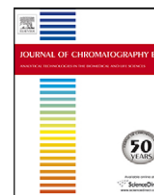
Journal of Chromatography B, 877 (2009) 3127–3136

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb



# A label-free nano-liquid chromatography–mass spectrometry approach for quantitative serum peptidomics in Crohn's disease patients

Paolo Nanni [a], Fredrik Levander [b], Giulia Roda [c], Alessandra Caponi [c], Peter James [b], Aldo Roda [a,*]

[a] Department of Pharmaceutical Sciences, University of Bologna, Via Belmeloro 6, Bologna 40126, Italy
[b] Department of Immunotechnology, Lund University, BMC D13, S-221 84 Lund, Sweden
[c] Gastroenterology Unit, University of Bologna, S.Orsola Hospital, Via Massarenti 9, Bologna 40138, Italy

ARTICLE INFO

*Article history:*
Received 6 May 2009
Accepted 3 August 2009
Available online 7 August 2009

*Keywords:*
Crohn's disease
Exoproteases
Label-free liquid chromatography–mass spectrometry
Protein profling
Proteomics

ABSTRACT

The identification of serum biomarkers for the diagnosis of inflammatory bowel diseases able to reduce the need for invasive tests represents a major goal in their therapy and follow-up. We report here a methodological approach for the evaluation of specific changes in the serum peptides abundance in healthy (H) and Crohn's disease (CD) subjects, based on a label-free LC ESI/Q-TOF differential mass spectrometry (MS) approach combined with targeted MS/MS analysis. The low molecular weight serum proteins were separated by RP nano-LC ESI/Q-TOF MS and the resulting datasets were aligned with msInspect software. The differently abundant peptides, evaluated using Proteios Software Environment, were identified by MS/MS analysis and database search. The identification of clusters of peptides resulting from proteins (such as fibrinogen-α) commonly involved in physiological processes lead to the evaluation of a possible role in CD of specific serum exoproteases. An assay based on synthetic peptides spiked into H, CD and ulcerative colitis (UC) serum samples as substrate, followed by MALDI MS and chemometric analysis of the metabolite patterns has been developed achieving a 100% discrimination between CD, UC and H subjects. The results are promising for the application of this approach as a simple tool for diagnostic aims and biomarker discovery in CD.

© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

The discovery of new diagnostic serum protein biomarkers requires powerful analytical tools able to identify, quantify and accurately evaluate their disease-related profile variations. Several comparative proteomics approaches based on the combination of different mass spectrometry (MS) techniques and the use of stable isotopes or mass tag peptides labeling have been developed. Among them, the most useful techniques are iTRAQ (Isobaric tags for relative and absolute quantitative), ICAT (Isotope coded affinity tags), and SILAC (Stable isotope labeling with amino acids in cell culture) [1]. These techniques allow an accurate relative quantification of proteins in several pooled samples simultaneously in a single run by using a different isobaric or isotopic tag for each sample. Due to the high cost of isotopic labeling reagents quantification methods based on a label-free strategy have been proposed as promising alternatives when large number of samples need to be analyzed [2], such as in clinical proteomics [3]. The label-free methods present many advantages when compared with the isotope-based ones. They are less expensive and require less steps during sample preparation. Furthermore they do not present any limit to the number of samples to be compared, do not increase the complexity of mass spectra and have a larger quantification dynamic range [2].

In our laboratory the main focus is on the discovery of new protein biomarkers for Crohn's and ulcerative colitis diseases, collectively termed inflammatory bowel diseases (IBDs) [4,5], which are chronic pathologies characterized by acute and remission phases. IBDs present non-specific symptoms and require endoscopic, histological or radiographic tests for a reliable diagnosis.

The recently proposed clinical serum antibodies-based tests do not present an adequate clinical sensitivity and specificity [6,7] and new serum biomarkers are highly requested in order to achieve an early diagnosis of IBDs and for drugs discovery and efficacy screening.

Despite the huge potential of MS-based proteomics for the identification of new protein biomarkers and the clinical importance of the low molecular weight (LMW) proteome for the diagnosis of diseases [8,9], only few papers have been so far published on proteomics and serum biomarker discovery in IBDs using mass spectrometry approaches. Meuwis et al. using SELDI-TOF MS combined with a multiple decision tree approach were able to identify

*Abbreviations:* CD, Crohn's disease; H, healthy; UC, ulcerative colitis; IBD, inflammatory bowel disease; LMW, low molecular weight.

* Corresponding author. Tel.: +39 051 343398; fax: +39 051 343398.
*E-mail address:* aldo.roda@unibo.it (A. Roda).
*URL:* http://www.anchem.unibo.it (A. Roda).

1570-0232/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2009.08.003

and purify serum biomarkers for IBDs with a relevant diagnostic value ((i.e.: platelet aggregation factor 4 (PF4), myeloid related protein 8 (MRP8), fibrinogen-α (FIBA) and Haptoglobin a2 (Hpa2)) [10]. In another paper our group used a serum protein profiling method based on selective solid-phase bulk extraction combined with MALDI-TOF MS and statistical chemometric analysis approaches for IBD classification and prediction [4]. A 96% correct prediction ability was achieved and the number of proteins involved in IBDs was reduced to 20, but with this MALDI approach, even if assisted by a group separation on different SPE principles, these proteins cannot be identified. Both studies allowed to obtain a greater clinical sensitivity and specificity (ability of a test to correctly identify disease and health, respectively) when compared with the use of conventional antibody based tests. These fingerprint statistical classification methodologies have been useful as a first approach able to minimize the number of serum proteins that might be involved in IBDs, but for diagnostic purposes their identification is necessary.

We report here the development of a label-free nano-liquid chromatography electrospray ionization/Quadrupole-Time-of-Flight (LC ESI/QTOF) differential mass spectrometry approach combined with targeted MS/MS analysis for the identification of the LMW serum proteins most affected in Crohn's disease as potential new candidates serum biomarkers.

Serum samples from healthy subjects (H) and Crohn's disease patients (CD) were analyzed and only the LMW proteins were studied as candidate biomarkers for CD. Serum is a complex matrix containing thousands of proteins in a wide range of concentrations, but the presence of few high abundant proteins such as albumin or IgG highly interfere during separation techniques with the separation of LMW and low abundant proteins which are supposed to be rich in diagnosis information. Immunoaffinity chromatography columns can be used for their depletion [11,12], but due to their high cost associated with problems of specificity [13,14] and with the potential loss of proteins [15] a more reproducible and cheaper method which can be clinically applied should be used. In this work the LMW proteins were enriched by ultrafiltration using molecular filter devices with a 10 kDa cut-off.

The LMW protein mixtures were analyzed without performing any enzymatic digestion. Proteins and peptides were separated by reverse phase (RP) nano-LC ESI/QTOF MS analysis with accurate alignment of the chromatographic runs using the software msInspect [16].

The relative quantification and the evaluation of the differently expressed peptides and proteins in the pooled samples will be achieved using the tryptic digest peptides of Alcohol dehydrogenase as internal standards for the alignment of the LC–MS analysis and for the normalization of the acquired data. The open-source Proteios Software Environment (ProSE) [17] was used for the statistical analysis while for the identification of the peptides RP nano-LC ESI/QTOF MS/MS analysis and database search were performed.

