Manfred Fussi
Department of Microbiology and Immunology
McGill University, Montreal

Using SELDI-TOF-MS To Discover Biomarkers In Leishmaniasis Patients

Submitted in August, 2007

“A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of M.Sc."

© Manfred Fussi, 2007
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>5</td>
</tr>
<tr>
<td>RESUME</td>
<td>5</td>
</tr>
<tr>
<td>GENERAL INTRODUCTION</td>
<td>6</td>
</tr>
<tr>
<td>Life cycle of Leishmania</td>
<td>7</td>
</tr>
<tr>
<td>Diagnosis in Leishmaniasis</td>
<td>8</td>
</tr>
<tr>
<td>Current Diagnostic Tests</td>
<td>9</td>
</tr>
<tr>
<td>Microscopic</td>
<td>9</td>
</tr>
<tr>
<td>Immunologic Examination</td>
<td>11</td>
</tr>
<tr>
<td>Molecular Methods</td>
<td>13</td>
</tr>
<tr>
<td>Surface Enhanced Laser Desorption Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS)</td>
<td>14</td>
</tr>
<tr>
<td>SELDI-TOF technology</td>
<td>16</td>
</tr>
<tr>
<td>From Sample to Chip</td>
<td>16</td>
</tr>
<tr>
<td>Desorption/Ionization-Process</td>
<td>16</td>
</tr>
<tr>
<td>Separating molecules</td>
<td>17</td>
</tr>
<tr>
<td>From Detector to Spectrum</td>
<td>17</td>
</tr>
<tr>
<td>Spectrum</td>
<td>17</td>
</tr>
<tr>
<td>Averaging Spectra</td>
<td>18</td>
</tr>
<tr>
<td>Baseline</td>
<td>18</td>
</tr>
<tr>
<td>Machine Parameters</td>
<td>18</td>
</tr>
<tr>
<td>How SELDI can Contribute to the Diagnosis of Leishmaniasis</td>
<td>19</td>
</tr>
<tr>
<td>Rational and Objective of the Research</td>
<td>20</td>
</tr>
<tr>
<td>Connecting Statement</td>
<td>21</td>
</tr>
<tr>
<td>Literature General Introduction</td>
<td>22</td>
</tr>
<tr>
<td>CHAPTER 1: USING SELDI-TOF-MS TO DISCOVER BIOMARKERS IN CUTANEOUS LEISHMANIASIS PATIENTS</td>
<td>32</td>
</tr>
<tr>
<td>Introduction</td>
<td>32</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>34</td>
</tr>
<tr>
<td>Patients and Samples</td>
<td>34</td>
</tr>
<tr>
<td>Fractionation and Binding of Sera</td>
<td>37</td>
</tr>
<tr>
<td>Reading of Arrays</td>
<td>38</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Discussion</td>
<td>81</td>
</tr>
<tr>
<td>Literature Article 2</td>
<td>86</td>
</tr>
<tr>
<td>Contributions of Authors</td>
<td>91</td>
</tr>
<tr>
<td>CONCLUSION AND SUMMARY</td>
<td>92</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>98</td>
</tr>
</tbody>
</table>
ABSTRACT

Leishmaniasis is an important parasitic disease for which no test with 100% sensitivity and/or specificity exists. We studied pre- and post-treatment sera (n=49) from patients with Old and New World cutaneous leishmaniasis (CL) as well as pre-treatment sera (n=45) from visceral leishmaniasis (VL) cases from Nepal. In all cases, samples were fractionated, bound on CM10 and IMAC30 arrays and read by surface-enhanced, laser desorption and ionization, time-of-flight mass spectrometry (SELDI-TOF-MS). Data were analyzed using CiphergenExpress (creating M/z scatter plots, p-values and ROC) and Biomarker Pattern Software (building decision trees). We attempted to visualize on gel all biomarkers detected in both software programs. Three of the 14 most promising candidate biomarkers for CL and 4 of 15 candidate biomarkers for VL were visible. These bands were cut and sequenced by LC-MS-MS. This allowed us to identify various proteins as possible biomarkers in CL and VL patients.

