Specific serum protein biomarkers of rheumatoid arthritis detected by MALDI-TOF-MS combined with magnetic beads

Qian Niu¹, Zhuochun Huang¹, Yunying Shi², Lanlan Wang¹, Xiaofu Pan¹ and Chaojun Hu³

¹Department of Laboratory Medicine and ²Department of Nephrology, West China Hospital, Sichuan University, Chengdu, No. 37, Guo Xue Xiang, Chengdu 610041, The People’s Republic of China
³Department of Clinical Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100032, The People’s Republic of China

Correspondence to: L. Wang; E-mail: wang_lanlan@126.com

Received 10 October 2009, accepted 16 April 2010

Abstract

Objectives. To identify novel serum protein biomarkers and establish diagnostic pattern for rheumatoid arthritis (RA) by using proteomic technology. Methods. Serum proteomic spectra were generated by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) combined with weak cationic exchange magnetic beads. A training set of spectra, derived from analyzing sera from 22 patients with RA, 26 patients with other autoimmune diseases and 25 age- and sex-matched healthy volunteers, was used to train and develop a decision tree model with a machine learning algorithm called decision boosting. A blinded testing set, including 21 patients with RA, 24 patients with other autoimmune diseases and 25 healthy people, was used to examine the accuracy of the model. Results. A decision tree model was established, consisting of four potential protein biomarkers whose m/z values were 4966.88, 5065.3, 5636.97 and 7766.87, respectively. In validation test, the decision tree model could differentiate RA from other autoimmune diseases and healthy people with the sensitivity of 85.71% and specificity of 87.76%, respectively. Conclusions. The present data suggested that MALDI-TOF-MS combined with magnetic beads could screen and identify some novel serum protein biomarkers related to RA. The proteomic pattern based on the four candidate biomarkers is of value for laboratory diagnosis of RA.

Keywords: autoimmune diseases, MALDI-TOF-MS, proteomics, rheumatoid arthritis, weak cationic exchange magnetic beads

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory autoimmune disease with persistent inflammation of multiple synovial joints, which results in progressive and irreversible tissue destruction of bone and cartilage (1). Consequently, patients with RA commonly suffer from joint deformity as well as joint dysfunction. Accurate diagnosis and proper treatment are therefore crucial for enhancing curative effects and improving life quality of RA patients. For many years, laboratory diagnosis of RA has relied on the detection of rheumatoid factor (RF), one of autoantibodies. However, this diagnostic test performs poorly with moderate sensitivity and low specificity (2). Although a new assay that detects antibodies to citrullinated peptides, called the anti-CCP assay, has a much higher specificity (>97%) (3) than the RF test, its sensitivity (68%) is still not satisfactory. Anti-keratin antibodies detection that is also commonly performed in laboratory is of similar clinical value with anti-CCP assay, and simultaneously detecting these three antibodies can improve the diagnosis of RA. However, all these laboratory diagnostic methods are still not perfect enough, and identifying novel biomarkers with high specificity and sensitivity is necessary for improving the diagnosis of RA.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) is an important proteomic technology, which has been used for analyzing serum proteomic spectra. Based on this technology, many protein biomarkers of certain diseases have been discovered (4–6). Magnetic beads have large surface so that they can capture more small-molecular peptides and proteins when compared with conventional protein chips (7, 8). The combination of MALDI-TOF-MS and magnetic beads can take advantages of each other, and it can detect more proteins of low molecular weight.
In this study, we used MALDI-TOF-MS combined with weak cationic exchange (WCX) magnetic beads to detect serum proteome of 43 patients with RA, 50 patients with other autoimmune diseases [19 with systemic lupus erythematosus (SLE), 15 with Sjögren's syndrome (SS) and 16 with systemic scleroderma (SSc)] and 50 healthy volunteers. Based on these data, we aimed to discover some novel potential biomarkers for RA and then established a diagnostic model in order to improve diagnostic efficiency for RA.

Methods

Patients and controls

This study was approved by the Ethics Committee of the Chinese Human Genome and the Ethics Committee of West China Hospital, and written informed consent was obtained from all participants. A total of 143 serum samples were included in this study, among which, there were 43 from patients with RA, 50 from patients with other autoimmune diseases (19 with SLE, 15 with SS and 16 with SSc) as disease control (DC) and 50 from healthy volunteers as healthy control (HC). All samples were collected at West China Hospital of Sichuan University and Peking Union Medical College Hospital from April 2006 to December 2007. All the patients were diagnosed according to the latest diagnosis criteria of American Rheumatism Association. Demographic features of all the patients and healthy volunteers were provided in Table 1 and sample numbers used for profiling (training set) and validation (blinded testing set) were listed in Table 2.