The identification of clusters of peptides which are thought to be generated by serum exoproteases (enzymes responsible for the cleavage of biologically active peptides) lead to the evaluation of their possible specific role in CD. To evaluate the activity and the specificity of these exoproteases a method for the analysis of the degradation products of synthetic peptides spiked into H, CD and ulcerative colitis (UC) serum samples, based on matrix assisted laser/desorption ionization (MALDI) MS followed by chemometric statistical analysis of the small peptides obtained after the enzymatic cleavage has been developed.

The involvement of specific enzymes in IBDs was already described in a previous paper where procoagulant peptides (fibrinopeptide A) and a resolution of experimental colitis correlated with dipeptidyl peptidase inhibition have been reported in mice [18].

Since recently serum platelet-derived microparticles (PDMPs) have been involved in IBDs and their serum levels were correlated with disease activities and sP-selectin levels [19] we speculated that platelet-derived MPs may also contain those enzymes. The study of the subproteome of the MP in inflammatory bowel diseases represents an open unexplored field and introduce the concept of functional proteomic in IBD where enzyme could be up or down regulated.

## 2. Experimental

### 2.1. Materials and reagents

All the analytical grade reagents, the alcohol deydrogenase tryptic digest, the Lowry protein assay kits and the solvents were purchased from Sigma–Aldrich (St. Louis, MO, USA).

### 2.2. Subjects studied

A total of 63 subjects, including 48 healthy donors volunteers (H; 22 M and 26 F) and 15 Crohn's disease patients (CD; 9 M and 6 F), were admitted to this study. All the subjects were ranging in age from 20 to 65 years. The diagnosis of the CD was achieved evaluating clinical symptoms, clinical chemistry data and conventional endoscopical procedures and only patients with active disease (Crohn's Disease Activity Index (CDAI) > 150 [20]) were included. The CD patients were not subjected to anticoagulant therapies.

Blood samples were provided by the Gastroenterology Unit at S.Orsola University Hospital of Bologna (Italy) after obtaining informed consent from the patient. From each subject 5 mL of blood sample were collected in the presence of serum beads clot activator, allowed to clot for 30 min at room temperature and centrifuged at 4000 rpm for 10 min [21,22]. The obtained serum samples were distributed into aliquots and stored frozen in plastic vials at −80 °C until use. Six pools of serum samples from healthy subjects (250 μL from 8 different serum samples each) and five pools of serum samples from Crohn's disease patients in the active phase of the disease (250 μL from 3 different serum sample each) were created and aliquoted. Haemolysis was not observed in the samples.

A table containing the demographic of the patients and the composition of the pools is available in the supplementary material.

### 2.3. Low molecular weight serum proteins enrichment

Protein quantification of each pooled serum sample was performed in triplicate by Lowry protein assay using bovine serum albumin (BSA) as a standard according to the manufacturers' instructions (Sigma–Aldrich).

For each sample 25 mg of proteins amount was diluted to 2 mL of 25 mM $NH_4HCO_3$ pH 8.2 containing 10% (v/v) acetonitrile (ACN) and incubated at room temperature for 30 min to disrupt the protein–protein interactions [23,24]. Ten picomoles of a tryptic digest of alcohol deydrogenase (ADH) was added as internal standard.

Each diluted serum sample was transferred to a Centricon centrifugal filters with a 10 kDa molecular weight cut-off (MWCO) (Millipore Corporation, Bedford, MA, USA) to deplete the high molecular weight proteins [25]. The samples were centrifuged at $3000 \times g$ until 90% of the diluted serum had passed through the membrane. At the end of the process approximately 1400 μL of LMW protein enriched serum samples were collected and distributed into aliquots. Two hundreds microliters of the pooled serum collected filtrates were lyophilized to dryness and resuspended in 30 μL of deionized $H_2O$ containing 0.1% formic acid (FA) for the following LC–MS analysis.

To evaluate the reproducibility of the MWCO ultrafiltration two hundreds microliters of a solution containing five standards proteins (alcohol dehydrogenase 1 (P00330; 36.8 kDa; 50 ng/μL); hemoglobin subunit beta (P68871; 16 kDa; 25 ng/μL); bovin serum albumin (P02769; 66 kDa; 75 ng/μL); alpha-S1-casein (P02662; 24.5 kDa; 50 ng/μL); myoglobin (P68082; 17 kDa; 100 ng/μL)) have been ultrafiltrated as described above and the filtrate has been subjected to over-night digestion with trypsin (protein:trypsin ratio 50:1 (w/w)). For the evaluation of the recovery of the peptides after the ultrafiltration process the five standard proteins have been previously digested with trypsin and afterwards ultrafiltrated and analyzed by LC–MS/MS. The LC–MS analysis of both the ultrafiltrated and not-ultrafiltrated peptide mixture was performed in duplicate for the quantification of the recovery.

To evaluate the reproducibility of LMW enrichment process in real serum protein samples the same experiment has been performed four times on serum aliquots of an H subject and of a CD patient and the resulting peptides/proteins were analyzed by LC–MS.

### *2.4. Liquid chromatography and mass spectrometry*

LC–MS analysis was performed on a CapLC (Waters, Manchester, UK) with flow splitting from 5 μL/min to 250 nL/min, connected with a nano electrospray interface to a QTOF Ultima (Waters) using MassLynx v4.0 software as operating software. An Atlantis dC18 NanoEase column (150 mm × 0.075 mm, 3 μm) (Waters) with a C18 Intersil precolumn (0.3 mm × 5 mm, 3 μm particle size) (LC-Packings, Skandinaviska Genetec AB) was used.

The mobile phase had a constant concentration of 0.1% formic acid, with an acetonitrile gradient (after a 10 min desalting step) from 5 to 60% over 125 min, followed by 10 min column cleaning at 80% acetonitrile, and 15 min equilibration. Two blank injections with a 60 min gradient plus column washing were run between samples to minimize sample carry over, and every five samples a quality control analysis (500 fmol of yeast enolase tryptic digest) was performed with a 45 min gradient to evaluate the experimental variation. For quantitative experiments, the QTOF was set to scan in profile mode *m*/*z* 400–1800 with 1.9 s per scan and 0.1 s of scan delay.

All the samples were analyzed in triplicate, each with a 3 μl injection. For targeted MS/MS, 6 μl sample was injected, the same LC gradient was run and the survey scan used 1 s scan time and a peak limit of 20 count to switch to MS/MS mode. For inclusion lists the time tolerance was set to 180 s.

### *2.5. MS data analysis and generation of a targeted MS/MS peptide mass list*

The whole process of MS data analysis, include list generation and feature peptides generation has been already described elsewhere by our research group [5]. Briefly, the LC–MS raw data files were converted to mzXML using massWolf (version 1.4, http://sashimi.sourceforge.net) and the peptide feature finding was performed using msInspect version 1.01 [16]. The resulting feature lists were aligned using msInspect (setting a mass window of 0.2 *m*/*z* and a time window of 75 scans) and the detailed results files were used for further processing. For the alignments of LMW serum protein analysis the triplicates of each pool of serum sample from healthy subjects were analyzed with the triplicates of each pool of serum sample from Crohn's disease patients. The normalization was performed using the msInspect algorithm during the alignment step.