RESUME

La leishmaniose est une maladie parasitaire pour laquelle aucun test de sensibilité et/ou spécificité égale à 100% n’existe. Nous avons étudié des sérum (n = 49) pré- et post-traitement de patients atteints de leishmaniose cutanée (LC) de l’ancien et du nouveau monde de même que des sérum (n = 45) prétraitement de cas de leishmaniose viscérale (LV) du Népal. Tous les échantillons ont été fractionnés et sont retenues sur une surface chromatographique phase solide appelée ‘Chip’ (CM10 et IMAC30) et lus par SELDI-TOF-MS (Surface Enhanced Laser Desortion and Ionization, Time-of-Flight Mass Spectrophotometry). Les résultats ont été analysés par les programmes Ciphergen Express (en utilisant des graphique M/z, p-values et ROC) et Biomarker pattern Software (en créant des arbres décisionnels). Nous avons tenté de visualiser sur gel tous les marqueurs biologiques détectés par les deux programmes d’analyse. Trois des 14 marqueurs biologiques pour LC et 4 des 15 marqueurs biologiques pour VL ont pu être visualisés. Les bandes ont été sectionnées et séquencées par LC-MS-MS. Ceux ci nous permis aidé à identifier des marqueurs biologiques dans les LC et LV patients.
GENERAL INTRODUCTION

Leishmaniasis describes a spectrum of diseases with important clinical and epidemiological diversity that exist in humans in three major forms, cutaneous (CL), mucocutaneous (MCL) and visceral leishmaniasis (VL). The diseases are caused by obligate intracellular protozoan parasites belonging to various species of the genus *Leishmania* (Saha *et al*., 2006). Leishmaniasis is endemic in 88 countries. 72 of them are developing countries, including 13 of the least developed (see figure 1). 90% of the reported CL cases occur in seven countries - Afghanistan, Algeria, Brazil, Iran, Peru, Saudi Arabia and Syria, while 90% of VL occurs in five countries - Bangladesh, India, Nepal, Sudan and Brazil (Desjeux, 2004). An estimated 1.5-2 million children and adults develop symptomatic disease every year (cutaneous 1-1.5 million; visceral 0.5 million). The incidence of infection is substantial when subclinical infections are included (Desjeux, 2004).

![Distribution of VL (top image) and CL (bottom image) in the world (WHO).](image)

Fig.1: Distribution of VL (top image) and CL (bottom image) in the world (WHO).
About 70 of around 1000 known sandfly species can transmit leishmaniasis (Murray et al., 2005) including *L. lutzomyia* in the Americas and *L. phlebotomous* elsewhere (Mandell et al., 2005). Sandflies inoculate the skin with flagellated promastigotes, which invade or are phagocytosed by tissue macrophages or immediately recruited immune cells, including neutrophils (see fig.2). Within phagolysosomes of resident macrophages, surviving promastigotes transform and replicate as amastigotes, that can infect additional macrophages either locally or in distant tissues after dissemination (Murray et al., 2005).
Diagnosis in Leishmaniasis

The clinical signs and symptoms of VL and CL are not pathognomic. Kala-azar may be confused with other similar conditions such as malaria, tropical splenomegaly, schistosomiasis or cirrhosis with portal hypertension, African trypanosomiasis, milliary tuberculosis, brucellosis, typhoid fever, bacterial endocarditis, histoplasmosis, malnutrition, lymphoma, and leukaemia (Singh & Sivakumar, 2003). CL should be differentiated from tropical ulcers, impetigo, infected insect bites, leprosy, lupus vulgaris, tertiary syphilis, yaws, blastomycosis, skin cancer, etc. (Herwaldt, 1999; Lainson & Shaw, 1987; Ashford & Bates, 1998; Singh et al., 2005; Singh & Sivakumar, 2003). In a suspected case of leishmaniasis only the laboratory diagnosis can give a final answer (Singh, 2006).

