Biomarkers for cystic fibrosis lung disease: Application of SELDI-TOF mass spectrometry to BAL fluid

Gordon MacGregor a, Robert D. Gray a, Thomas N. Hilliard b, Margaret Imrie a, A. Christopher Boyd a, Eric W.F.W. Alton b, Andrew Bush b, Jane C. Davies b, J. Alastair Innes a, David J. Porteous a, Andrew P. Greening a,*

a School of Molecular and Clinical Medicine, University of Edinburgh, EH4 2XU, United Kingdom
b Imperial College, London, SW3 6LR, United Kingdom

Received 25 April 2007; received in revised form 5 December 2007; accepted 17 December 2007
Available online 1 February 2008

Abstract

Background: For cystic fibrosis (CF) patients there is a lack of good assays of disease activity and response to new therapeutic interventions, including gene therapy. Current measures of airways inflammation severity are insensitive or non-specific.

Methods: Bronchoalveolar lavage fluid from 39 CF children and 38 respiratory disease controls was obtained at bronchoscopy and analysed by surface enhanced laser desorption ionisation time of flight (SELDI-TOF) mass spectrometry. Recognized proteins were assessed for CF disease specificity. Individual protein identification of specific peaks was performed.

Results: 1277 proteins/peptides, >4 kDa, were detected using 12 different surfaces and binding conditions. 202 proteins/peptides were differentially expressed in the CF samples (p < 0.001), 167 up-regulated and 35 down-regulated. The most discriminatory biomarker had a mass of 5.163 kDa. The most abundant, with a mass of 10.6 kDa, was identified as s100 A8 (calgranulin A).

Conclusions: The application of SELDI-TOF mass spectrometry allows evaluation of proteins in BAL fluid avoiding the limitations of only analysing predetermined proteins and potentially identifying proteins not previously appreciated as biomarkers. Its application to cystic fibrosis should enable appropriate evaluation of evolving illness, of gene therapy and other new therapies.

Keywords: Paediatric; Proteomics; Pulmonary disease; Bronchoscopy

1. Introduction

Cystic Fibrosis (CF) is the most common fatal single gene defect in Caucasian populations [1]. CF is characterised by airways inflammation, which occurs early [2], chronic bacterial infection, frequent exacerbations and ultimately respiratory failure and death. Gene therapy is a logical ambition to prevent the fatal lung progression, and proof-of-principle for gene transfer has been reported in the nasal and pulmonary epithelia following non-viral gene therapy [3–5].

Assessing the effects of new therapies in CF is difficult since most standard clinical measurements lack adequate sensitivity and specificity. Physiological measurements such as first second forced expired volume (FEV1) reflect long term lung damage and are less sensitive to changes in airways inflammation. Assays of individual markers of inflammation, such as interleukin (IL)-8, allow quantitative assessment of airways inflammation but are not specific for CF. Surrogate markers of CF airways inflammation in exhaled breath have been reported but these too are not disease specific [6–8]. Therefore sensitive and specific biomarkers of CF airways inflammation are required.

Proteomics is the characterisation of the proteome (protein compliment) of a given biological system. Proteomics techniques can be applied to identify disease specific proteins and protein patterns in biological samples. In CF lung disease protein
biomarkers may reflect disease severity (phenotype) more sensitively than genetic data (genotype). Knowledge of specific “proteotypes” could be used to develop assays for the assessment of inflammation in the CF lung and changes to therapeutic interventions. Recent work utilising the proteomic method of two dimensional polyacrylamide gel electrophoresis (2D PAGE) has been reported demonstrating biomarkers of CF exacerbation in sputum [9]. This is an established technique in proteomics but can be labour intensive and time consuming. Furthermore, it may be less sensitive in demonstrating low abundance and low molecular weight proteins which tend to be underrepresented or absent on 2D PAGE [10]. These less abundant, lower molecular weight proteins are best investigated with mass spectrometry techniques.

Mass spectrometry separates proteins by mass and net electrical charge. Active selection of proteins in a biological sample by adherence to chromatographic surfaces and application of mass spectrometry (surface enhanced laser desorption ionisation time of flight; SELDI-TOF) allows analysis of proteins without reliance on antigen–antibody interactions. Disease related proteins can therefore be identified without any predetermined selection process. This may allow the development of specific and novel assays, and facilitate better clinical monitoring. The technique was recently reported in application to sarcoidosis [11] and most recently in CF [12]. We describe its application to CF, seeking identification of appropriate biomarkers to monitor the use of new treatments such as gene therapy. Assessments were made in young children to increase the chance of detecting primary changes consequent upon altered cystic fibrosis transmembrane regulator (CFTR). However, the control children also had active lung disease, including suppurative problems, which acted as a control for the bacterial colonization found in CF children.