The aligned peak lists were analyzed using newly developed plug-ins for the ProSE 2.1 platform [17]. First the MsInspect Details Analyser (version 0.91) plug-in was used to find features which were significantly upregulated in the CD samples and to automatically produce include lists for MS/MS identification. For a feature to be identified as significantly upregulated it had to be present in at least two sample replicates and have a *p*-value of less than 0.05 in a homoscedastic Student's *t*-test. For the *t*-test the total intensities, which represent the integrated peak volumes, were used. For features where peaks could not be found in the healthy control samples, a value of 50 ion counts was used, which was an estimate for the detection level in the present setup. Features which were upregulated at least 1.5 times in the CD sample compared to the healthy control and with a significant *p*-value were sorted according to intensity and put into include lists with a maximum of 300 peaks per include list. The retention time of the last acquired sample was used in the include list.

### *2.6. Feature peptide identification*

To generate peak lists for peptide identification, ProteinLynx Global Server 2.2 (Waters) was used. The xml format peak lists were converted to mzData using ProSE. Mascot version 2.1.02 (www.matrixscience.com) was used for peptide identification.

The Sprot human database, version 53.1 was used, 191 913 sequences in total. The search settings were 0.1 Da precursor and fragment tolerances, no fixed modification carbamidomethylation of cysteine and none enzyme. The search results were exported as XML and matched with MS features using a ProSE plug-in, with a retention time tolerance of 100 s and a mass tolerance of 0.12 Da.

Protein were considered correctly identified when at least two different peptides (with significant individual score, i.e., $p<0.05$) were present.

### *2.7. Protein ratio calculation*

To calculate the mean and standard deviations for protein fold-changes, all individual peptide peak intensities of identified proteins were considered. First the standard deviation of mean was calculated for the intensities of each peptide group. The fold-change and fold-change variance was then calculated for each individual peptide. To calculate the whole protein fold-change and variance, peptides with small variances in fold-change were used. The calculations are included in the supplemental data tables.

### *2.8. Exoprotease activity assay*

The exoprotease activity assay was performed on 20 μL aliquots from 3 CD, 3 UC and 3 H individual serum samples. These samples were not included in the set of pool samples used for the hypothesis-generating experiments.

For the confirmation of the results by other method, two synthetic peptides were used as substrate for the exoprotease activity, both of them purchased from Bachem Distribution Service GmbH (Weil am Rhein, Germany): Fibrinopeptide A without the N-terminal Ala (FPA(-A)) and Complement 3f (C3f). FPA(-A) was chosen because of the fast degradation time of normal FPA [26,27]. FPA(-A) solution was prepared in water at a concentration of 20 pmol/μL, while C3f solution at a concentration of 8 pmol/μL in water.

The whole protocol is adapted from the sequence-specific exopeptidase activity test developed by the group of Villanueva [27]. Briefly, three series of aliquots of 20 μL of serum from 3 CD, 3 UC and 3 H individual serum samples were thawed on ice. One series was spiked with 18 μL (144 pmol in total) of C3f, one series with 18 μL (360 pmol in total) of FPA(-A) and one with 18 μL of $H_2O$ (to evaluate the exoprotease activity on the endogenous peptides). After an optimization phase the incubation times were fixed

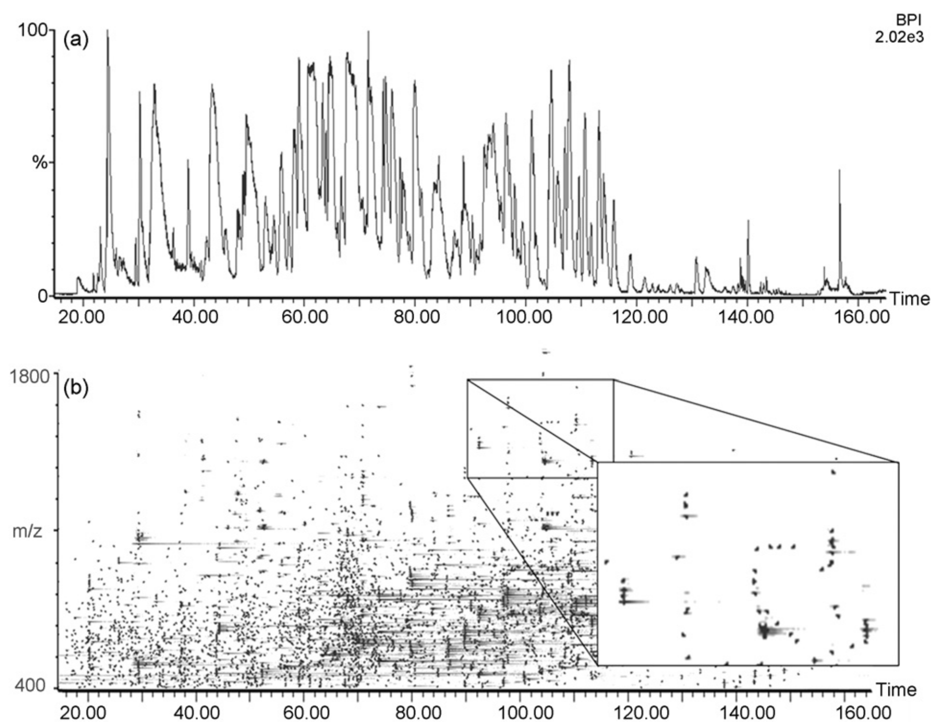



**Fig. 1.** Example of LC ESI/QTOF MS analysis of a serum protein sample. (a) BPI visualization of a chromatogram and (b) msInspect bi-dimensional data visualization. The spot dimension is the ion intensity. In the zoom box an example of extracted features (peaks) is shown.

to 0 min (t0) for C3f, FPA(-A) and $H_2O$, to 1 h for C3f and $H_2O$, and to 5 h for FPA(-A) and $H_2O$. Two microliters of [Glu]-Fibrinopeptide B (GFP, Sigma) at a concentration of 100 pmol/μL in water was used as internal standard for the normalization of the results. GFP is always spiked at the end of the incubation time in order to avoid its degradation. Immediately at the end of the incubation time the peptides are extracted using a standard C18 solid-phase bulk extraction (Silica C18/Corasil, Bondapak® 37–50 μm) and eluted with 6 μL of 50% ACN/0.1% TFA. For each sample 1 μL has been loaded on a MALDI-target plate, mixed with 1 μL of α-cyano-4-hydroxycinnamic acid (ACCA) matrix solution (10 mg/mL, 50% ACN/0.1% TFA, Sigma) and analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) using an Applied Biosystems Voyager-DE PRO mass spectrometer equipped with a 337-nm nitrogen laser. Analysis were performed in reflector positive-ion mode, using delayed extraction. The acceleration voltage was 20 kV, guide wire was 0.002 of the accelerating voltage, grid voltage was 75%, and the delay time was 100 ns. For each spectrum, collected manually in the $m/z$ range 500–2500, five hundreds laser shots delivered in five sets of 100 shots were averaged. External mass calibration was performed using the calibration mixture 1 of the Sequazyme peptide mass standard kit (Applied Biosystems, Darmstadt, Germany), containing des-$Arg^1$-Bradikynin (1 pmol/μL), Angiotensin I (1.3 pmol/μL), $Glu^1$-Fibrinopeptide B (1.3 pmol/μL) and Neurotensin (50 fmol/μL).

