Discovery of serum proteomic biomarkers for prediction of response to infliximab (a monoclonal anti-TNF antibody) treatment in rheumatoid arthritis: an exploratory analysis

Ortea, Ignacio; Roschitzki, Bernd; Ovalles, Juan Gabriel; Longo, Javier López; de la Torre, Inmaculada; González, Isidoro; Gómez-Reino, Juan J; González, Antonio

Abstract: Biologics such as TNF antagonists are a new class of drugs that have greatly improved Rheumatoid Arthritis (RA) treatment. However, for unknown reasons, individual patients with RA respond to one of these drugs but not to others even those targeting the same molecule. Methods to predict response are sorely needed because these drugs are currently selected by trial and error, what is very inefficient and prejudicial for the patient and the healthcare system. Here, we have explored the discovery of protein biomarkers in serum from patients treated with infliximab, one of the major anti-TNF drugs. The study was based in a quantitative proteomics approach using 8-plex iTRAQ labeling. It combined depletion of the most abundant serum proteins, two-dimensional LC fractionation, protein identification and relative quantification with a hybrid Orbitrap mass spectrometer. This approach allowed the identification of 315 proteins of which 237 were confidently quantified with two or more peptides. The detection range covered up to 6 orders of magnitude including multiple proteins at the ng/mL level. A new set of putative biomarkers was identified comprising 14 proteins significantly more abundant in the non-responder patients. The differential proteins were enriched in apolipoproteins, components of the complement system and acute phase reactants. These results show the feasibility of this approach and provide a set of candidates for validation as biomarkers for the classification of RA patients before the beginning of treatment, so that anticipated non-responders could be treated with an alternative drug.

DOI: https://doi.org/10.1016/j.jprot.2012.09.011

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-90550
Accepted Version

Originally published at:
DOI: https://doi.org/10.1016/j.jprot.2012.09.011
Discovery of serum proteomic biomarkers for prediction of response to infliximab (a monoclonal anti-TNF antibody) treatment in Rheumatoid Arthritis: an exploratory analysis

Ignacio Ortea\textsuperscript{1}, Bernd Roschitzki\textsuperscript{2}, Juan Gabriel Ovalles\textsuperscript{3}, Javier López Longo\textsuperscript{3}, Inmaculada de la Torre\textsuperscript{3}, Isidoro González\textsuperscript{4}, Juan J. Gómez-Reino\textsuperscript{1,5}, Antonio González\textsuperscript{1}

\textsuperscript{1}.- Laboratorio Investigacion 10 and Rheumatology Unit. Instituto de Investigacion Sanitaria - Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain.
\textsuperscript{2}.- Functional Genomics Center Zurich, University and ETH Zurich, Switzerland.
\textsuperscript{3}.- Rheumatology Unit. Hospital Gregorio Marañon, Madrid, Spain.
\textsuperscript{4}.- Rheumatology Unit. Hospital de la Princesa, Madrid, Spain.
\textsuperscript{5}.- Department of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain.

Corresponding author:
Ignacio Ortea
e-mail: nachoog@iim.csic.es
Laboratorio Investigacion 10 and Rheumatology Unit. Instituto de Investigacion Sanitaria - Hospital Clínico Universitario de Santiago, c/Choupana, ES15703, Santiago de Compostela, Spain.
Tel.:+34 981 955 090; fax: +34 981 950 538
Abstract

Biologics such as TNF antagonists are a new class of drugs that have greatly improved Rheumatoid Arthritis (RA) treatment. However, for unknown reasons, individual patients with RA respond to one of these drugs but not to others even those targeting the same molecule. Methods to predict response are sorely needed because these drugs are currently selected by trial and error, what is very inefficient and prejudicial for the patient and the healthcare system. Here, we have explored the discovery of protein biomarkers in serum from patients treated with infliximab, one of the major anti-TNF drugs. The study was based in a quantitative proteomics approach using 8-plex iTRAQ labeling. It combined depletion of the most abundant serum proteins, two-dimensional LC fractionation, protein identification and relative quantification with a hybrid Orbitrap mass spectrometer. This approach allowed the identification of 315 proteins of which 237 were confidently quantified with two or more peptides. The detection range covered up to 6 orders of magnitude including multiple proteins at the ng/mL level. A new set of putative biomarkers was identified comprising 14 proteins significantly more abundant in the non-responder patients. The differential proteins were enriched in apolipoproteins, components of the complement system and acute phase reactants. These results show the feasibility of this approach and provide a set of candidates for validation as biomarkers for the classification of RA patients before the beginning of treatment, so that anticipated non-responders could be treated with an alternative drug.