2. Methods

2.1. Subjects

Subjects were children undergoing flexible bronchoscopy, for clinical reasons, in a single paediatric centre. Bronchoalveolar lavage (BAL) was performed in 39 children with CF and 38 non-CF children with a range of other respiratory diseases (Table 2). The reasons for bronchoscopy in the CF children were as follows:

to detect bacterial infection after diagnosis of CF (8 cases), failure to respond to antibiotic therapy during a pulmonary exacerbation [24], lobar consolidation [5], and microbiological surveillance while undergoing a surgical procedure [2]. The majority of CF patients and control subjects were taking antibiotic therapy at the time of bronchoscopy. The study was approved by the Local Research Ethics Committee and formal written consent was obtained from patients’ parents. The patients in this paper reflect an unselected subset of patients in a BAL/bronchoscopy study which will be reported elsewhere (Hilliard et al., in submission).

2.2. Procedures

Flexible bronchoscopy was carried out under general anaesthesia, with BAL performed in the middle lobe using 3 aliquots of 1 ml/kg of normal saline. The return fluid was pooled. BAL fluid was centrifuged at 2000g × 10 min to separate cellular and fluid phases. The supernatant was stored at −80 °C until analysis. The cell pellet was resuspended, treated with 0.1% dithiothreitol and differential cell counts performed on a cytospin preparation stained with May–Grunwald–Giemsa. IL-8 concentrations were measured by commercial ELISA (R&D, Minneapolis, USA).

2.3. Surface enhanced laser desorption/ionisation mass spectroscopy

Samples were thawed and applied to a range of chromatographic chip surfaces (Table 1) under specific binding conditions. The use of a range of surfaces and binding conditions avoided the possibility of unintended “pre-selection” of proteins that might occur if only one or two chip surfaces were used. For the metal affinity surface 2 μL of sample was incubated directly on to the chip surface. For other chip surfaces 20 μL of sample was incubated on the chip surface via a bioprocessor. Following incubation the chip surfaces were washed, sinapinic acid matrix applied, and air dried. Samples were analysed automatically on the Protein Biology System 2 SELDI-TOF mass spectrometer (Ciphergen Biosystems, Freemont, USA) with a laser intensity of 205, deflector at 4000 Da and a focus mass of 11,000 Da. Data were then processed in the following fashion. Baseline correction was performed to enhance the contrast of peaks to baseline using a

<table>
<thead>
<tr>
<th>Chip surface</th>
<th>Binding properties</th>
<th>Number of subjects in analysis*</th>
<th>Differential peaks</th>
<th>Highly differential peaks</th>
<th>Up regulated in CF</th>
<th>Down regulated in CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM10 at pH4</td>
<td>Cationic</td>
<td>24 21</td>
<td>81</td>
<td>21 (p&lt;0.001)</td>
<td>15 6</td>
<td></td>
</tr>
<tr>
<td>CM10 at pH6</td>
<td>Cationic</td>
<td>22 21</td>
<td>31</td>
<td>9 (p&lt;0.001)</td>
<td>9 0</td>
<td></td>
</tr>
<tr>
<td>CM10 at pH8</td>
<td>Cationic</td>
<td>23 10</td>
<td>33</td>
<td>7 (p&lt;0.001)</td>
<td>5 2</td>
<td></td>
</tr>
<tr>
<td>Q10 at pH6</td>
<td>Anionic</td>
<td>20 20</td>
<td>148</td>
<td>77 (p&lt;0.001)</td>
<td>71 6</td>
<td></td>
</tr>
<tr>
<td>Q10 at pH8</td>
<td>Anionic</td>
<td>24 20</td>
<td>27</td>
<td>20 (p&lt;0.001)</td>
<td>14 6</td>
<td></td>
</tr>
<tr>
<td>Q10 at pH10</td>
<td>Anionic</td>
<td>21 23</td>
<td>103</td>
<td>46 (p&lt;0.001)</td>
<td>33 13</td>
<td></td>
</tr>
<tr>
<td>IMAC Nickel</td>
<td>Metal ion</td>
<td>21 14</td>
<td>33</td>
<td>12 (p&lt;0.001)</td>
<td>12 0</td>
<td></td>
</tr>
<tr>
<td>IMAC Copper</td>
<td>Metal ion</td>
<td>19 17</td>
<td>24</td>
<td>10 (p&lt;0.001)</td>
<td>8 2</td>
<td></td>
</tr>
</tbody>
</table>