On the acquired spectra the ratio of the intensities of all the peaks derived from the degradation of C3f and FPA(-A) and of the intensities of internal standard GFP were calculated, and the normalized intensities of the degradation products of the endogenous C3f and FPA(-A) at the same incubation time were subtracted. The obtained data matrices were used for clustering analysis using the chemometrical software package V-PARVUS [28]. For clustering analysis the Euclidean distance was used, computing the similarity on the basis of the closest object (single linkage).

## 3. Results and discussion

### 3.1. Serum samples

In the CD serum proteomics approach the variations of the abundance of the LMW proteins and their potential use as new biomarkers were investigated. The study was limited to LMW proteins since they might contain the highest potential diagnosis ability [8,9].

According to already published studies that describe the influence of clotting time, blood collection tubes and therapy of the patients on the peptides composition of serum, an optimized standard protocol has been developed to control the production of degradation peptides due to ex vivo enzymatic activity on serum proteins [21,22]. The blood samples were allowed to clot for 30 min, since the use of longer clotting times lead to consistent changes in the LMW composition and to the decrease of peptides such as fibrinopeptide A. Afterwards the serum samples were centrifuged, aliquoted and stored at −80 °C in order to avoid any enzymatic activity without the need of proteases inhibitors.

### 3.2. Preanalytical step

The preanalytical step has been simplified to prevent loss of peptides and proteins and the serum samples are just slightly denatured using 10% acetonitrile and ultrafiltrated using a 10 kDa cut-off molecular filter device.

The reproducibility of the LMW protein enrichment process has been evaluated with three set of experiments. In the first experiment ($n$ = 5) the ability of the MWCO to enrich only the proteins with MW lower than 10 kDa was evaluated performing the ultrafiltration on a standard solution containing five standards proteins with MW higher than the molecular filter cut-off. The filtrate has been subjected to tryptic digestion and analyzed by LC–MS/MS and

no proteins were identified. In the second experiment ($n = 5$) the recovery of the peptides after the ultrafiltration process was evaluated performing the tryptic digestion of the five standard proteins followed by ultrafiltration and LC–MS/MS analysis. The identification of the proteins was always achieved with a total of 38 identified peptides (30 of them evaluated in all the analysis) (data shown in the supplementary material). The LC–MS analysis of both the ultrafiltered and not-ultrafiltered peptide mixture and the comparison of the intensities of 10 peptides peaks allowed to estimate an average $90.3 \pm 5.2\%$ recovery.

In the last series of experiments the reproducibility of the LMW enrichment process has been evaluated using real serum protein samples. The LMW protein enrichment step has been performed four times with an identical aliquot of healthy subject serum sample and four times with an identical aliquot of CD patient serum sample, and the obtained peptide/protein mixtures were analyzed by LC–MS. The general profiles were comparable (supplementary material, figure S-1).

### 3.3. LC–ESI-mass spectrometry

In shotgun proteomics studies, where the final purpose is the identification of a high number of proteins, a multidimensional chromatography approach combined with data dependent MS/MS analysis is commonly used. Otherwise during tandem MS analysis not all the eluting peptides can be fragmented and a loss of information can occur. In this work the aim was the identification of only the peptides/proteins whose abundance change in CD and H samples and for this reason a monodimensional chromatography approach combined with MS analysis has been developed, thus reducing the time of analysis and allowing to perform the differential statistical analysis on all the eluting peptides (and not only on the fragmented ones).

Peptide/protein samples were analyzed in triplicate by label-free nano-LC–MS. The selected instrumental conditions were optimized to achieve a reasonable LMW protein separation combined with an optimal electrospray performance operating in positive mode. A typical chromatogram is reported in Fig. 1a.

Due to the length of the LC analysis (165 min) and to the flow that is passively split in the CapLC (from 5 μL/min to 250 nL/min) little variations in the retention time among the LC runs can be observed (less than 1%). The small non-linear variations in the retention time have been evaluated adding 10 pmol of tryptic digest of Alcohol dehydrogenase (ADH) as internal standard to each pooled serum sample. For all the peaks on all the LC–MS analysis the alignment was performed using a quite wide retention time windows (75 scans), with an associated $m/z$ window of only 0.1 Da. ADH has been used also to evaluate LC–MS variability and the reproducibility of the peak intensities among the different LC–MS runs. The coefficient of variation (CV) of the intensities of eight replicates peaks of ADH has been calculated and CV values ranging from 17% to 24% have been obtained (data reported in Table 1).

These data are relative to a standard mixture of tryptic peptides (characterized by a quite specific polarity and charge) spiked into serum sample, and for this reason are not representative of the whole serum proteome. However since the only role of ADH was to monitor the retention time and intensity shifts during the multiple runs and it was not used for any calculation of fold-change variations, its potential different polarity and charge was not considered as relevant. For this reason these results can be used for a general estimation of the reproducibility of the LMW enrichment step in a complex protein samples such as serum.

The LC–MS experimental data were imported into the open-source software msInspect and the bi-dimensional maps of the LC–MS runs were created for each sample (an example is shown in Fig. 1b). A specific algorithm for feature finding was used to fil-

**Table 1**
Evaluation of the reproducibility of the peak intensities among the different LC–MS analysis. Ten picomoles of tryptic digest of alcohol dehydrogenase I (yeast) has been added to each serum samples as internal standard and LC–MS analysis was performed. The retention time shifts and the intensity variations of eight selected peaks during all the LC–MS analysis were calculated.

| $m/z$ | Mr (Da) | Charge | LC information | | | | | | MS information | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total # scan | Retention time (min) | Scan range | | Δ scan | Δ scan/total scan (%) | % of samples | Average intensity | Standard deviation | CV% | Peptide sequence |
| 844.4 | 843.47 | 1 | 9800 | 38.4 | 631 | 701 | 70 | 0.7 | 100.0 | 44.29 | 9.0 | 20.4 | S(acet)IPETQK |
| 447.3 | 892.54 | 2 | | 40.7 | 696 | 770 | 74 | 0.8 | 100.0 | 135.47 | 28.9 | 21.3 | DIPVPKPK |
| 418.7 | 835.4 | 2 | | 48.7 | 921 | 990 | 69 | 0.7 | 100.0 | 120.11 | 26.8 | 22.3 | IGDYAGIK |
| 568.8 | 1135.6 | 2 | | 53 | 1068 | 1115 | 47 | 0.5 | 86.4 | 15.02 | 2.6 | 17.0 | EALDFFAR |
| 693.9 | 1385.8 | 2 | | 70 | 1547 | 1612 | 65 | 0.7 | 95.5 | 31.28 | 6.0 | 19.1 | EELFR |
| 507.3 | 1012.6 | 2 | | 75.4 | 1715 | 1771 | 56 | 0.6 | 95.5 | 70.06 | 16.9 | 24.2 | ANELLINVK |
| 626.3 | 1250.7 | 2 | | 80.3 | 1869 | 1917 | 48 | 0.5 | 77.3 | 16.66 | 4.0 | 24.2 | SISIVGSYVGNR |
| 484.7 | 967.5 | 2 | | 98.7 | 2424 | 2457 | 33 | 0.3 | 90.1 | 14.41 | 3.5 | 24.0 | GVIFYESHGK |

**Table 2**
Proteins identified in CD and H samples by LC–MS/MS analysis and Mascot database search ("–" and "+" correspond respectively to lower and higher ion intensities of the derivative peptides observed in CD samples during ESI-QTOF MS analysis).