Table 1. Demographics of RA patients and control subjects

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Number of samples</th>
<th>Male/female</th>
<th>Age (mean ± SD)</th>
<th>Age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA(^a)</td>
<td>43</td>
<td>10/33</td>
<td>48.0 ± 10.7</td>
<td>16–61</td>
</tr>
<tr>
<td>Control</td>
<td>100</td>
<td>17/83</td>
<td>39.5 ± 10.2</td>
<td>12–65</td>
</tr>
<tr>
<td>DCs</td>
<td>50</td>
<td>8/42</td>
<td>40.8 ± 13.3</td>
<td>11–59</td>
</tr>
<tr>
<td>SLE</td>
<td>19</td>
<td>4/15</td>
<td>35.5 ± 9.9</td>
<td>15–60</td>
</tr>
<tr>
<td>SS</td>
<td>15</td>
<td>2/13</td>
<td>43.2 ± 10.3</td>
<td>19–58</td>
</tr>
<tr>
<td>SSc</td>
<td>16</td>
<td>2/14</td>
<td>39.4 ± 10.6</td>
<td>10–58</td>
</tr>
<tr>
<td>Healthy controls</td>
<td>50</td>
<td>9/41</td>
<td>38.9 ± 11.6</td>
<td>21–68</td>
</tr>
</tbody>
</table>

\(^a\) Twenty-three active RA and 20 inactive RA included.

Table 2. Number of serum samples used in training set and blinded testing set

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Training set</th>
<th>Blinded testing set</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA patients</td>
<td>22</td>
<td>21</td>
<td>43</td>
</tr>
<tr>
<td>DCs</td>
<td>26</td>
<td>24</td>
<td>50</td>
</tr>
<tr>
<td>SLE</td>
<td>10</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>SS</td>
<td>9</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>SSc</td>
<td>7</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Healthy controls</td>
<td>25</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>70</td>
<td>143</td>
</tr>
</tbody>
</table>

Table 3. Storage time of each study group

<table>
<thead>
<tr>
<th>Group</th>
<th>Storage time (day, mean ± SD)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA patients</td>
<td>317 ± 105</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>DCs</td>
<td>282 ± 110</td>
<td></td>
</tr>
<tr>
<td>SLE</td>
<td>264 ± 131</td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>321 ± 98</td>
<td></td>
</tr>
<tr>
<td>SSc</td>
<td>296 ± 112</td>
<td></td>
</tr>
<tr>
<td>Healthy controls</td>
<td>338 ± 87</td>
<td></td>
</tr>
</tbody>
</table>
To minimize potential bias that would affect the accuracy of results, all procedures including sample pretreatment and MALDI-TOF-MS analysis were performed by a special technician and all samples from patients and control groups were processed in the same way. Analysis of all samples was completed within 5 days.

Statistical analysis

The spectra of 143 serum samples were separated by a stratified random sampling into ‘training set’ data (a total of 73, 22 RA, 10 SLE, 9 SS, 7 SSc and 25 healthy volunteer samples) and ‘blinded testing set’ data (a total of 70, 21 RA, 9 SLE, 6 SS, 9 SSc and 25 healthy volunteer samples). The training set data were used for developing a boosting decision classification, which could screen specific protein biomarkers and accordingly establish a diagnosis model of RA. The resulting boosting decision tree classification was then applied to the blinded testing set data for evaluating the accuracy of the decision tree model. CIPHERGEN ProteinChip Software 3.0.2, Biomarker Wizard Software 3.1.0 and Biomarker Patterns Software 5.0 (Ciphergen Company) were available supporting software provided by Ciphergen Company for data analysis.

All the results were expressed as ‘mean ± SD’ and P values <0.05 were considered statistically significant.

Results

Identification of serum protein spectra

Protein spectra of 143 serum samples were detected by MALDI-TOF-MS combined with WCX magnetic beads. The combination was particularly effective in resolving proteins and peptides of low molecular weight (Fig. 1).

Comparison of serum protein spectra between RA and control subjects (disease controls and healthy controls)

A total of 102 protein peaks were detected by MALDI-TOF-MS in the training set including 22 RA, 10 SLE, 9 SS, 7 SSc and 25 healthy volunteer samples. In the control subjects, 34 differential protein peaks, 9 were over-expressed and 25 were under-expressed. Furthermore, among the 34 differential protein peaks, 9 were over-expressed and 25 were down-regulated in RA subjects.