Keywords: rheumatoid arthritis, infliximab, serum biomarkers, response prediction.

Running title: Biomarkers for prediction of response to infliximab in RA.
Introduction

Rheumatoid arthritis (RA) is a chronic systemic autoimmune disease of complex etiology comprising genetic and environmental factors that is characterized by inflammation in multiple joints [1]. Left without treatment, it progresses to disability, deformities due to bone erosion and life shortening. RA prevalence is about 1% of the world population. Classical treatments are still commonly used, but they are not sufficiently effective for many patients. In the last decade, new drugs became available in the group of biologics (monoclonal antibodies, soluble receptors or other complex molecules targeting specific players in the disease process). The first that were available for RA treatment were the tumor necrosis factor (TNF) antagonists. Among them, infliximab, a chimeric antibody comprising a human IgG1 constant fraction and a murine variable region targeting membrane and soluble TNF [2], has become one of the biologics most commonly used in RA.

Biologics have greatly improved RA treatment but none of them is effective in all patients. For unknown reasons, about a third of the patients in whom one of these drugs is assayed fail to show significant improvement. These patients can respond to an alternative biologic targeting the same or a different molecule [3]. Currently, clinical or laboratory methods for the prediction of patients response are not available. Therefore, the only approach to select biologics for a particular RA patient is by trial and error. This approach is associated with notable inefficiency and prejudices because responsiveness can only be assessed after three to six months of treatment. During this time, patients suffer uncontrolled disease with the potential of irreversible damage, and the healthcare system expends large amounts in ineffective drugs. Thus, it is necessary to find biomarkers that make possible the identification of non-responder patients in advance, to treat them with an alternative drug from the beginning. Many studies have already tried to identify this type of biomarkers in the genetics, functional genomics, proteomics, autoantibody and clinical fields, but no reproducible and informative findings have yet been reported [4].

Very few proteomic studies have attempted to identify biomarkers for prediction of response to biologics in RA. A couple of studies analyzed selected cytokines or cytokines plus RA autoantibodies showing that some of them were associated with clinical response to the TNF antagonist etanercept [5,6]. By contrast, not a single cytokine was associated with response to a different biologic, rituximab, in a similar
The unique previous agnostic proteomic study was done by Trocmé et al. [8]. These authors used SELDI-TOF technology to identify plasma biomarkers for prediction of response to infliximab. Six potential biomarkers were detected, although only two proteins were identified. None of the previous studies have been independently replicated. This paucity of proteomic research on biomarkers for prediction of response contrasts with the multiple studies searching RA diagnostic and prognostic biomarkers [9-20], and with several proteomic studies monitoring the changes that take place after the administration of different drugs [7,16,21,22].

Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) is a quantitative proteomic approach ideally suited for biomarker discovery. It provides quantification, identification and multiplexing in a single assay. However, it has been scarcely used to study human serum and plasma. This is unfortunate because serum and plasma are informative for many diseases, especially for systemic diseases like RA, and easily available. The latter is very important when validating the potential biomarkers and also for their future widespread use. For example, patients with RA are not routinely subjected to synovial tissue biopsies and it will pose significant difficulties to implement them for drug selection. Unfortunately, discovery of biomarkers among the serum or plasma proteins is limited by their great complexity and wide dynamic range. Protein concentrations extend for more than 11 orders of magnitude with the top 10 most abundant plasma proteins accounting for ~ 90% of the total proteins [23]. Disease biomarkers are usually present at low concentrations (~ ng/mL) [23], being masked by higher abundance proteins in 2-DE and being blurred in MS due to competitive ionization and signal suppression. Therefore, the quantitative and qualitative analysis of low abundance proteins is challenging. To overcome these problems, there is a need for (i) prefractionation methods to specifically remove the high abundance proteins; (ii) good separation techniques to further decrease protein complexity; and (iii) MS equipments with high sequencing speed and sensitivity. In this study, we have explored the performance of an approach including these characteristics. It was applied to the discovery of biomarkers for prediction of response in serum of patients with RA that had been prospectively evaluated during treatment with infliximab. The proteomics approach comprised an immunodepletion prefractionation step, a thorough 2-D LC fractionation and quantification of the differentially isotopic labeled peptides by MS/MS. Results were satisfactory because this approach allowed the identification of a large number of proteins, covering a wide dynamic range and including many proteins.
in the ng/mL level, and identifying 14 putative biomarkers for prediction of response to infliximab that are consistent with our knowledge of the disease.