In CL, serum antileishmanial antibodies can be detected in many subjects using sensitive reference assays (Romero et al., 2005). However, such tests are not widely available and, in practice, the diagnosis is made microscopically by identification of amastigotes in biopsies, scrapings, or impression smears of lesions (Vega-Lopez, 2003). The highest yield is usually obtained from material from the ulcer base (Weina et al., 2004; Ramirez et al., 2000). Diagnostic sensitivity can be increased to more than 85% when microscopy and culture are combined (Blum et al., 2004; Ramirez et al., 2000). In addition, culture (or DNA analysis) allows species identification (Vega-Lopez, 2003; Ramirez et al., 2000). Detection of parasite DNA in lesion material by PCR is generally considered to be the most sensitive test for the diagnosis of both cutaneous and mucocutaneous leishmaniasis (Vega-Lopez, 2003; Weina et al., 2004; de Oliveira et al., 2003; Faber et al., 2003; Oliveira et al., 2005), but is seldom the only test producing positive results (Weina et al., 2004; Oliveira et al., 2005). Culture and PCR testing are relatively difficult laboratory techniques that are not currently practical in developing countries (Murray et al., 2005).

For VL, direct visualisation of amastigotes in clinical specimens is the diagnostic gold standard in regions where deep tissue aspiration is routinely performed and the microscopy and technical skills are available. Diagnostic sensitivity for splenic, bone
marrow, and lymph node aspirate smears is >95%, 55-97%, and 60%, respectively (Guerin et al., 2002; Sundar et al., 2002; Davidson, 1998; Sundar & Benjamin, 2003; Ziljstra & El-Hassan, 2001; Pagliano et al., 2003). Elsewhere, including epidemic settings (Guerin et al., 2002; Marlet et al., 2003; Collin et al., 2004; Davidson, 1998), the diagnostic standard for VL is serum antileishmanial immunoglobulin G in high titre, measured primarily with direct agglutination tests or other serological assays (Desjeux, 2004; Herwaldt, 1999; Guerin et al., 2002; Gama et al., 2004; Davidson, 1998; Sundar & Benjamin, 2003; Abdallah et al., 2004). Freeze-dried antigen, which doesn’t need refrigeration (Abdallah et al., 2004) and rapid detection of anti-K39 antibody with fingerstick blood using an immunochromatographic strip test (Sundar et al., 2002) have advanced field serodiagnosis. In symptomatic patients, the sensitivity of the anti-K39 strip-test is high (90-100%) (Sundar et al., 2002; Veeken et al., 2003; Boelaert et al., 2004), but specificity varies by region (Sundar et al., 2002; Davies et al., 2003; Weina et al., 2004; Boelaert et al., 2004).

**Current Diagnostic Tests**

Microscopic, immunologic and molecular methods are used to diagnose leishmaniasis. On the following pages several methods will be discussed.

**Microscopic**

To confirm the diagnosis of leishmaniasis, the microscopic demonstration of *Leishmania* amastigotes in tissue aspirates or biopsies from bone marrow, spleen, lymph nodes or liver, skin slit smears or biopsies (Singh & Sivakumar, 2003) or in the peripheral blood buffy coat (Liarte et al., 2001) is needed.