Screening of serum protein biomarkers of RA and construction of boosting decision tree

The initial data of serum protein peaks of training set group were processed by the Biomarker Wizard Version 3.1.0 Software. After that, the construction of the decision tree classification algorithm was performed by BPS (Ciphergen Biosystems, Inc.) using the processed data. The BPS produced many decision trees and evaluated the error cost (represented as ‘relative cost’ in the software) for each tree. Of the classification trees analyzed by BPS software, the most optimal classification tree with the lowest error cost (relative cost being 0.135 and BPS settings as follows: method = 1.50, advanced = 10 and testing = 18) was eventually established, and 4966.88, 5065.3, 5636.97 and 7766.87 (m/z) peaks were selected as the best biomarkers of RA in the classification tree. The protein peak m/z
which were 85.71 and 87.76%, respectively. Sensitivity and specificity than that in the training set group, the blinded set group, the decision tree model showed lower accuracy of 85.7, 87.5 and 88.0%, respectively (Table 5). In other autoimmune diseases and healthy people with the acc.

diseases and 23 of 25 healthy people, which suggested discrimination of each sample. In the training set group, the decision tree applied to the logical relationship within the tree (Fig. 3). Samples differentiated into terminal nodes 1 and 4 were assigned to RA; terminal nodes 3 and 6 to DC and terminal nodes 2, 5 and 7 to HC. For example, there was an unknown sample with protein peak m/z 5065.3 (intensity < 1.241) and m/z 4966.88 (intensity > 6.618); then, the sample was placed in terminal node 4 and it would be classified as RA. If the sample was placed in terminal node 7, then it would be assigned to HC. Based on the stochastic nature of reality, the possibility of misclassification of a new sample cannot be ruled out even for a pure node that contains only one sample type. To make sure that whether an unknown sample could be classified correctly or not, the expected probability was calculated for each class in the seven terminal nodes (Table 4). For example, the expected probability for HC samples to be misclassified as RA in terminal node 4 is 4.5%, whereas there was 86.4% chance for RA specimens to be correctly classified to terminal node 4. The corresponding Receiver Operating Characteristics curve of the optimal decision tree was also supplied by the Biomarker Patterns Software 5.0. The Receiver Operating Characteristics integral was 0.983 (Fig. 4). Besides, statistical analysis showed that the sensitivity and specificity of this decision tree model in training set group were 86.36% and 92.16%, respectively.

Diagnostic characteristics of the decision tree model
The accuracy of the boosting decision tree model for recognizing RA among people with other autoimmune diseases and healthy individuals was generated by the comparison between the decision tree judgment and the clinical diagnosis of each sample. In the training set group, the decision tree could identify 19 of 22 RA, 24 of 26 other autoimmune diseases and 23 of 25 healthy people, which suggested a high accuracy in identifying RA patients and control subjects. Results of validation test performed in the blinded set group showed that the decision tree could differentiate RA, other autoimmune diseases and healthy people with the accuracy of 85.7, 87.5 and 88.0%, respectively (Table 5). In the blinded set group, the decision tree model showed lower sensitivity and specificity than that in the training set group, which were 85.71 and 87.76%, respectively.

Fig. 1. Representative protein spectrum of a single RA serum sample generated from MALDI-TOF MS combined with WCX magnetic beads, showing the protein mass to charge (m/z) between 2000 and 30000.

4966.88 was over-expressed, while m/z 5065.3, 5636.97 and 7766.87 were down-regulated in RA group compared with control groups (Fig. 2). All 73 training set serum samples were differentiated into seven terminal nodes according to the logical relationship within the tree (Fig. 3). The capability of MALDI-TOF-MS to detect low-molecular weight peptides and proteins has helped to open a new realm of plasma components, comprising the fragmentome or peptidome, which were largely overlooked by traditional techniques. Furthermore, magnetic bead purification for the analysis of low-abundance proteins in the blood or body fluids can facilitate identifying new candidate biomarkers by MALDI-TOF-MS. Generally, protein profiling by MALDI-TOF-MS after proteome fractionation with magnetic beads is a robust, precise and high-throughput technique for the investigation of complex blood samples (13). For these reasons, MALDI-TOF-MS combined with magnetic beads is being applied widely to analyze potential disease markers in serum or plasma specimens.

Although proteomic technology has been applied to identify novel biomarker in various diseases including cancer, cardiovascular diseases, some autoimmune diseases such as SLE and Sjögren’s syndrome and so on (14–18), the application of proteomics in RA is still in its infancy. Several discriminating acute-phase proteins have been identified in serum, plasma and synovial fluid of patients with different rheumatic pathologies using classic proteomic approaches (19–23). Serum amyloid A protein is present in the synovial fluid and plasma of RA patients, but it is undetectable in plasma or synovial fluid of osteoarthritis (OA) patients. This acute-phase protein plays a crucial role in the early organization of host defense, but it might have a destructive effect in chronic inflammation (21). Certain fibrinogen isofoms and the calgranulin protein isofoms have been suggested to have a disease-associated expression or processing in biological fluid matrices of RA and OA patients (21, 23). However, to our knowledge, there is no study about scanning biomarkers of RA patients and establishing a diagnostic mode for RA by proteomic technique remains.