**Materials and Methods**

**Ethics approval**

The project was approved by the Ethics Committee for Clinical Research of Galicia and carried out according to the Helsinki Declaration Principles. All participating subjects gave their written informed consent.

**Sample collection**

Patients with RA according to the American College of Rheumatology (ACR) classification criteria [24] from a single center (Gregorio Marañón Hospital, Madrid, Spain) were enrolled in the study. All were naive for any biologics before the start of the enrollment period. Blood was collected into 8 ml Vacuette Z Serum Sep Clot Activator tubes (Greiner Bio-One GmbH, Frickenhausen, Germany) before starting infliximab administration, left to clot at room temperature for 2 hours and then centrifuged at 3000 g for 10 min. The collected serum was aliquoted and stored at -80°C. Infliximab (Remicade; Centocor Inc., Malvern, PA) was given following the standard dose and administration schedule. Clinical response was determined 6 months after infliximab initiation according with the European League Against Rheumatism (EULAR) criteria based in the Disease Activity Score 28 (DAS28) [25]. Only patients classified as non-responder (NR, n = 4) or good responder (R, n = 4) were compared to increase the chances of finding differences. The intermediate class of moderate responders was excluded.

**Immunoaffinity depletion of high-abundance proteins**

The six most abundant proteins in serum were depleted using the Hu-6 Multiple Affinity Removal System kit (Agilent Technologies, Wilmington, USA) following manufacturer’s instructions. Afterwards, the remaining proteins were concentrated using 5000 MWCO spin concentrators (Agilent Technologies) and acetone precipitation.
resulting air-dried pellets were dissolved in 0.5 M triethylammonium bicarbonate (TEAB) pH 8.5 buffer and protein concentration was measured.

**Protein digestion and iTRAQ labeling**

For each sample, 40 µg of protein, at 1 mg/mL in 0.5 M TEAB pH 8.5 and 0.1% SDS, were reduced with 5 mM tris-(2-carboxyethyl) phosphine (TCEP) at 60°C for 1 hour and cysteine-blocked with 10 mM methyl mehanethiosulfonate (MMTS) at RT for 10 min. The proteins were then digested with Sequencing Grade Modified trypsin (Promega, Madison, WI, USA) at a trypsin-to-protein ratio of 1:50, at 37°C for 6 hours. Each digest was labeled at 25°C for 3 h with one of the 8-plex iTRAQ reagents previously solubilized in 100 µl isopropanol, according to the manufacturer’s instruction (AB Sciex, Framingham, MA, USA). The labeling reactions were stopped by adding phosphoric acid to reach pH below 4.0, and all iTRAQ-labeled samples were combined into one tube. Tryptic peptides were then dried by centrifugal evaporation.

**Peptide fractionation with SCX chromatography**

Peptides were fractionated using a PolyLC SCX Polysulphoethyl A 200 mm x 2.1 mm, 5 µm, 200 A column (PolyLC, Columbia, MD, USA), on a high-pressure LC pump (1100-series, Agilent Technologies). Dried peptides were reconstituted in 600 µl buffer A, the pH was adjusted to 2.7 with H₃PO₄. The flow rate was kept at 0.2 ml/min, and the sample was fractionated using a two-buffer system (buffer A, 7mM KH₂PO₄ in 25% ACN pH 2.7; buffer B 7mM KH₂PO₄, 500 mM KCl in 25% ACN pH 2.7). The gradient employed was 0% B for 20 min, 0% to 5% B in 5 min, 5% to 35% B in 35 min, 35% to 100% B in 10 min, 100% B for 5 min, and then 0% B for 25 min. Eluted peptides were monitored at 214 and 280 nm and collected from 20-25.7 min, 25.7-31.3 min, 31.3-37.0 min, then every 2.4 min until 95 min. A total of 27 fractions were collected and then dried. An aliquot of 0.5 µl of each fraction was desalted by C18 ZipTip tips (Millipore, Billerica, MA, USA) and analyzed by MALDI-TOF-TOF (4800, AB Sciex) MS to check its peptide complexity. The first three fractions were discarded, and the other 24 fractions were mixed in 12 final fractions according to their complexity.

