Alzheimer's disease versus normal aging and frontotemporal dementia. *Dement Geriatr Cogn Disord*. 2007; 24:434–440. [PubMed: 17971664]
38. Kentsis A, Lin YY, Kurek K, Calicchio M, et al. Discovery and Validation of Urine Markers of Acute Pediatric Appendicitis Using High-Accuracy Mass Spectrometry. *Ann Emerg Med*. 2009
39. Schulz-Knappe P, Schrader M, Zucht HD. The peptidomics concept. *Comb Chem High Throughput Screen*. 2005; 8:697–704. [PubMed: 16464157]
40. Ivanov VT, Yatskin ON. Peptidomics: a logical sequel to proteomics. *Expert Rev Proteomics*. 2005; 2:463–473. [PubMed: 16097881]
41. Thompson A, Schafer J, Kuhn K, Kienle S, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry*. 2003; 75:1895–1904. [PubMed: 12713048]
42. Ross PL, Huang YN, Marchese JN, Williamson B, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*. 2004; 3:1154–1169. [PubMed: 15385600]
43. Kajita H, Morishima S, Shirakata Y, Kotera T, et al. A mini Cl⁻ channel sensitive to external pH in the basolateral membrane of guinea-pig parietal cells. *J Physiol*. 1995; 488(Pt 1):57–64. [PubMed: 8568665]
44. Paulo JA, Lee LS, Wu B, Repas K, et al. Identification of Pancreas-Specific Proteins in Endoscopically (Endoscopic Pancreatic Function Test) Collected Pancreatic Fluid with Liquid Chromatography- Tandem Mass Spectrometry. *Pancreas*. 2010

Statement of clinical relevance

Gastroduodenal diseases have a significant global impact on healthcare. Improved strategies for the diagnosis and treatment of such upper digestive tract diseases are necessary to reduce healthcare and patient burdens. Identifying perturbations in the protein profiles of healthy and diseased patients will provide insight into disease pathogenesis at the molecular level. Mass spectrometry-based proteomic analysis of gastroduodenal fluid offers a novel strategy to study upper digestive tract diseases including peptic ulcer disease and gastric cancer. We show that the ePFT (endoscopic pancreas function test) collection method can safely collect gastroduodenal fluid for mass spectrometry-based proteomic analysis. Application of this robust and stringent analysis allowed us to generate a list of gastroduodenal fluid proteins, which to our knowledge is the largest gastroduodenal fluid proteome described to date. Using three different precipitation methods, we eliminated bias at the protein extraction level, enabling the elucidation of a core set of gastroduodenal proteins, which can be further investigated with orthogonal methodologies. The ePFT technique in tandem with our mass spectrometry strategy provides a valuable foundation upon which future proteomic investigations of gastroduodenal fluid can be developed further. We envision future applications for comparative analyses searching for biomarkers of gastroduodenal disease.

Figure 1.

Experimental workflow. Gastroduodenal fluid is collected from six individuals, particulates are pelleted. Samples from the six individuals are pooled and subsequently divided into 3 aliquots to be precipitated by three different chemical (acetone, ethanol, and TCA) precipitation techniques. Samples are analyzed by GeLC-MS/MS methodology.

Figure 2.

SDS-PAGE analysis on gastroduodenal fluid of each of the six individuals. Patient-to-patient variability is observed among the six samples. Subsequent proteomic analyses use a pooled sample of gastroduodenal fluid from these six individuals.

Figure 3. Protein extraction from gastroduodenal fluid. A) SDS-PAGE analysis of gastroduodenal fluid precipitated by (1) acetone, (2) ethanol, and (3) trichloroacetic acid (TCA). B) Quantitative gel densitometry measurements of gel lanes (98kDa to 7kDa molecular weight marker) as determined using ImageJ software. Data points were normalized with respect to the maximum intensity value.

Figure 4. Venn diagram depicting the number of proteins identified by GeLC-MS/MS of gel lanes in gastroduodenal fluid precipitated by acetone, ethanol, and TCA. Tallied proteins were identified with two or more peptides that were of 95% or greater confidence as determined by the Paragon Algorithm [19].

Figure 5. Gene ontology analysis of A) subcellular localization and B) molecular function. Gene ontology characterization using the categories listed of the proteins identified in 2 of the 3 samples was performed manually with the UniProt [21] database or using the GoFact online tool [22, 23].

Table 1

Demographic and clinical data on chronic abdominal pain patients.