| Protein name | Entry name | Mr (Da) | No of AA | Derivative peptides ion intensities in CD | Fold-change |
|---|---|---|---|---|---|
| Complement C3 | C3 | 187148 | 1663 | + | 2.3 |
| Fibrinogen alpha chain | FIBA | 94973 | 866 | + | 3.0 |
| Fibrinopeptide A | FPA | 1537 | 16 | + | 2.7 |
| Complement C3f fragment | C3f | 2021 | 17 | + | 2.5 |
| Complement C4-A | C4-A | 192771 | 1744 | + | 1.7 |
| Apolipoprotein E | APO E | 36154 | 317 | – | 2.5 |
| Apolipoprotein A-IV | APOA4 | 45399 | 396 | – | 1.9 |
| Alpha-2-HS-glycoprotein (Fetuin A) | FETUA | 39325 | 367 | – | 1.7 |

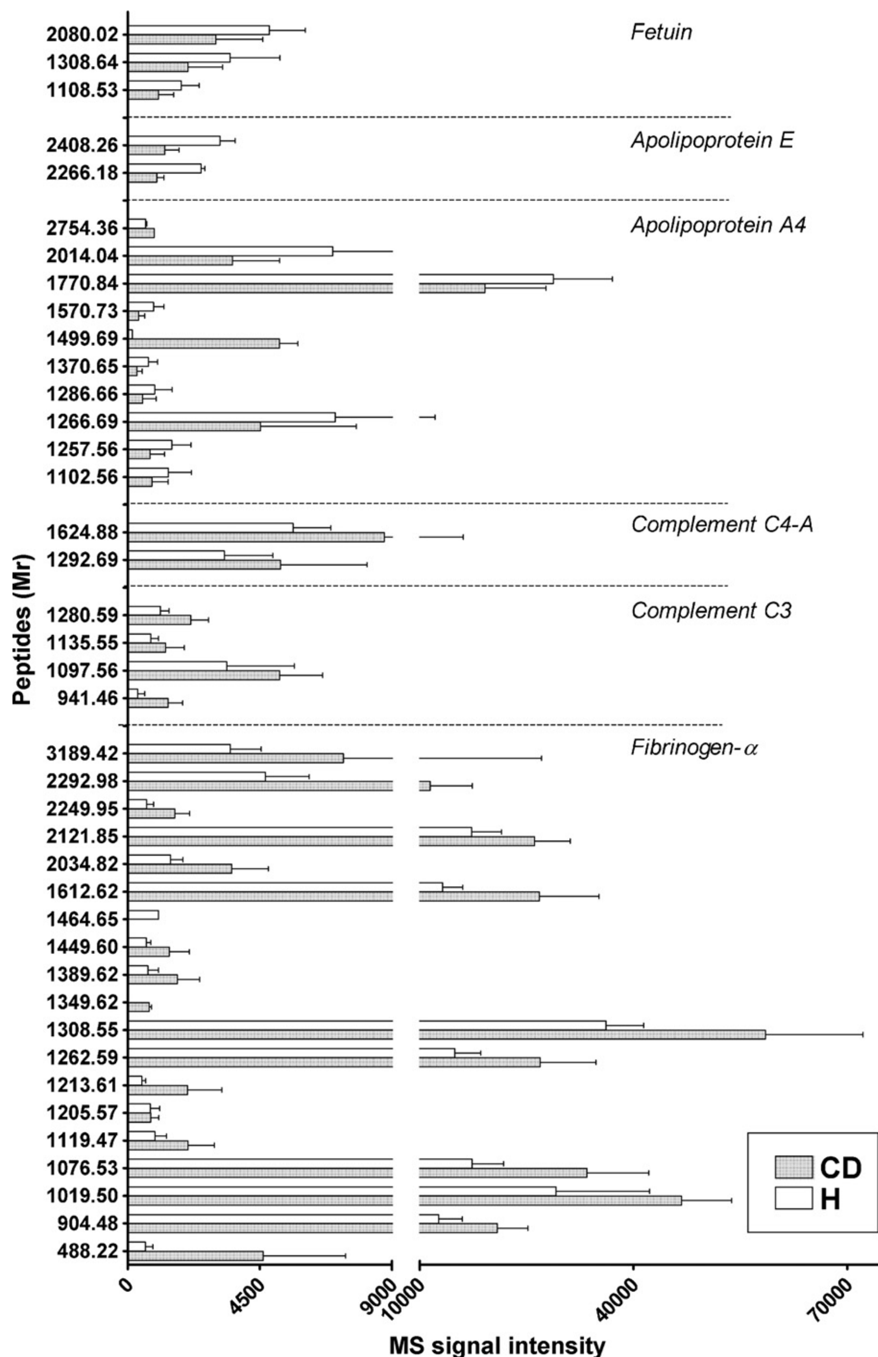



**Fig. 2.** Intensities of the differently expressed peptides in serum samples from healthy subjects and Crohn's disease patients. The monoisotopic masses of the peptides are reported.

**Table 3**
Serum peptides patterns in Crohn's disease patients and healthy subjects, compared to serum peptides signature for breast, prostate and bladder cancer evaluated in another research work by Villanueva et al. [26] The peptides are monoisotopic. Coloured box corresponds to peptide presence, "+" to overexpression, "–" to lower expression, "=" and "/" respectively to equally and randomly observed.