Discussion
Nowadays, the diagnosis of RA is mainly based on clinical symptoms and some laboratory indices such as RF, C-reactive protein, erythrocyte sedimentation rate and anti-CCP antibodies. However, these approaches are not effective in accurately diagnosing RA. In our study, we described the application of a new technology, MALDI-TOF-MS combined with magnetic beads, in identifying biomarkers and diagnosing RA.

Proteomic pattern analysis is one of the most promising approaches for the discovery and subsequent identification of proteins and peptides associated with various diseases (4–6, 12). The capability of MALDI-TOF-MS to detect low-molecular weight peptides and proteins has helped to open a new realm of plasma components, comprising the fragmentome or peptidome, which were largely overlooked by traditional techniques. Furthermore, magnetic bead purification for the analysis of low-abundance proteins in the blood or body fluids can facilitate identifying new candidate biomarkers by MALDI-TOF-MS. Generally, protein profiling by MALDI-TOF-MS after proteome fractionation with magnetic beads is a robust, precise and high-throughput technique for the investigation of complex blood samples (13). For these reasons, MALDI-TOF-MS combined with magnetic beads is being applied widely to analyze potential disease markers in serum or plasma specimens.

Fig. 2. Protein mass to charge (m/z) distribution across the MALDI-TOF-MS spectra generated from RA and HC serum samples.
In our study, 34 proteins were identified with discrepancy between RA patients and control objects by using MALDI-TOF-MS coupled with WCX magnetic beads, which suggested that a wide range of proteins might be involved in disorder of humoral immunity. These proteins could be those already known ones such as RF, anti-CCP antibodies and complements, although there could also be some unidentified proteins whose roles in the pathogenesis of RA remained obscure. With the identification of 34 protein peaks, the potential protein biomarkers may be used to diagnose RA.

With the help of Biomarker Patterns Software, the optimal boosting decision tree was established by analyzing ‘training set’ data. Judgment rules within this tree generated by the software well formed a model for recognizing RA. This model could provide one of the three following results including ‘RA sample’, ‘other autoimmune diseases sample’ and ‘normal sample’ for an unknown sample.
Since a test specimen cannot be identified as a RA sample with 100% accuracy, we thus calculated the expected probabilities to predict the chance of a sample being classified correctly or misclassified. Data showed that the chance of RA samples being misclassified into the HC terminal node was only 4.3%, and it did not affect the efficiency of this decision tree model for RA diagnosis. ‘Blinded testing set’ data were used for validating the performance of the decision tree model in diagnosing RA, and the results suggested that the diagnostic model could effectively distinguish RA patients from individuals with other autoimmune diseases or healthy people. Compared with anti-CCP assay, this decision tree model had a comparable specificity (87.76%) but a higher sensitivity (85.71%) in diagnosing RA. Therefore, the good performance indicates that this decision tree model could be a potential diagnostic tool for RA.

Four protein peaks at m/z 4966.88, 5065.3, 5636.97 and 7766.87 were members of the decision tree. Among the four, the protein peak m/z 4966.89 was over-expressed in RA cases, which may help to enhance the immune response of RA patients. On the contrary, the other three protein peaks were down-regulated in RA, which possibly contribute to suppressing overreactive immune response. The disorder of these proteins may play a critical role in promoting dysfunction of the autoimmune system. However, MALDI-TOF-MS cannot identify proteins or allow for absolute protein quantification. In this context, identifying the protein species represented by the peaks on the spectra would provide further evidences that they are indeed biologically meaningful molecules related to RA. Therefore, our further work will include searching the empirical proteomic ontology knowledge base to identify these differential proteins and then validating them by western blot or ELISA.
This is the first report about the discovery of four potential protein biomarkers for RA and the establishment of a novel RA diagnostic model by MALDI-TOF-MS combined with
magnetic beads and Biomarker Patterns Software. This diagnostic model comprised four proteins that may be helpful in accurately diagnosing RA, and further identification of these four proteins may have important therapeutic implications for RA patients.

Acknowledgements

None of the authors has any potential financial conflict of interest related to this manuscript.

References

10 Mei, J., Kolbin, D., Kao, H. T. and Porter, B. 2006. Protein expression profiling of postmortem brain in schizophrenia. Schizophr. Res. 84:204.