ID	Reason for referral	Age (yr)	Gender	CT Scan	MRI	Pancreatic elastase-1 (units/mg)	EUS CP Score (0-9)	ePFT Peak Bicarbonate (meq/L)	Endoscopic Biopsy
CAP1	abdominal pain	28	male	normal	normal	n/a	3	84	negative
CAP2	abdominal pain	53	female	normal	normal	148 (diarrhea)	1	92	negative
CAP3	abdominal pain	54	male	normal	normal	202	n/a	114	negative
CAP4	abdominal pain	51	male	n/a	normal	70 (diarrhea)	4	84	negative
CAP5	abdominal pain	58	male	normal	n/a	n/a	2	101	negative
CAP6	abdominal pain	23	female	normal	normal	n/a	4	81	negative

Table 2
Proteins identified by mass spectrometry in at least 2 of the 3 precipitated samples

Bold protein names and IPI accession numbers were identified from all three precipitated samples. Also listed in the table are: % Cov, percent coverage (number of unique amino acids identified divided by the total number of amino acids in the sample); Peptides (95%), number of peptides that were identified with 95% confidence for that particular protein.

Name	Accession #	Acetone precipitation			Ethanol precipitation			TCA precipitation		
		% Cov	Peptides (95%)	% Cov	Peptides (95%)	% Cov	Peptides (95%)	% Cov	Peptides (95%)	Peptides (95%)
A2M Alpha-2-macroglobulin	IPI00478003.1	56.5	7	52.9	5	49.5	5	49.5	5	5
Anti-(ED-B) scFV	IPI00929201.1			62.6	3	73.1	3	73.1	2	2
C20orf70, epithelium carcinoma-associated protein 2	IPI00304557.2	57.4	5	55.4	2					
cDNA FLJ60007, highly similar to Pepsin A	IPI00908544.1			66.7	29	64.3	23			23
cDNA FLJ60927, highly similar to Mucin-1	IPI00909818.1	43.1	3	38.7	3	27.8	6			6
CEACAM5 Protein	IPI00880101.1	33.5	2	23.5	2					
CEACAM8	IPI00013972.1			37.5	2	31.2	3			3
CELA2A Putative uncharacterized protein	IPI00829925.1			57.5	6	49	4			4
CELA3A Elastase-3A	IPI00295663.1	75.6	7			58.9	7			7
CELA3B 33 kDa protein	IPI00871512.1	63.2	8	47.7	4	70.3	9			9
CST5 Cystatin-D	IPI00002851.1	46.5	2	60.6	3					
CTRB2 chymotrypsin B2	IPI00515087.2	97	9	68.4	3					
DEFA3 Neutrophil defensin 3	IPI00021827.3	67	2	60.6	2	69.2	6			6
DMBT1 Putative uncharacterized protein DMBT1	IPI00871790.1			63	28	55.9	18			18
FCGBP IgGfC-binding protein	IPI00242956.5	43.6	9	41.7	6	36.5	4			4
FNI Isoform 14 of Fibronectin	IPI00022434.4	87.1	27	54.7	17	64.2	18			18
GIF Gastric intrinsic factor	IPI000018169.4	41	7	60.4	2	42.5	4			4
GP2 Alpha of Pancreatic secretory granule	IPI00914943.1	50.4	9			54.3	4			4
IGHA1, highly similar to Ig alpha-1 chain C region	IPI00449920.1			53.6	18	51.7	20			20
IGHA1 Putative uncharacterized protein	IPI00423462.5	65.9	24			51.3	17			17
IGHG3 Putative uncharacterized protein	IPI00930684.1	63.6	3	52.8	3	55.3	3			3
IGKV3D-15 Myosin-reactive immunoglobulin light chain	IPI00549330.4	62.1	3			100	3			3
IGKV4-1 Similar to Ig kappa chain V-IV region precursor	IPI00026197.7			33.3	2	40	4			4
LCN2 Neutrophil gelatinase-associated lipocalin	IPI00299547.4	90.4	29			93.9	29			29
LOC100126583:IGHA2 Putative protein	IPI00784950.1	60.1	11	58.5	7	56.6	7			7