**Nano-reverse-phase LC-MS/MS**
Combined fractions were desalted onto C18 spin tips (StageTips, Thermo Fisher Scientific, Waltham, MA, USA) lyophilized and dissolved on buffer A. Analysis by nanoHPLC-MS/MS was done using a NanoLC-Ultra system (Eksigent, Dublin, CA, USA) coupled to an Orbitrap Velos hybrid mass spectrometer (Thermo-Finnigan, San Jose, CA, USA). The separation was performed on a inhouse-made tip column (75 µm id x 8 cm) packed with Magic RP C18 AQ, 200A, 3 µm beads (Bischoff GmbH, Leonberg, Germany), at a flow rate of 250 nl/min. Water with 1% ACN and 100% ACN, both containing 0.2% formic acid, were used as solvents A and B, respectively. Peptides were loaded on the column for 16 min with 2% of solvent B and a flow rate of 500 nl/min. Peptide elution was started using the following gradient of solvent B: 0 to 1 min 2% to 10% B, 1 to 65 min 10% to 30% B, 65 to 70 min 30% to 45% B, 70 to 74 min 45% to 97% B. The column was washed for 7 min with 97% B and then re-equilibrated for 9 min with 2% solvent B before the next run. The mass spectrometer was operated in data-dependent mode with the following ion scanning parameters: survey MS scan in FT mode from 300 to 2000 m/z (resolution 30000), followed by top eight peaks collision induced dissociation (CID) fragmentation (isolation width 2 m/z, normalized collision energy 35%) for identification, plus the same top eight peaks higher-energy collisional dissociation (HCD) fragmentation (isolation width 2 m/z, normalized collision energy 52%) and readout in the FT analyzer (resolution 7500) for quantification. Fragmented peptide masses were set in dynamic exclusion for 60 s and singly charged ions were excluded from MS/MS analysis. To improve sensitivity of the MS/MS analysis for peptides of low-abundance proteins, each fraction was run a second time excluding previously fragmented precursors.

Protein identification and protein relative abundance

Peptide and protein identification was performed with ProteinPilot software v4.0 (AB Sciex) and the Paragon algorithm [26] by comparison with the human Swissprot/TrEMBL database (downloaded in December 2011). Paragon method parameters were: peptide labeled with iTRAQ 8plex, fixed modification of methylmethanethiosulfonate on Cys (+46 Da), digestion with trypsin, instrument Orbi/FT MS (1-3ppm) LTQ MS/MS, and ID focus on biological modification. Proteins having at least one peptide above the 95% confidence level as determined by Protein Pilot were
recorded. False discovery rates (FDR) were estimated using a concatenated target-decoy database [27]. For the estimation of the protein abundance ratio, the intensities of iTRAQ reporter ions for each MS/MS spectra were extracted from ProteinPilot and the sum ratio for each protein was calculated across the spectra matched to the corresponding peptides. Data were normalized for loading error by bias corrections using ProteinPilot. The statistical significance of the differences between the means for each group (R vs. NR) was determined on the transformed data (arc sin hyperbolic) using the two–tailed t-test. Threshold for significance was set at p < 0.05. A receiver operation characteristic (ROC) curve analysis was performed for each protein to obtain the area under curve (AUC). AUC was used as the summary statistic reflecting the overall predictive accuracy of each protein [28]. A model or test with perfect discriminatory ability will have an AUC of 1.0, while a model of random outcomes will have an AUC of 0.5. Interactions and pathways of proteins with fold differences between NR and R patients higher than 1.5 were analyzed with Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA, USA).

Results and Discussion

Figure 1 shows the discovery-driven 8-plex iTRAQ workflow used in this study. This workflow combines depletion of the highest abundance serum proteins, intensive fractionation of the depleted serum, and MS/MS based identification and quantification of the fractionated proteins. In the first step, sera were immunodepleted from the six highest abundance proteins on a human MARS-6 spin column. Reproducibility and protein recovery of this step were evaluated by 1-DE (Figure 2) and by total protein quantification. These analyses revealed uniform reduction of major protein bands and compensatory increases in other bands of lower concentration together with significant reduction of the total protein concentration. The mean protein recovery rate was 11.5%, which is in agreement with the column’s manufacturer specifications. Subsequent MS analysis showed that this depletion step was effective because our protocol allowed the identification with strict criteria (at least 2 peptides and 95% confidence) of medium and low concentration proteins. Examples of identified proteins that are in the ng/mL range according with the bibliography were hepatocyte growth factor activator, sex hormone-binding globulin, alpha synuclein and retinoic acid receptor responder protein [29-32]. However, our protocol was not able to identify very rare proteins in the pg/mL range.
range. Since concentrations of the most abundant identified proteins are in the mg/mL level, the workflow used in this study allowed the detection of serum proteins with a dynamic range of up to $10^6$. A wide detection range is a necessary characteristic for effective proteomic analysis of serum given the complexity and heterogeneity of this biologic fluid.