**Spleen, Liver, Lymph Node and Bone Marrow Aspirate and Biopsy**

It is extremely important that spleen puncture is performed with utmost care in order not to tear the spleen capsule (Singh, 2006). The capsule of the spleen should only be
penetrated for a fraction of a second with the thinnest needle possible [preferably 21 gauge (0.8 mm)] using the correct technique and equipment with confidence to minimize the risk of complications like haemorrhage of the spleen (Bryceson, 1982; Chuley & Bryceson, 1983). But even in experienced hands the risk cannot be eliminated and fatal bleeding can occur in 2/10,000 patients (Bryceson, 1982). The sensitivity of liver biopsies is between 40% and 90% but, again, aspiration must only be attempted with the utmost care in order to avoid tearing the capsule (Singh, 2006). For bone marrow aspiration, material obtained from sternal or iliac crest puncture is a much safer, but more painful method. Parasites are also less likely to be demonstrated in direct stained films of marrow aspirates (Singh et al., 2005; Liarte et al., 2001; Chulay & Bryceson, 1983) with a sensitivity range in most studies between 76% and 85%. Culture of marrow aspirates can achieve sensitivities of up to 80% (Singh et al., 2005; Chulay & Bryceson, 1983). In the case of lymph node aspiration, the aspirate is extracted from any enlarged lymph gland after injecting sterile normal saline. To have the best chance of diagnosis, the aspirate should be used for both direct examination and culture (Singh, 2006). In Kala-azar cases, the sensitivity of lymph node aspiration may be up to 40-50%, in CL cases it can be higher (58.6%, Romero et al., 1999).

**Blood Buffy Coat**

Amastigotes can be rarely demonstrated in the buffy coat of peripheral blood. This finding is more common in severely immunocompromised AIDS patients (Martinez et al., 1993) and patients on immunosuppressive therapy (Maggi et al., 2004). While some authors have found sensitivity up to 53% (Delgado et al., 1998) others have found this method to perform very poorly (7.6% sensitivity) (Navin et al., 1990).

**Tegumentary Leishmaniaisis**

The routine diagnosis of CL patients depends on the examination of skin lesions using smears and cultures of dermal scrapings or examination of sections obtained from a skin biopsy. Normally 3-5 aspirates from different lesions or portions of lesions are used
(Singh, 2006). For dermal scrapings, 3-5 samples from different lesions should be taken. The first collections should be used for microscopy and last for culture to minimize the risk of contamination (Singh, 2006). The sensitivity of direct microscopic identification of *Leishmania* amastigotes from tegumentary disease forms can be up to 60-65% while the sensitivity of culture remains less than 42%. Lack of sensitivity is typically attributed to the elevated chance of culture contamination with these samples (Singh & Sivakumar, 2003).

**Culture Examination**

Up to 90% of the active Kala-azar cases will grow promastigotes in their splenic and liver aspirates (Manson-Bahr, 1987). Culture-based diagnosis of MCL has low sensitivity as the organisms are often scant. The biggest handicap is culture contamination at early stages, even in best laboratory setups (Singh *et al*., 2005; Weigle *et al*., 1987).

**Immunologic Examination**

The hallmark of VL is hyperimmunoglobulinaemia. Therefore a number of methods to detect antibodies have been developed. In case of CL however, the humoral immune response is often extremely poor and therefore no serological method is routinely helpful for cutaneous or mucocutaneous leishmaniasis (Singh, 2006). The following methods are used to detect antibodies:

**Fluorescent Antibody Test**

The indirect fluorescent antibody test is commonly practised for anti-leishmanial antibody detection using fixed promastigotes. Antibodies can be detected during the early stages of visceral infection and are undetectable after 6-9 months after cure. Titres above 1:20 are significant and above 1:128 are diagnostic, but there is a possibility of a cross-reaction with trypanosomal sera (Singh *et al*., 2005; Boelaert *et al*., 2004). The reported sensitivity of serum antibody tests varies from as low as 28.4 % (Boelaert *et al*., 2004) to 86.6 %
Low sensitivity can be overcome to some extent by using *Leishmania* amastigotes as the antigen instead of the promastigotes (Singh, 2006). Direct fluorescence is more useful to diagnose CL, MCL and PKDL, where fluorescent dye conjugated antibodies can be used as tracers to detect the antigen (amastigotes) in the tissue sections or smears (Singh, 2006).