| | | | Cancer | | | Crohn's disease | |
|---|---|---|---|---|---|---|---|
| | | | Prostate | Bladder | Breast | Crohn's disease | Healthy |
| FPA | **1536.68** | ADSGEGDFLAEGGGVR | - | - | - | | |
| | 1465.65 | DSGEGDFLAEGGGVR | - | - | | + | |
| | 1350.64 | SGEGDFLAEGGGVR | - | - | - | + | |
| | 1263.6 | GEGDFLAEGGGVR | - | - | - | + | |
| | 1206.57 | EGDFLAEGGGVR | - | - | | + | |
| | 1077.53 | GDFLAEGGGVR | - | - | | + | |
| | 1020.47 | DFLAEGGGVR | - | - | | + | |
| | 905.5 | FLAEGGGVR | | - | | + | |
| | 758.45 | LAEGGGVR | - | - | | | |
| Fibrinogen α | 3261.43 | (K) SSSYSKQFTSSTSYNRGDSTFESKSYKMA | - | - | | | |
| | 3190.36 | (K) SSSYSKQFTSSTSYNRGDSTFESKSYKM | - | - | | | |
| | 2931.2 | (K) SSSYSKQFTSSTSYNRGDSTFESKSY | | - | | | |
| | 2768.26 | (K) SSSYSKQFTSSTSYNRGDSTFESKS | | - | | | |
| | 2553.01 | (K) SSSYSKQFTSSTSYNRGDSTFES | = | = | = | | |
| | 2379.03 | .SSYSKQFTSSTSYNRGDSTFE | | - | | | |
| | 1390.62 | NRGDSTFESKSY | | | | + | |
| | 1120.47 | GDSTFESKSY | | | | + | |
| | 1063.45 | DSTFESKSY | | | | + | |
| | 1450.53 | STSYNRGDSTFES | | | | + | |
| | 2816.25 | (R) GSESGIPTNTKESSSHHPGIAEFPSRG (K) | | | | | |
| | 2293.97 | MADEAGSEADHEGTHSTKRGHA | | | | + | |
| | 1872.74 | KMADEAGSEADHEGTHST | | | | + | |
| | 2035.33 | YKMADEAGSEADHEGTHST | | | | + | |
| | 2022.87 | SYKMADEAGSEADHEGTHST | | | | + | |
| | 3239.22 | SYKMADEAGSEADHEGTHSTKRGHAKSRPV (R) | / | / | / | | |
| | 2659.03 | DEAGSEADHEGTHSTKRGHAKSRPV (R) | | | | | |
| C3f | 1098.55 | HWESALLR | | | | + | |
| | 2021.06 | SSKITHRIHWESASLLR | = | = | = | | |
| | 1864.95 | SSKITHRIHWESASLL. | | | - | | |
| | 1777.93 | SKITHRIHWESASLL. | | | | | |
| | 1690.9 | KITHRIHWESASLL. | | | | | |
| | 1562.84 | ITHRIHWESASLL. | | | - | | |
| | 1449.76 | THRIHWESASLL. | | | | | |
| | 1348.7 | HRIHWESASLL. | / | / | / | | |
| | 1211.7 | RIHWESASLL. | | | | | |
| | 1055.6 | IHWESASLL. | | | | + | |
| | 942.43 | HWESASLL. | | | - | | |
| | 1851.88 | SSKITHRIHWESASL.. | | | | | |
| | 1136.55 | THRIHWESA. | | | | + | |
| C4a | 1895.99 | RNGFKSHALQLNNRQI (R) | | | | | |
| | 1739.93 | NGFKSHALQLNNRQI (R) | | | | | |
| | 1626.85 | NGFKSHALQLNNRQI. | | | | | |
| | 1498.91 | NGFKSHALQLNNR.. | | | | | |
| | 1625.85 | GFKSHALQLNNRQI. | | | | + | |
| | 1293.68 | SHALQLNNRQI. | | | | + | |
| | 3200.52 | (R) GLEEELQFSLGSKINUKVGGNSKGTLKVLR | / | / | / | | |
| | 2704.13 | (R) GLEEELQFSLGSKINUKVGGNSKGTL | | | | | |
| | 2305.2 | (R) GLEEELQFSLGSKINUKVGGNS | | | | | |
| | 1762.87 | (R) GLEEELQFSLGSKINUKV | | - | | | |
| ApoA-IV | 2508.16 | ISESEEELRQRLAPLAEDVRGNL (K) | | | | | |
| | 2755.2 | (K) GNTEGLQKSLAELGGHLDQQVEEFR | | | | | |
| | 1927.94 | SLAELGGHLDQQVEEFR | | | | | |
| | 1771.81 | SLAELGGHLDQQVEEF. | | | | | + |
| | 1571.75 | AELGGHLDQQVEEF. | | | | | + |
| | 1500.71 | ELGGHLDQQVEEF. | | | | | + |
| | 1371.66 | LGGHLDQQVEEF. | | | | | + |
| | 1258.57 | GGHLDQQVEEF. | | | | | + |
| ApoE | 2565.45 | (R) AATVGSLAGQPLQERAQAWGERLR | | | | | + |
| | 2409.13 | (R) AATVGSLAGQPLQERAQAWGERL. | | | | | |
| | 2267.12 | TVGSLAGQPLQERAQAWGERL. | | | | | + |

ter the protein/peptides signals according to the criteria described in the Experimental section, obtaining a distribution of peptides ranging from 398.2 ($m/z$ 399.2, $1^+$) to 10672.2 ($m/z$ 1779.7, $6^+$) Da.

The differently expressed LMW proteins in CD and H serum samples (at least 1.5-fold-change) were evaluated by statistical analysis using ProSE. Two alignment were performed for every set of data to be compared: one to the last acquired H subjects sample and one to the last acquired CD sample. This allowed to create for each category an include list containing the overexpressed peptides (266 peptides for H and 271 peptides for CD), reducing the problem of retention time shift due to the lag of time between MS and MS/MS analysis (data reported in the supplementary Tables 3 and 4).

The use of targeted MS/MS analysis instead of data dependent MS/MS analysis confers the advantage of performing the fragmentation of only the peptides that are differently expressed avoiding to fragment the peptides not relevant for the clinical issue.

### 3.4. Serum proteomics

The nano-LC–MS/MS analysis of the proteins/peptides evaluated as differently expressed followed by database search allowed the identification of many peptides. Most of them resulted from few proteins commonly involved in the inflammatory and the coagulation processes while few of them were not associated to any protein. Table 2 lists the identified proteins and the fold-change variations in the ion intensities of the derivative peptides observed during ESI-QTOF MS analysis. Since the use of targeted MS/MS analysis of only differently expressed peptides lead to a small number of identified peptides the confidence in protein identification was evaluated using the Mascot $p$-value (supplementary material, Table 5) instead of the target-decoy database searching, which could be less accurate.

In Fig. 2 the variations of the intensities of every single peptide in H and CD serum samples is shown (data reported in the supplementary Table 5). For each protein most of the single peptides ratios were very similar, thus confirming the biological meaning of the results. The divergent relative abundance ratios of few peptides used for the identification of proteins (i.e. peptide AEGGGV from FIBA, having fold-change variation 7.7 instead of the average FIBA fold-change of 2.7) can be explained by the presence of peptides shared among multiple protein isoforms having different abundance [5,29] or by the presence of post-translational modification of the protein with a different concentration of the modified peptide.

Several studies have been already published describing the importance of the inflammation, fibrinolysis and coagulation network in the pathogenesis of IBDs; activation of coagulation pathways in IBDs is well described but it is unclear if it represents a consequence or a cause of the disease [30].

Among the identified proteins fibrinopeptide A (FPA), a peptide released from fibrinogen during clotting of the blood, has been evaluated as the peptide whose concentration mostly increase in CD. This data is in agreement with the already available data on LMW biomarkers [10] but it is poor specific. In CD serum samples an increased concentration of peptides from complement 3 protein (C3), that plays a central role in the activation of the complement system, and of its fragment C3f have been found too The identification in CD of an higher amount of peptides from apolipoprotein A-IV is in agreement with recent studies that showed how apolipoprotein A-IV inhibits experimental colitis in vivo [31] and can be considered as an independent predictor of disease activity in patients with IBD [32].

Peptides from apolipoprotein E have been identified as less abundant in CD, confirming the role of the protein as “inflammatory imbalance” between pro- and anti-inflammatory mediators [33].