* Number of subjects refers to the number of spectra compared for each surface after data processing and normalisation (see Methods).
Fig. 1. Heat Map of protein expression on IMAC Ni surface. Areas of red represent overexpressed proteins. Hierarchical clustering separates most CF subjects from control subjects. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
fitting width of 4 times the expected width. Noise was automatically measured from 4 to 50 kDa and spectra corrected accordingly. Data were then normalised for total ion current. The total ion current for an individual spectrum was divided by the average total ion current over all spectra and thus each spectrum was awarded a normalisation coefficient. A normalisation coefficient of 1 reflected individual AUC the same as average, suggests a greater AUC than average, and a smaller AUC than average. Spectra with a normalisation coefficient of 2 were excluded from further analysis to ensure only good quality spectra were compared.

2.4. Protein identification

Protein identification was carried out by peptide mass fingerprinting of trypsin digested fragments following protein purification and gel electrophoresis. BAL fluid was bound to hydrophobic beads (Polymer Labs, Shropshire, UK) and eluted with increasing concentrations of acetonitrile. This allowed separation of proteins by chemical properties before application to 18% tricine/glycine polyacrylamide gel for one dimension electrophoresis. After Coomassie staining, bands were punched out, samples destained and protein passively eluted with formic acid into solution. An aliquot of eluate was reapplied to SELDI-TOF for confirmation of molecular weight of peak of interest, and trypsin digestion of the main sample performed. Digest fragments were applied for further mass spectrometry analysis on SELDI-TOF. The proteins were identified using on-line database recognition of the peptide fragments (Profound 4.10.05, Rockefeller University, USA).

2.5. Data analyses

Mass spectral data analyses were carried out using Ciphergen Express (Ciphergen Biosystems, Freemont, USA), a platform specific software package which allows automatic peak identification, clustering and data analysis. Peaks were subjected to a cluster wizard using Ciphergen Express (Ciphergen Biosystems, Freemont, USA). A cluster window of 0.3% was employed allowing the comparison of protein peaks across groups by their signal intensity, only peaks with a signal to noise ratio of 3 being used. Statistical analysis between the two groups (CF vs. Non-CF) was performed using Mann Whitney testing similar to the method used by Kriegova et al. [11]. The analysis was restricted to peptide peaks >4 kDa. No correction was made for multiple comparisons.

Table 2

<table>
<thead>
<tr>
<th>Measurement</th>
<th>CF</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>6.4 (0.7)</td>
<td>6.1 (0.6)</td>
</tr>
<tr>
<td>FEV1 (% predicted)</td>
<td>61 (3.3)</td>
<td>80 (3.4)</td>
</tr>
<tr>
<td>FVC (% predicted)</td>
<td>79 (3.8)</td>
<td>92 (4.2)</td>
</tr>
<tr>
<td>BAL Neutrophils (%)</td>
<td>49.2 (5.3)</td>
<td>21.8 (3.6)</td>
</tr>
<tr>
<td>BAL IL-8 (pg/ml)</td>
<td>1308.1 (56.4)</td>
<td>*716 (89.9)</td>
</tr>
<tr>
<td>Male (n)</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>Antibiotic therapy at time of BAL (n)</td>
<td>23</td>
<td>22</td>
</tr>
</tbody>
</table>

Figures represent mean (SEM). (FEV1)/FVC not recorded in all patients due to difficulty of procedure in the younger age group. There were 39 CF subjects and 38 control subjects.

* Control patients consisted of lower respiratory tract infection n=11, chronic cough 8, primary ciliary dyskinesia 7, croup 3, others 9.

* p<0.001.

Table 3

<table>
<thead>
<tr>
<th>Marker molecular weight (Da)</th>
<th>Correlation</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5163</td>
<td>No correlation</td>
<td>ns</td>
</tr>
<tr>
<td>11,589</td>
<td>No correlation</td>
<td>ns</td>
</tr>
<tr>
<td>10,545</td>
<td>No correlation</td>
<td>ns</td>
</tr>
<tr>
<td>5025</td>
<td>No correlation</td>
<td>ns</td>
</tr>
<tr>
<td>10,590 (s100 A8)</td>
<td>Negative correlation</td>
<td>0.0025</td>
</tr>
<tr>
<td>5321</td>
<td>No correlation</td>
<td>ns</td>
</tr>
<tr>
<td>12,269</td>
<td>Negative correlation</td>
<td>0.012</td>
</tr>
<tr>
<td>12,162</td>
<td>No correlation</td>
<td>ns</td>
</tr>
<tr>
<td>11,454</td>
<td>No correlation</td>
<td>ns</td>
</tr>
<tr>
<td>10,186 (s100 A12)</td>
<td>Positive correlation</td>
<td>0.0035</td>
</tr>
<tr>
<td>6214</td>
<td>Positive correlation</td>
<td>0.0004</td>
</tr>
<tr>
<td>21,066</td>
<td>Positive correlation</td>
<td>0.0025</td>
</tr>
</tbody>
</table>
3. Results