**Direct Agglutination Test (DAT)**

The DAT assay is inexpensive and simple making it ideal for both field and laboratory use. DAT uses whole, stained promastigotes either as a suspension or in a freeze-dried form, which are heat stable, facilitating the use in the field (Abdallah *et al*., 2004). The DAT has been found to be 72-100% specific and 91-100% sensitive (Liarte *et al*., 2001; Tavares *et al*., 2003). However, the major disadvantage of DAT is the long incubation time of 18 h and the need for serial dilutions of blood or serum. Furthermore, this assay has no prognostic value for evaluating the parasitological cure of the disease because the results may remain positive for several years after cure (Singh, 2006).

**Enzyme Linked Immunosorbent Assay (ELISA)**

ELISA is one of the most sensitive tests for the serodiagnosis of VL. The test is useful for laboratory analysis or field applications and to screen large numbers of samples. The sensitivity and specificity of this test format is greatly influenced by the antigen used (Singh, 2006). For example, the excretory, secretory and metabolic antigens released by *L. donovani* promastigotes into a protein-free medium used for serodiagnosis of VL showed 100% specificity and sensitivity (Singh, 2006). The recombinant antigen, rK39 is specific for antibodies against VL caused by *L. donovani* complex members. It is highly sensitive and predictive for onset of acute disease (Singh *et al*., 2002) and has a high predictive value for detecting VL in immunocompromised persons like AIDS patients (Singh *et al*., 2005). However, there are also reports from Sudan and other countries were assays based on this antigen have much lower sensitivity and specificity. For example, in
Sudan, 7% of parasitologically proven cases were missed by the rK39 ELISA (Ziljstra & El Hassan, 2001).

**Antigen Detection**

Circulating antigen levels are expected to theoretically correlate with the total body parasite load. Therefore antigen detection may be an ideal test in immunocompromised patients, in whom the antibody response is often poor. However, the detection of antigen in a patient’s serum can be complicated by the presence of high levels of antibodies, circulating immune complexes, serum amyloid, rheumatoid factor and autoantibodies, because they may mask immunologically important antigenic determinants or competitively inhibit the binding of free antigen (Singh, 2006). Nevertheless a latex agglutination test developed to detect leishmanial antigens in the urine of VL patients has been reported to have 100% specificity and 68-100% sensitivity (Attar et al., 2001).

**Leishmanin skin test (LST)**

Delayed hypersensitivity is an important feature of human CL and can be measured by the Leishmanin test, also known as the Montenegro reaction. No cross-reaction occurs with Chagas disease, but some cross-reactions are found with cases of glandular tuberculosis and lepromatous leprosy (Singh, 2006).

**Molecular Methods**

PCR has been proved to be the most sensitive and specific technique amongst the molecular methods used for clinical diagnosis. Various gene targets and nucleic acids can be used in PCR. The most important gene targets identified to date are 18S-rRNA, small subunit rRNA (SSU rRNA), a repetitive genomic sequence of DNA, the miniexon (spliced ladder) gene repeat, the b-tubulin gene region, gp63 gene locus, internal transcribed spacer (ITS) regions, micro-satellite DNAs such as maxi- and minicircles of kinetoplast DNA (da Silva et al., 2004; Pizzuto et al., 2001; Wortman et al., 2001; Gangneux et al., 2003). The specificity of PCR on bone marrow aspirates has been
reported to be up to 100 % (Martin-Sanchez et al., 2004) with sensitivity of 80-93.3 %. These numbers compare favorably with the 50-60 % reported sensitivity of smear and culture examination (Tavares et al., 2003).