Name	Accession #	Acetone precipitation			Ethanol precipitation			TCA precipitation		
		% Cov	Peptides (95%)	% Cov	% Cov	Peptides (95%)	% Cov	% Cov	Peptides (95%)	Peptides (95%)
LOC100132635 similar to Mucin-5AC precursor mucin	IP100886878.1	66.9	43	65.1	49	51.7	44			
LOC100291464 hypothetical protein XP_002346449	IP100829804.1	58.5	5			72.9	4			
LYZ Lysozyme C	IP100019038.1	74.3	5	95.3	11	91.2	13			
MUC13 Mucin-13	IP100011448.1	55.5	2	41.8	2					
MUC5AC Gastric mucin	IP100816128.2			64.7	30	48.7	28			
MUC5AC Mucin 5AC, oligomeric mucus/gel-forming	IP100918002.1	48.6	54	56.3	48	45.3	35			
MUC5AC Mucin-5AC	IP100103397.2	59.6	63	57.1	71	52.9	52			
MUC6 mucin 6, gastric	IP100401776.9	38.6	7	32.4	5	24.3	3			
NPC2, highly similar to Epididymal secretory protein	IP100301579.4	75.1	2	33.3	2	46.8	2			
OLF4 Olfactomedin-4	IP100022255.1	57.5	2			41.8	3			
PGA5;PGA4;PGA3 pepsinogen 3, group I	IP100736755.2	61.9	48			46.7	23			
PGA5;PGA4;PGA3 Pepsinogen 5, group I	IP100181304.5	61.6	42	44.6	24	46.7	22			
PGC Gastricsin	IP100022213.1	44.6	29	35.3	29	37.4	21			
PLA2G1B Phospholipase A2	IP100021792.1	85.1	4	75	4	75	8			
PNLIP Pancreatic triacylglycerol lipase	IP100027720.1	63.9	4	54.8	4	54.4	3			
PRB2 Basic salivary proline-rich protein 2	IP100552432.3	87.5	2	71.9	4	47.6	2			
PRB3 PRB3 protein	IP100883885.1	31.5	3	37	4					
Proline-rich protein HaeIII subfamily 1	IP100847261.1	62.3	19	49.2	13					
PRSS1 PRSS1 protein	IP100815665.1	46.6	15	55.9	12	42.5	11			
PRSS1 Trypsin-1	IP100011694.1	74.5	25	61.9	23	66.8	19			
PRSS3 Isoform A of Trypsin-3	IP100015614.4	73.7	16	96.7	10	56.9	7			
PRSS7 Enteropeptidase	IP100023788.1	48.6	9	49.3	11	39.2	5			
Putative uncharacterized protein	IP100550731.2	93.3	2			55.2	3			
REG1A Lithostathine-1-alpha	IP100009027.1			57.2	4	38	4			
Rheumatoid factor C6 light chain	IP100829956.1	69	5	65.5	2	64.7	4			
Rheumatoid factor D5 light chain	IP100816799.1			54.2	2	64.4	3			
SERPINA1 Isoform 1 of Alpha-1-antitrypsin	IP100553177.1	81.1	23	85.9	16	82.5	18			
SERPINA3, similar to Alpha-1-antichymotrypsin	IP100550991.3	73.2	9	72.8	11					
SERPINE1 Leukocyte elastase inhibitor	IP100027444.1	70.7	8	71.8	10	65.4	10			
SERPINE3 Isoform 1 of Serpin B3	IP100022204.2	82.3	2	76.2	2	75.1	2			

Name	Accession #	Acetone precipitation		Ethanol precipitation		TCA precipitation	
		% Cov	Peptides (95%)	% Cov	Peptides (95%)	% Cov	Peptides (95%)
SERPINB6 Putative uncharacterized protein	IP100413451.1	85.6	2	86.1	2	84.4	3
SERPINC1 Antithrombin-III	IP100032179.3	69.2	2			68.1	4
SERPING1, highly similar to Plasma protease inhibitor	IP100879931.1			68.9	5	62.1	5
SLP1 Antileukoprotease	IP100008580.1	95.5	6	78	5	81.8	6
SPRR1A Putative uncharacterized protein	IP100914840.2	66.3	4			60.7	2
SPRR1B Cornifin-B	IP100304903.4	86.5	3			58.4	3
TFF2 Trefoil factor 2	IP100010675.1	73.6	12	64.3	9	56.6	8
TGM1 Protein-glutamine gamma-glutamyltransferase K	IP100305622.4	56.2	5	63.3	3	59.5	4
TGM3 Protein-glutamine gamma-glutamyltransferase E	IP100300376.4	56.4	3	68.4	2	59.7	9
TMPRSS11D Transmembrane protease, serine 11D	IP100003542.1	62.7	3	50.7	3	57.7	3
TXN Thioredoxin	IP100216298.6	89.5	2			64.8	7
ZG16B Zymogen granule protein 16 homolog B	IP100060800.5	79.8	16	85.1	13	54.8	14