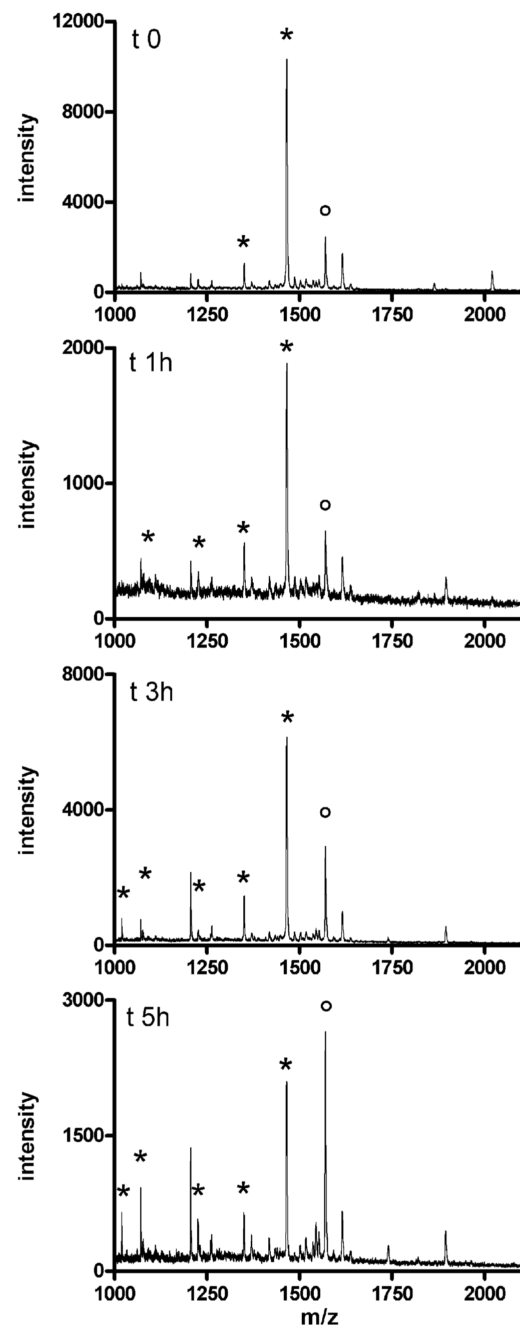



**Fig. 3.** MALDI-TOF MS spectra of serum samples spiked with Glu-fibrinopeptide B (GFP) as internal standard and with Fibrinopeptide A(-A) acquired at different incubation times (t0 and 1, 3, 5 h). It is possible to observe the decrease in the FPA(-A) intensity (*) compared to GFP (°) and an increase in some degradation products (*).

Despite the differential LC–MS analysis has been performed only on the LMW proteome enriched during the preanalytical step all the identified proteins have a MW much higher than 10 kDa (the MWCO applied to the serum). This might be explained by the activity of serum protease which acts on these proteins and produces their degradation peptides.

Furthermore these differently expressed peptides falls into tight clusters that are thought to be generated by exopeptidase activities and seem to allow the discrimination between CD patients and H subjects (Table 2).

### 3.5. Exoprotease activity

Recently Villanueva et al. demonstrated that some biomarkers are not directly expressed by the diseased tissues, but are ex vivo products from exoprotease activity acting on proteins involved in the process of coagulation and of activation of the complement cascade [26,27]. In his works he focused on different types of cancers (breast, bladder and prostate) because they involve the transformation of altered cells types that produce specific proteins. Since the same theory can be applied to every type of disease it has been supposed that specific serum exoproteases may be involved in CD and that the obtained peptide signatures may be potentially used as biomarkers.

Most of the identified peptides are the same found in both this and Villanueva work (Table 3 lists and compares these clusters of peptides). The overlapping among peptides involved in different diseases underlines the importance of the abundance of the proteins in the sample as substrate for protease activity. Indeed the concentration of these peptides is not a measure of the concentration of the parent proteins, since high abundance proteins such as fibrinogen-α, complement 3 and apolipoproteins are only the target of the exoproteases and are not themselves really overexpressed.

To confirm the results with an alternative method and to evaluate the specificity of the peptide clusters for CD an exoprotease activity assay (adapted from the sequence-specific exopeptidase activity test (SSEAT)) [27] has been performed on individual serum samples from subjects not included in the set used for the generation of the hypothesis and from patients affected by ulcerative colitis.

Briefly, two randomly selected standards of cluster precursor peptide (FPA(-A) and C3f) were added individually in the same amount into each CD, UC and H serum sample. After a short incubation the peptides were extracted, the MALDI-TOF MS analysis was performed and the degradation product signatures acquired. Fig. 3 shows the MALDI-TOF MS spectra acquired for FPA(-A) at different incubation time. Because of the semi-quantitative nature of MALDI-TOF MS due to the crystallization and ionization processes (matrix application, laser intensity...), the reproducibility of the method has been evaluated. The serum sample of an H subject was used for inter- ($n = 5$) and intra-day ($n = 5$) reproducibility studies and for the intensities of all the peptides in the spectra a medium CV of 17% ranging from 0.4 to 30.4% was achieved. The reproducibility of the whole exoprotease activity test (spike of the peptides, C18 solid-phase extraction, and MALDI MS analysis) has been evaluated too, obtaining a medium CV value of 23.1%.

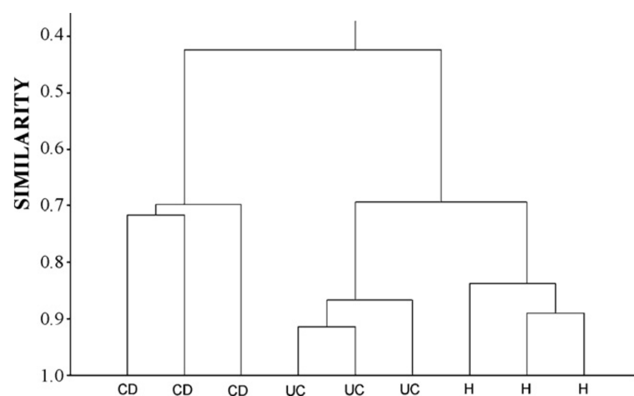



**Fig. 4.** Dendrogram obtained after cluster analysis of the normalized intensities of the degradation products of FPA(-A) and C3f at 5 h and 1 h of incubation respectively, spiked in 9 individual serum samples (3 from healthy subjects, 3 from Crohn's disease patients and 3 from ulcerative colitis patients).

For each sample the degradation peptides intensities of both FPA(-A) and C3f cluster precursors were normalized to the internal standard GFP intensity and the contribute of the degradation peptides from the endogenous precursors were subtracted (supplementary material, Table 6). Multivariate clustering analysis was performed on these peptide fingerprints allowing to achieve a 100% discrimination among the three categories of subjects. By using in combination the two cluster precursor peptides it was possible in particular to create a fingerprint of peptides able to discriminate the CD patients from the UC patients and the H subjects (which can be grouped in the same cluster), thus demonstrating the specificity of this activity. In Fig. 4 the dendrogram obtained after cluster analysis of these few samples is shown.

## 4. Conclusions

The developed label-free nano-LC ESI-QTOF MS comparative analysis approach combined with the LMW serum proteins enrichment step proved to be useful for the identification of the LMW serum proteins whose abundance changes of at least 1.5-fold in healthy subjects and Crohn's disease patients and for the discrimination among H subjects and CD patients serum samples.

This reproducible analytical method allowed to identify clusters of peptides that are thought to be generated by unidentified exoproteases acting on proteins involved in the coagulation and complement activation pathways, thus leading to the evaluation of a possible specific role of serum exoprotease in CD. Particularly interesting was the discovery of a cluster of eight peptides from fibrinopeptide A (FPA) and three from complement 3f (C3f), all of them more abundant in CD patients. Their ability to specifically discriminate the CD patients from H subjects and patients affected from other inflammatory diseases (such as UC) was confirmed developing a simple approach based on the use of synthetic peptides spiked as surrogate substrates into H, UC and CD serum samples combined with direct MALDI-TOF MS analysis of the obtained degradation products. The multivariate cluster analysis applied on the acquired metabolites fingerprints allowed a 100% discrimination among the three categories of samples and a clear distinction among CD patients and H and UC subjects.