**Surface Enhanced Laser Desorption Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS)**

Hutchens & Yip (1993) introduced the concept of SELDI because they wanted to develop a direct mass spectrometric technology for the detection of proteins in heterogenous samples that avoided many of the sample preparation problems inherent to matrix-assisted laser desorption-ionization (MALDI)-TOF. SELDI is a type of affinity-based mass spectrometry (Hutchens & Yip, 1993; Merchant & Weinberger, 2000; Davies, 2000; Fenn et al., 1989; Wright et al., 1999) in which the protein sample is directly applied to a pretreated surface termed the “ProteinChip” (Ciphergen Biosystems ProteinChip, Fremont, CA). SELDI-TOF is a MALDI-TOF-related research system that is more sensitive than other types of MS. One of the unique strengths of SELDI-TOF-MS is its ability to analyze proteins from a variety of crude sample types, with minimal sample consumption and processing. While SELDI provides a unique sample preparation platform, it is similar to MALDI MS in that a laser is utilized for the ionization of samples that have been co-crystallized with a matrix on a target surface. Unlike MALDI target surfaces, however, the SELDI protein chip surfaces are uniquely designed to retain proteins from complex mixtures according to their specific properties using chromatographic-based selectivity (Xiao et al., 2005). The on-chip purification of SELDI-TOF-MS as well as the small amount of starting material needed are great advantages in comparison to MALDI-TOF-MS (Engwegen et al., 2006). The disadvantages of SELDI-TOF are that the resolution of the instrumentation is usually quite low (many peptides can exist within each peak), and the ion peaks of interest cannot be directly identified. At the current time, the MALDI-TOF approach cannot routinely detect proteins in the low abundance range (<4 pg mL⁻¹), which almost all immunoassays can reach with ease (Petricoin et al., 2006).
The profiling of differentially-expressed proteins in healthy and diseased conditions is the most common application of clinical “proteomics” (Banks et al., 2000). SELDI-TOF is an emerging proteomic technology in biomarker discovery that allows for rapid and sensitive analysis of complex protein mixtures. This novel technology has the potential to contribute greatly to the discovery and clinical exploitation of clinically relevant biomarkers (Gretzer et al., 2003).

Bioinformatics is an emerging field focused on the organization and interpretation of massive biological datasets produced by high-throughput platforms such as SELDI (Srinivas et al., 2001). In the context of a biomarker discovery program, bioinformatics is used to assemble and search protein databases and to develop sophisticated analysis software (Gretzer et al., 2003). Such software permits the simultaneous analysis of multiple ProteinChips along with the ability to compare the data from different samples (Fung et al., 2001; Wilkins et al., 1996; Gras et al., 1999; Barclay et al., 1997). Once a protein or set of proteins has been identified as either unique or differentially expressed compared to control samples, these proteins are candidate biomarkers (Gretzer et al., 2003).

As many as 20,000 proteins have been estimated to populate human serum with an overall serum protein concentration maintained at about 60-80 mg/mL (Anderson & Anderson, 2002). For basic biomedical research and routine clinical assays, serum is one of the most valuable specimens. A huge range of serological assays have been developed based on the detection of changes in serum proteins that signal pathological abnormalities for diagnosis and prognosis in virtually all human conditions (Zhen Xiao et al., 2005). In the context of infectious diseases, these assays almost always target the detection of antibodies as markers of current or past infection. SELDI technology offers the first realistic opportunity to discover and exploit other serum proteins (of host and pathogen origin) for diagnostic and prognostic purposes.
SELDI-TOF technology

From Sample to Chip

SELDI chip types have different surfaces, which range from chromatographic chemistries that bind many different molecules, to surfaces with a specific biomolecular affinity (e.g. antibodies, receptors, enzymes and ligands), that bind one specific molecule or class of molecules. After putting the sample on a chosen chip type (Merchant & Weinberger, 2000), and washing weakly-bound proteins away (Dijkstra et al., 2006), the sample is overlaid with a solution of small photosensitive molecules (e.g.: sinapinic acid). This creates a crystalline matrix on each sample as it dries. The photosensitive (matrix) molecules facilitate desorption and ionization of proteins. Proteins compete for sites in the matrix crystals (Tang et al., 2004) if the protein concentration is so high that not all proteins can be incorporated in the matrix. Proteins that are more easily embedded will then have a higher concentration in the matrix. The sample is then put into a vacuum chamber, the so-called flight tube (Dijkstra et al., 2007).