Identification of serum proteins

The next steps involved trypsin digestion of the proteins, labeling of the resulting peptides with the iTRAQ reagents, pooling samples in a single mix and fractionation of the labeled peptides by SCX. The 27 initial fractions were collected and their peptide complexity was determined by MALDI-TOF MS. Low complexity fractions were pooled, given a total of 12 final fractions. These 12 fractions were subjected to nanoHPLC-MS/MS in an Orbitrap Velos hybrid mass spectrometer. Spectra were analyzed with ProteinPilot leading to the identification of 247 proteins with at least 2 peptides (315 proteins with a single peptide) (Table S-1 in the Supplementary Material). An experiment with 300 µg of depleted serum treated in the same conditions except for iTRAQ labeling, led to the identification of 235 proteins (results not shown). This result indicates that iTRAQ labeling with the analytical conditions used in this study does not substantially modify peptide ionization efficiency and protein identification. This finding is relevant because the effect of iTRAQ labeling seems to depend on the analytical system used. For example, iTRAQ labeling increases the number of identified proteins when using MALDI-TOF-TOF [33], but decreases the number with pulsed-Q dissociation (PQD) on a dual-pressure linear ion trap LTQ Velos [34] or with CID on a non-dual-pressure linear ion trap [35]. CID shows better performance for identification on linear ion traps, so quantification methods that use isobaric tags usually combine CID fragmentation for identification with PQD or HCD scans for quantification [36-38]. We have achieved similar protein identification for 8-plex labeled samples than for label-free samples, showing that the combination of dual-pressure LIT CID for identification with HCD for quantification, as implemented in the Orbitrap Velos hybrid mass spectrometer, is advantageous for isobaric tag quantification workflows.
Gene ontology analysis of the identified proteins showed that most of them had an extracellular location (Figure 3a), as expected. The main activities were defined by functional classification as binding, catalytic and enzyme regulator (Figure 3b).

**Quantitative analysis**

The iTRAQ label ratios were used for relative quantification of 289 proteins, 237 of them with at least two peptides (Table S-2 in the Supplementary Material). Comparison of protein abundance between responder and non-responder patients was used to identify potential biomarkers for prediction of response to infliximab. Statistical significance rather than fold difference is preferred for selection of proteins showing differential abundance [39]. Therefore, we have used p-values < 0.05 from t-test comparisons between the two groups of patients to discover 14 differential proteins (Table 1 and Figure S-1 in the Supplementary Material). As expected, the AUC values derived from ROC analyses were concordant with the p-values from the t-tests: with the most discriminant proteins being identified by the two analyses (Table 1 and Table S-2 in the Supplementary Material). Each value of AUC can be interpreted as the probability that the biomarker will rank a randomly chosen non-responder higher than a randomly chosen responder [28]. The AUC value for a perfect test is 1.0, whereas the value for a completely random test is 0.5. It is commonly accepted that tests with AUC below 0.75 are unlikely to have interest for clinical use, whereas those with values over 0.75 could be of utility. All the differential proteins in our study were in this later group with high predictive potential (Table 1) and compare favourably with those previously identified for prediction of response to treatment in RA [8], with other predictors of response [40-42] and with other predictors in RA research [43-45]. All the differential proteins, except for adipocyte plasma membrane-associated protein (APMAP), have been previously related to RA or to the inflammatory response. These relationships with inflammation and the disease process could explain that all of them were more abundant in the non-responder subset of patients than in the responder group. In addition, as the serum samples were obtained before the start of infliximab treatment, all of them came from patients with active inflammation because this is a requirement for the prescription of this drug. The fold difference values should be interpreted considering that iTRAQ quantification underestimates the abundance differences [46].
and therefore some of the results may be possibly more disparate between the two groups of patients when assessed with other techniques.