Further studies need to be performed using this screening and diagnostic serum assay on a larger cohort of serum samples from H subjects and patients affected by CD and other inflammatory diseases to confirm the specificity of these clusters. In order to develop a serum diagnostic assay able to reduce or delete the need of invasive procedures for the diagnosis of Crohn's disease a comparison of the results achievable by this test and by the already available serum antibody tests will be performed too.

Since the expression levels and activity of such postulated exoproteases seem to be related with the inflammatory level their identification and the evaluation of their biological role have to be performed. We expect to find exoprotease activity mostly inside the microparticles derived from platelet activation (PDMPs). An appropriate preanalytical step will be necessary to perform this subproteomics approach in which these microparticles will be separated with non invasive methods allowing to keep unaltered their structure and the enzymatic activity within them.

## Acknowledgments

The authors kindly thank Prof. Enrico Roda and Dr. Andrea Belluzzi from the Gastroenterology Unit of S.Orsola University Hospital of Bologna (Italy) for their helpful support.

This research was facilitated by access to the CIRB-CRBA Proeotomic Facility Services (LaP, c/o St. Orsola University Hospital of Bologna, Italy).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2009.08.003.

## References

[1] A. Iliuk, J. Galan, W.A. Tao, Anal. Bioanal. Chem. 393 (2009) 503.
[2] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, B. Kuster, Anal. Bioanal. Chem. 389 (2007) 1017.
[3] L.A. Liotta, E.F. Petricoin, Clin. Chem. 54 (2008) 3.
[4] P. Nanni, D. Parisi, G. Roda, M. Casale, A. Belluzzi, E. Roda, L. Mayer, A. Roda, Rapid Commun. Mass Spectrom. 21 (2007) 4142.
[5] P. Nanni, L. Mezzanotte, G. Roda, A. Caponi, F. Levander, P. James, A. Roda, J. Proteomics 72 (2009) 865.
[6] W. Simon, S.W. Beaven, M.T. Abreu, Curr. Opin. Gastroenterol. 20 (2004) 318.
[7] M. Papp, G.L. Norman, I. Altorjay, P.L. Lakatos, World J. Gastroenterol. 13 (2007) 2028.
[8] L.A. Liotta, E.F. Petricoin, J. Clin. Invest. 116 (2006) 26.
[9] L.A. Liotta, M. Ferrari, E. Petricoin, Nature 425 (2003) 905.
[10] M.A. Meuwis, M. Fillet, P. Geurts, D. de Seny, L. Lutteri, J.P. Chapelle, V. Bours, L. Wehenkel, J. Belaiche, M. Malaise, E. Louis, M.P. Merville, Biochem. Pharmacol. 73 (2007) 1422.
[11] L.J. Dekker, J. Bosman, P.C. Burgers, A. van Rijswijk, R. Freije, T. Luider, R. Bischoff, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 847 (2007) 65.
[12] M. Gao, J. Zhang, C. Deng, P. Yang, X. Zhang, J. Proteome Res. 5 (2006) 2853.
[13] N. Zolotarjova, J. Martosella, G. Nicol, J. Bailey, B.E. Boyes, W.C. Barrett, Proteomics 13 (2005) 3304.
[14] J. Brand, T. Haslberger, W. Zolg, G. Pestlin, S. Palme, Proteomics 6 (2006) 3236.
[15] T. Liu, W.J. Qian, H.M. Mottaz, M.A. Gritsenko, A.D. Norbeck, R.J. Moore, S.O. Purvine, D.G. Camp 2nd, R.D. Smith, Mol. Cell. Proteomics 5 (2006) 2167.
[16] M. Bellew, M. Coram, M. Fitzgibbon, M. Igra, T. Randolph, P. Wang, D. May, J. Eng, R. Fang, C. Lin, J. Chen, D. Goodlett, J. Whiteaker, A. Paulovich, M. McIntosh, Bioinformatics 22 (2006) (1902), Free available at http://proteomics.fhcrc.org.
[17] J. Häkkinen, G. Vincic, O. Månsson, K. Wårell, F. Levander, J. Proteome Res. 8 (2009) 3037, http://www.proteios.org.
[18] M.S. Geier, D. Tenikoff, R. Yazbeck, G.W. McCaughan, C.A. Abbott, J. Cell Physiol. 204 (2005) 687.
[19] A. Andoh, T. Tsujikawa, K. Hata, Y. Araki, K. Kitoh, M. Sasaki, T. Yoshida, Y. Fujiyama, Am. J. Gastroenterol. 100 (2005) 2042.
[20] W.R. Best, J.M. Becktel, J.W. Singleton, F. Kern Jr., Gastroenterology 70 (1976) 439.
[21] J. Villanueva, J. Philip, C.A. Chaparro, Y. Li, R. Toledo-Crow, L. DeNoyer, M. Fleisher, J Proteome Res. 4 (2005) 1060.
[22] N.I. Govorukhinaa, M. de Vriesb, H.T. Reijmersc, P. Horvatovicha, G.J. Ate, A.G. van der Zeed, R. Bischoff, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 877 (2009) 1281.
[23] X. Zheng, H. Baker, W.S. Hancock, J. Chromatogr. A 1120 (2006) 173.
[24] H.L. Huang, T. Stasyk, S. Morandell, M. Mogg, M. Schreiber, I. Feuerstein, C.W. Huck, G. Stecher, G.K. Bonn, L.A. Huber, Electrophoresis 26 (2005) 2843.
[25] R.S. Tirumalai, K.C. Chan, D.A. Prieto, H.J. Issaq, T.P. Conrads, T.D. Veenstra, Mol. Cell. Proteomics 2 (2003) 1096.
[26] J. Villanueva, D.R. Shaffer, J. Philip, C.A. Chaparro, H. Erdjument-Bromage, A.B. Olshen, M. Fleisher, H. Lilja, E. Brogi, J. Boyd, M. Sanchez-Carbayo, E.C. Holland, C. Cordon-Cardo, H.I. Scher, P. Tempst, J. Clin. Invest. 116 (2006) 271.
[27] J. Villanueva, A. Nazarian, K. Lawlor, S.S. Yi, R.J. Robbins, P. Tempst, Mol. Cell. Proteomics 7 (2008) 509.
[28] M. Forina, S. Lanteri, C. Armanino, M.C.C. Oliveros, C. Casolino, V-PARVUS (2004), Freely available at http://www.parvus.unige.it.
[29] A.I. Nesvizhskii, R. Aebersold, Mol. Cell. Proteomics 4 (2005) 1419.
[30] S. Danese, A. Papa, S. Saibeni, A. Repici, A. Malesci, M. Vecchi, Am. J. Gastroenterol. 102 (2007) 174.
[31] T. Vowinkel, M. Mori, C.F. Krieglstein, J. Russell, F. Saijo, S. Bharwani, R.H. Turnage, W.S. Davidson, P. Tso, D.N. Granger, T.J. Kalogeris, J. Clin. Invest. 114 (2004) 260.
[32] U.C. Broedl, V. Schachinger, A. Lingenhel, M. Lehrke, R. Stark, F. Seibold, B. Göke, F. Kronenberg, K.G. Parhofer, A. Konrad-Zerna, Inflamm. Bowel Dis. 13 (2007) 391.
[33] L. Jofre-Monseny, A. Loboda, A.E. Wagner, P. Huebbe, C. Boesch-Saadatmandi, A. Jozkowicz, A.M. Minihane, J. Dulak, G. Rimbach, Biochem. Biophys. Res. Commun. 357 (2007) 319.