Desorption/Ionization-Process

When a short laser pulse hits the crystal structure, it excites the matrix molecules. The energy of the excited matrix molecules is converted to thermal energy which heats up the crystal locally to around 1000 K within a fraction of 1 nanosecond (Zenobi & Knochenmuss, 1998). Excited matrix molecules can protonate closely-associated proteins including adjacent matrix proteins. The overheated part of the crystal explodes together with the embedded proteins into a plume. There are two types of ionization processes: primary ion formation that occurs during and immediately after the laser pulse, and secondary ion-forming reactions that take place in the plume (Zenobi & Knochenmuss, 1998). The plume expands but keeps a rather high density and a temperature of 500 K for ~100 ns (Zenobi & Knochenmuss, 1998), causing many collisions between proteins and matrix molecules or matrix clusters. In the gas phase, non-matrix proteins typically have
higher proton affinities than the matrix molecules themselves (Zenobi & Knochenmuss, 1998).

For SELDI analysis, several different matrix molecules can be used. One of the most common is sinapinic acid. It enables efficient laser desorption and ionization of large proteins, e.g. >10 kDa (ProteinChip Software 3.1 Operation Manual, Fremont CA Ciphergen Biosystems Inc 2002, www.ciphergen.com).

Separating molecules

The physical principle of the TOF analyzer in SELDI is that sublimated molecules, which have a different mass (m) over charge (z) ratio (m/z) are accelerated differently and traverse the flight tube with different velocities. Therefore, the time for an ion to pass through the fixed-length flight tube to the detector depends on its m/z (Dijkstra et al., 2007).

From Detector to Spectrum

Basically, the detector measures the period between the moment the electric field switches on (i.e.: laser pulse) and the moment that a particle hits the detector (Dijkstra et al., 2007). The detector gain is defined by the ratio of the eventually released electrons and the number of molecules striking the first detector plate. The detector gain strongly depends on the kinetic energy of the detected molecules (Mueller, 2003). The user can scale the detector gain by a machine setting called ‘detector sensitivity’ (Hellsing et al., 1985) which corresponds to the potential difference between the detector plates.

Spectrum

According to Dijkstra et al (2007), the computer screen displays the counted totals of electrons per time interval in the spectrum. Because the time intervals are small, the spectrum can be interpreted as an almost continuous, ‘smoothed’ histogram (time versus
numbers of electrons). Within the spectrum, a ‘baseline’ is typically distinguished with zero or more peaks. Individual molecules that all have the same chemical formula and occur in the same charge state, are detected as a signal, which is defined as a singleton peak. A singleton peak may show overlap with other singleton peaks. The peak area (area under the curve) is assumed to be proportional to the numbers of molecules present in the protein mixture (Merchant & Weinberger, 2000).

**Averaging Spectra**

A final spectrum is an average of spectra acquired by several individual laser pulses. Each single laser pulse is fired on a different position on the dried droplet. The crystal density and size varies with the position within the dried droplet (Dijkstra et al., 2007). The Ciphergen software automatically removes pulses that generate signals that are either too high or too low to be interpretable (ProteinChip Software 3.1 Operation Manual).

**Baseline**

Dark current (which is coincidental electron emission from each plate inside the detector due to thermal energy) and detected air molecules together form the sample-independent part of the baseline. The chemical noise forms the sample-specific part of the baseline and is more abundant when higher laser energies are applied, probably as a result of greater fragmentation. Chemical noise is also more abundant in more complex samples comprised of many different protein species because of fragmentation and metastable decay due to phenomena like collisions with the background gas (Mueller, 2003; Zenobi & Knochenmuss, 1998). With TOF, the chemical noise shows exponential decay (Dijkstra et al., 2006).