1277 clustering peaks, >4 kDa, were detected using 12 different surfaces and binding conditions. 202 peaks were differentially expressed in the CF samples ($p<0.001$) (Table 1). 167 of these were of higher signal intensity in CF compared to control and 35 had lower signal intensity than control. Fig. 1 demonstrates the separation of CF from non-CF subjects using a Heat Map (see figure legend). The statistically most significant biomarker with a predicted mass of 5.163 kDa (Fig. 2) is as yet unidentified. This biomarker completely discriminates CF from control and was expressed most efficiently on the IMAC nickel chip surface.

Thus far we have identified three proteins. The most abundant is s100 A8 (calgranulin A) at 10.6 kDa. This was increased significantly in the CF samples compared with controls, and was higher in controls with PCD than non bronchiectasis-associated controls (Fig. 3). s100 A8 was most readily found on the IMAC nickel chip surface but was also seen on the other chip surfaces. The other two identified proteins are s100 A9 (calgranulin B) and s100 A12 (calgranulin C). The identity of s100 A8 was confirmed by western blot and direct antibody capture on SELDI-TOF. Other proteins are in the process of identification.

BAL neutrophil and IL-8 levels are shown in Table 2. Regression analyses were performed for the IMAC nickel surface proteins and 5/12 correlated with BAL neutrophils, two negatively and three positively (Table 3; Fig. 4). The 5.163 kDa protein correlated both with BAL neutrophils ($p<0.01$) and BAL IL-8 concentration ($p<0.001$).

4. Discussion

Using SELDI-TOF technology we have been able to demonstrate a large panel of peptide biomarkers that differentiate, with high statistical significance, CF lung disease from a disease control population. Although these represent cross sectional data we would suggest they potentially represent biomarkers that may prove useful in monitoring CF airways inflammation and possibly CFTR function. Longitudinal data are required for evaluation of robustness and value of individual proteins once formally identified.

SELDI-TOF allows the identification of large numbers of potential biomarkers in a biological sample, based on molecular weights and chemical characteristics. In essence it provides high throughput screening for biomarkers, particularly when present in low abundance, avoiding the limitations of antibody binding and of only analysing predetermined proteins. It is able, therefore, to identify proteins not previously appreciated to be potentially valuable biomarkers. The technology has been applied to serum and urine to identify disease specific biomarkers [13–15], but its application to BAL fluid allows a valuable evaluation of proteins in the areas directly involved in airways inflammation, without confounding effects of, for example, inflammation of the liver or pancreas. This technology has recently been applied to BAL in chronic obstructive pulmonary disease, in a small number of adult patients to demonstrate specific biomarkers [10] and findings have also now been reported in CF BAL [16].

Even with the application of very restrictive parameters on data analysis (in order to avoid excessive claims of significance) we have ascertained a large number of potential biomarkers. These require specific identification. Thus far we have identified three using a combination of reverse phase chromatography, one dimension gel electrophoresis and trypsin digestion. The finding that s100A8 (calgranulin A) was the most abundant and one of the most significant markers of CF was not predictable prior to this study being performed although it has since been reported in the literature [16]. We believe that this finding highlights the value of the proteomics approach. However, >15 years ago s100 A8 and A9 (calgranulins A and B), as heterodimers, were described in the serum of patients with CF and in the mouse model of CF [17,18], and were referred to as CF antigen. S100A8 is known to regulate neutrophil, monocyte and lymphocyte migration [19–21], and CF antigen has been reported in non-CF diseases, including squamous carcinoma of skin, rheumatoid arthritis and dermatitis. In addition, plasma calprotectin (heterodimeric s100 A8/9) has been suggested as a marker of inflammation in CF [22]. It is produced by macrophages, epithelial cells and neutrophils and is a potent chemoattractant of neutrophils. It therefore appears to be a good candidate biomarker in CF airways inflammation.
We have applied very restricting limits on the data we have acquired using SELDI-TOF. This was deliberate, in order to avoid making unsustainable claims about the number/value of biomarkers thus identified. The utility of SELDI-TOF mass spectrometry as a diagnostic test for ovarian carcinoma has been criticised [23–26]. In Petricoin’s original paper problems with reproducibility in data analysis related to sample and data processing were suggested and furthermore the most discriminating proteins were in the mass range of <500 Da, a range where mass spectrometry is less effective in discriminating genuine protein peaks from background noise. Baggerly et al. [25] drew particular notice to the impact of different modes of data preparation and analysis on eventual results following SELDI-TOF mass spectrometry. They demonstrated that inaccuracies in sample collection protocol and mass calibration as well as inconsistent data manipulation could lead to misinformed results. Therefore, to ensure uniformity in our study we ensured that all sample preparation was uniform and that all data in our analyses were subject to the same steps of preparation prior to statistical testing.