Results of pathway analysis showed that the differential proteins were significantly enriched in the inflammatory response and protein synthesis networks (scores 46 and 41, respectively). These results highlight a potential role for apolipoproteins and complement factors in the differential response to infliximab (Figure 4). Among the apolipoproteins, three were significantly more abundant in non-responder patients: apoB-100, apoA-II and apoM. In addition, apoA-I, apoC-I and apoC-II were near the significance threshold ($P$ values < 0.08). Proteins of this family bind lipids to form lipoprotein particles and transport these lipids to tissues. ApoB-100 is the hepatic isoform of apoB. It is the main protein component of low-density lipoproteins (LDL) that carry most of the cholesterol that is transported to the tissues. ApoB is the recognition signal for the cellular binding and internalization of LDL, and it is a very good biomarker for cardiovascular risk [47]. ApoA-I and apoA-II are the main apolipoproteins of high-density lipoproteins (HDL). ApoA-I is associated with protection from cardiovascular risk because it promotes cholesterol efflux from tissues to the liver for excretion. It has been much more extensively studied than apoA-II, which seems to play a crucial role in triglyceride catabolism [48]. ApoM is also a HDL-associated apolipoprotein with anti-atherogenic potential probably through an antioxidant effect [49]. ApoC-I is a component of HDL and is also associated with triglycerides and cholesterol in very low density lipoproteins (VLDL). ApoC-II is also a component of VLDL, and it is incorporated to these particles once they are circulating in the blood. The differences we have found in apolipoproteins should be considered in the context of the complex interactions between chronic inflammation and lipid metabolism in RA that we still do not completely understand. An increased mortality of RA patients due to cardiovascular events has been reported [50]. This increase is partly due to the adverse serum lipid profile found in untreated patients, low levels of HDL cholesterol and increased LDL/HDL cholesterol and apoB/apoA-I ratios [51,52]. This adverse lipid profile could be partially explained by changes in the expression of genes in the liver or in the reticuloendothelial system as part of the acute phase response. In addition, inflammation induces changes in the composition of the HDL and LDL particles and in the clearance of cholesterol [52,53]. The lipid profile improves during effective treatment, probably as consequence of the control of inflammation [54,55]. Together with these effects of inflammation on apolipoproteins there are also effects of
apolipoproteins on inflammation. The most studied effects have been those of apoA-I, which is anti-inflammatory [56], and apoB, which is proinflammatory [57], but it has been also reported that apoA II has an anti-inflammatory role [58] and apoC-I a proinflammatory one [59]. These contrasting effects pose a problem for a simple interpretation of our results because all the differential apolipoproteins showed higher levels in the non-responder group than in the responder one. However, recent evidence indicates that the known roles of apolipoproteins are reversed in a subset of RA patients that have proinflammatory HDL particles in place of anti-inflammatory HDL, due to changes in composition [60]. This change of properties has been shown to be a consequence of the acute phase response [61] and might mean that all the observed differences in our study are directed towards a more damaging and proinflammatory lipid profile.

Another protein related with lipid metabolism in our significant findings is APMAP. This protein is a transmembrane protein necessary for adipocyte differentiation that is increased in obesity models [62], although no relationship with inflammation or RA have been described.

Regarding the other overrepresented functional pathway, there were five complement-related proteins among the proteins showing significant differences: C4B-alpha chain, complement factor H-related protein 4 (CFHR4), mannan-binding lectin serine protease 2 (MASP2), and inter alpha trypsin inhibitor heavy chain H1 (ITIH1) and H2 (ITIH2). Other four proteins in this pathway showed fold differences over 1.5 but were not significantly different: C8 beta, C8 alpha, C5 and complement factor H-related protein 3 (CFHR3). C4B is one of the two isotypes of C4, each of them encoded by a different gene. Upon secretion, the C4 molecules are cleaved in three chains that remain together as a trimer. The classic and the lectin complement activation pathways include cleavage of the C4 alpha chain, which we have found different, into the C4a anaphylotoxin and C4b, which continues with the activation cascade. The different complement activation pathways converge in C3, whose activation is followed by the cleavage of C5 in C5a, another anaphylotoxin, and C5b, which initiates the membrane attack complex. This complex is the effector cytolytic endproduct of the complement system and includes, among others, complement factors C5b and C8. The latter is made of three subunits encoded in separated genes: C8 alpha, beta and gamma. CFHR4 is a member of the CFH family that has recently been identified as promoting complement activation via the alternative and classic pathways [63]. CFHR3 is a member of the
same family that seems to be involved in regulation of the complement system by
inhibiting the activation of C3 [64]. MASP2, in turn, is a serum protease that activates
the lectin pathway via the cleavage of C4 and C2 [65]. Finally, the inter alpha trypsin
inhibitor (IαI) complex is made of two heavy chains ITIH1 and ITIH2 plus bikunin, and
it is a broad spectrum proteinase inhibitor. Among its many targets, it inhibits the early
phases of complement activation by the three pathways: classic, alternative and lectin
[66]. All these differential proteins related to complement can have a role in RA, since
the complement system is activated in the inflamed joints of patients with RA. This
activation state has been shown by the relative consumption of C3 and C4 in the
synovial fluid together with increased concentrations of C5a and the membrane attack
complex [67]. The higher abundance of all these proteins in the non-responder patients
makes it tempting to propose that this subset of patients is characterized by more
complement activation. However, this interpretation should be very cautious because
regulation of the complement system is mainly done at the activation levels and not at
the transcription level, and because ITIH1 and ITIH2 are inhibitors of complement
activation. An additional element to consider is that many components of the
complement system are acute phase reactants and they could be increased as a
consequence of inflammation [68].