We excluded proteins below 4 kDa from our analyses. This is almost certainly overly cautious and “misses” proteins that may be relevant. For example, at the lower MW ranges there were visible on the spectrographs the typical patterns of the neutrophil defensins. We also set a limit of a signal to noise ratio of <5, which more than excludes the potential for over interpretation of small peaks. The absolute number of potential biomarkers demonstrated by this approach is not critical, although it is clearly very large. The numbers we indicate in Table 1 could be altered if we applied other restrictions. For example, the technology allows further exclusion from analysis proteins with relatively low absolute signal intensity (even if the signal to noise exclusion we used has already eliminated the possibility of over interpretation of small peaks). If we had imposed such a filter at a mean signal intensity of <4 we would have excluded another 30–40% of the markers we list. We also recognise that some markers, such as s100 A8, will bind on to, and therefore be detected on, more than one chip surface/condition. Thus, some peptides may appear (falsely) twice or more. A proportion of some proteins may be doubly protonated. Since the “time of flight” is related to both MW and charge, such peptides will appear “twice”, once with half the MW of the original peptide. Nevertheless, despite all these restrictions this approach appears to be a very powerful tool in revealing potential biomarkers of lung disease in a completely non-biased fashion. We must however consider that different peaks may represent cleavage products of the same protein and contribute to higher numbers of potentially differentiating markers being recorded in the CF group.

A further criticism of this present study may be the failure to utilise a protease inhibitor during the preparation of BALF samples for storage prior to analysis as has been performed in other studies [16]. This may lead to breakdown products of proteins rather than true protein biomarkers to be identified. In spite of this all samples in this study were processed in uniform fashion an exposure to such factors as repeated freezing and thawing was minimized. We do draw the reader’s attention however to the importance of sample preparation in any proteomics study. As SELDI-TOF is a relatively new technology the reproducibility of this method is still to be fully assessed. Reproducibility was not formally assessed in this study but has been previously described for other body fluids serum (intra-assay CV of 15.6%, inter-assay variation of 24.4% [15]; urine (intra-assay CV 8–30%) [27]; and saliva (intra-assay CV 18% and inter-assay CV 31%) [28].

The biomarkers discovered in this study suggest disease specificity for CF vs. control, but may in reality reflect markers that are related to airways inflammation rather than CF per se, although the control group did contain subjects with other inflammatory airways conditions, including primary ciliary dyskinesia which also leads to bronchiectasis. Nevertheless we believe that further validation of these markers should be pursued in respect to other inflammatory lung diseases, in particular, non-CF related bronchiectasis.

We believe this proteomics approach to identification of specific disease/inflammation biomarkers is valid and effective. To allow translation into clinical practice we appreciate that bronchoalveolar lavage has limited potential for repeated sampling. We have preliminary data suggesting that many of the biomarkers can be identified by SELDI-TOF in induced sputum. These data are, therefore, complimentary to recent studies that have demonstrated the presence of specific biomarkers in the sputum of patients with CF, as demonstrated with 2D PAGE [9] and more recent data from CF BAL fluid [16]. If further studies confirm this there is indeed a firm base for clinical application of the markers identified. This may lead to the development of assays to monitor the effects of new therapies, including gene therapy in Cystic Fibrosis lung disease, and clearly has the potential for application to a wide range of lung diseases in general.

Acknowledgements

We gratefully acknowledge the support of the Cystic Fibrosis Trust acting through the UK CF Gene Therapy Consortium. GM, JCD and ACB were funded by the Cystic Fibrosis Trust. RG was supported by the MRC (programme grant G9313618). We thank Nathan Harris for his technical advice and support in the use of SELDI-TOF.

References