Other differential proteins in our results were also part of the acute phase
response: ceruloplasmin (CP), thrombospondin (THBS1), vitamin D-binding protein
(GC) and fibronectin (FN1). CP is a ferroxidase enzyme involved in iron transport and
metabolism that has been proposed as a link between iron metabolism and the immune
system [69]. THBS1 is a multifunctional protein that mediates cell-to-cell and cell-to-
matrix interactions. It has been related to RA pathogenesis via the activation of
transforming growth factor beta (TGFβ), and more generally to acute and chronic
inflammation through a variety of mechanisms [70]. GC is involved in the transport of
vitamin D and its metabolites to target tissues [71]. It is also the precursor of vitamin D-
binding protein-derived macrophage-activating factor (GcMAF) that results from
enzymatic modifications taking place in B and T cells [72]. GcMAF is a potent activator
of monocyte and macrophage functions that has not yet been specifically studied in RA.
Finally, FN1 is one of the most abundant proteins in inflamed joints of patients with RA
[73], and it is also related with RA as citrullinated FN1. This posttranscriptional
modification and antibodies against it are present in synovial tissue, synovial fluid and
sera of RA patients [74]. In this regard, it is interesting to note that protein citrullination
and anti-citrullinated protein antibodies play a central role in the pathophysiology of RA [75]. Most of the proteins in this group of acute phase reactants have been found at elevated levels in serum of patients with RA [15,73,76]. However, it is uncertain if their higher level in the non-responder patients of our study could be ascribed to a more severe inflammation. The reasons for this doubt are that clinical studies have not identified any sign of inflammation as predictive of response to infliximab [77], and that other prominent acute phase reactants, serum amyloid A protein and C-reactive protein, showed a non-significant excess in the responder group (fold differences 0.8 and 0.65, respectively). Therefore, we cannot exclude the possibility that the higher abundance of these proteins is due to specific mechanisms unrelated with the intensity of inflammation.

Finally, gelsolin (GSN) is a protein that does not fit in any of the previous groups. In blood, it contributes to limit inflammatory responses by acting as a scavenger binding the actin released by tissue damage, bioactive lipids and proinflammatory mediators [78]. It is decreased in multiple acute and chronic inflammatory diseases including RA [79]. Possible mechanisms of this decrease include clearance of the GSN complexes and trapping of GSN-actin into the inflamed joints in RA. As commented for the other groups of proteins, it is impossible to know at this stage what could be behind the higher levels of GSN found in the non-responder group of patients.

The only previous study with similar aims has been already mentioned [8]. It identified six potential proteins that were different between the non-responder and responder RA patients treated with infliximab. These proteins showed AUC ranging from 0.761 to 0.846, but not statistical tests or fold differences were reported. Five of them were more abundant in the non-responder group and only one in the responder patients, which is a direction of differences similar to our findings. Four of the proteins were not further characterized because the authors used SELDI-TOF-MS and needed purification for identification. The two identified proteins were also present in our analysis. Platelet factor 4 was more abundant in the non-responder patients in both studies (fold difference NR/R = 1.39; AUC = 0.81 in our samples), but the difference was not significant in our analysis (P = 0.2). It has been also highlighted as a biomarker for prediction of response to infliximab in a Crohn’s disease study showing also higher concentration in the non-reponder patients [80]. Therefore, platelet factor 4 seems a good candidate biomarker although it did not came in the top list of our study. The second protein, apoA-1 showed contrasting results. It was almost significantly more
abundant in the non-responder patients in our study, as already discussed, whereas it
was more abundant in responder patients than in non-responder ones in Trocmé et al.
[8]. This discordant result serves us to remind that all these results need to be replicated
in new studies because of the exploratory nature of the previous and the current studies.

Conclusions

This report is the first label-based quantitative proteomics study aimed at
discovering potential serum biomarkers for prediction of response to biologics in RA.
The results have been encouraging in several respects: the number of identified proteins,
the dynamic range of concentrations they covered, and the number of differential
proteins between the two groups compared. The panel of differential proteins needs
further verification and clinical validation, but already it has been possible to notice that
most of these proteins are known to be related with important processes in RA and
many of them are known to show altered levels or function related with RA or chronic
inflammation. Therefore, we expect that they will help establish approaches to classify
RA patients as responders and non-responders to infliximab before the beginning of the
treatment, in order to treat anticipated non-responders with an alternative drug.

Conflict of interest statement
All authors declare there are no financial/commercial conflicts of interest.

Acknowledgments
This work has been supported by PRIME-XS project (grant agreement number 262067)
funded by 7th Framework Program of the European Union, by grant 11/01048 and by
RETICS Program, RD08/0075 (RIER) of the Instituto de Salud Carlos III (Spain) that
are partially financed by the European Regional Development Fund of the European
Union.

References


[61] Van Lenten BJ, Hama SY, de Beer FC, Stafforini DM, McIntyre TM, Prescott SM, La Du BN, Fogelman AM, Navab M. Anti-inflammatory HDL becomes pro-inflammatory during the acute phase response. Loss of protective effect of HDL


Figure captions.

Figure 1. Discovery-driven experimental workflow of the study. R, responder; NR, non-responder; CID, collision-induced dissociation; HCD, higher-energy collisional dissociation.

Figure 2. Depletion of the 6 highest abundance proteins from serum. 10% NuPAGE Bis-Tris 1-DE gel image of the different protein fractions.

Figure 3. Gene ontology (GO) annotation of identified serum proteins according to (a) cellular location; and (b) molecular function.

Figure 4. Pathway analysis of proteins that showed differential abundance between responder and non-responder patients with RA treated with infliximab. Direct interactions are shown. Proteins with R/NR ratios > 1.5 are in green, and with NR/R ratios > 1.5 are in red. Merged two top score networks are shown (immune and inflammatory response, score 46, and protein synthesis, score 41), centered on low-density lipoprotein (LDL) and high-density lipoprotein (HDL) as reflected by the number of interactions.
Table 1. Serum proteins showing differential abundance between the responder and non-responder patients with RA treated with infliximab.

<table>
<thead>
<tr>
<th>Accession number</th>
<th>Protein name</th>
<th>Non-responder/Responder</th>
<th>p-val</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>D6RAK8</td>
<td>Vitamin D-binding protein splicing variant GC-006</td>
<td></td>
<td>5.359</td>
<td>0.0386</td>
</tr>
<tr>
<td>P00450</td>
<td>Ceruloplasmin</td>
<td></td>
<td>3.375</td>
<td>0.0178</td>
</tr>
<tr>
<td>P04114</td>
<td>Apolipoprotein B-100</td>
<td></td>
<td>2.198</td>
<td>0.0348</td>
</tr>
<tr>
<td></td>
<td>Inter-alpha-trypsin inhibitor heavy chain H2</td>
<td></td>
<td>2.072</td>
<td>0.0230</td>
</tr>
<tr>
<td>P07996</td>
<td>Thrombospondin-1</td>
<td></td>
<td>1.994</td>
<td>0.0319</td>
</tr>
<tr>
<td>Q6U2E9</td>
<td>Complement C4-B alpha chain</td>
<td></td>
<td>1.971</td>
<td>0.0315</td>
</tr>
<tr>
<td>P19827</td>
<td>Inter-alpha-trypsin inhibitor heavy chain H1</td>
<td></td>
<td>1.785</td>
<td>0.0388</td>
</tr>
<tr>
<td>P06396</td>
<td>Gelsolin</td>
<td></td>
<td>1.718</td>
<td>0.0328</td>
</tr>
<tr>
<td>P02652</td>
<td>Apolipoprotein A-II</td>
<td></td>
<td>1.718</td>
<td>0.0414</td>
</tr>
<tr>
<td>P02751-7</td>
<td>Fibronectin isoform 7</td>
<td></td>
<td>1.682</td>
<td>0.0340</td>
</tr>
<tr>
<td>C9J7J7</td>
<td>Complement factor H-related protein 4</td>
<td></td>
<td>1.680</td>
<td>0.0486</td>
</tr>
<tr>
<td>O95445</td>
<td>Apolipoprotein M</td>
<td></td>
<td>1.601</td>
<td>0.0325</td>
</tr>
<tr>
<td>Q9HDC9</td>
<td>Adipocyte plasma membrane-associated protein</td>
<td></td>
<td>1.508</td>
<td>0.0219</td>
</tr>
<tr>
<td>O00187</td>
<td>Mannan-binding lectin serine protease 2</td>
<td></td>
<td>1.336</td>
<td>0.0246</td>
</tr>
</tbody>
</table>