**Machine Parameters**

Machine settings can influence both the quality and reproducibility of the data generated. A higher applied voltage (that creates an electric field with a larger potential difference)
generates faster ions, which have a shorter TOF to the detector and induce more electrons when striking the detector, leading to higher peaks. Ion clouds, which have a longer TOF expand more and generate broader and lower peaks (Dijkstra et al., 2007). The detector with higher sensitivity settings generates more electrons when detecting the same number of ions, leading to higher peaks.

The energy of the laser is the only machine parameter that can strongly influence the desorption/ionization-process and the number of ions generated. Dijkstra et al (2007) report that a 210 µJ laser pulse generates almost no ions which leads to small peaks, while a 250 µJ laser pulse generates many ions which leads to large peaks with low resolution. Higher laser energy increases the thermal energy of the ions, resulting in more and more violent collisions between the ions and increasing fragmentation and chemical noise (Dijkstra et al., 2007). To minimize the impact of the machine-based variability, we meticulously calibrated the machine used in our study before running each sample sets. For the calibration we used the protein MW standards kit C100-0001 (bound on an NP20 array) and an insulin array. This kit consisted of the following proteins (MW in Dalton): Insulin, bovine (5733.6); Ubiquitin, bovine (8564.8); Cytochrome C, bovine (12230.9); Superoxide dismutase, bovine (15591.4); Myoglobin, equine (16951.5); Beta-lactoglobulin A, bovine (18363.3); Horseradish peroxidase (43240); Serum albumin, bovine (66.410); Conalbumin, chicken (77490); IgG, bovine (147300).

How SELDI can Contribute to the Diagnosis of Leishmaniasis

Since proteins represent the preponderance of the biologically active molecules responsible for cellular functions, it is believed that the direct measurement of protein expression can more accurately indicate cellular dysfunction underlying the development and progression of disease because often there is no predictive correlation between mRNA abundances and the quantity of the corresponding functional protein present within a cell (MacNeil, 2004).
Recent advances in mass spectrometry (MS) such as SELDI-TOF MS (Isaaq et al., 2002) have revolutionized the high-throughput proteomic profiling of serum. The technique is, in principle, applicable to any disease state, and mass-spectral signatures have recently been reported for alcoholism (Nomura et al., 2004), infarctive and haemorrhagic strokes (Allard et al., 2004), rheumatoid arthritis (Uchida et al., 2002), ectopic pregnancy (Gerton et al., 2004) and HIV-related cognitive impairment (Luo et al., 2003). Additionally SELDI-TOF MS has recently been applied to the search of biomarkers in ovarian (Kozak et al., 2003), prostate (Banez et al., 2003) and lung cancer (Zhukov et al., 2003) as well as human African trypanosomiasis (Agranoff et al., 2005). During the last two decades, advances in biomarker research have also benefited from the introduction of new proteomic analytical techniques (Adam et al., 2001; Williams & Hochstrasser, 1997).

Apart from its remarkable accuracy, other intriguing features of this technology are that it requires no major prior assumptions about the nature of the proteins that contribute to the signature and that its diagnostic performance does not depend on knowledge of their identities (Agranoff et al., 2005).

Papadopolous et al (2004) mention that disease states might be associated with distinctive configurations of circulating proteins: so-called ‘proteomic fingerprints’. More than 1000 distinct gene products can be discerned in human plasma (Anderson et al., 2004), although only a handful of protein detection assays are currently in routine clinical use (Agranoff et al., 2005).

**Rational and Objective of the Research**

No test on the market can deliver 100 % sensitivity and specificity in all leishmaniasis cases. Therefore the need for new diagnostic tests still exists. Using SELDI-TOF-MS to search for biomarkers in serum of leishmaniasis patients has promise because the technology is very sensitive and allows high throughput analysis. The aim of this study was to test if SELDI-TOF-MS can be used to detect diagnostically-useful or otherwise informative biomarkers in both cutaneous and visceral leishmaniasis.
Connecting Statement

One of the geographic hotspots of CL is the Near- and Middle-East. Because of military operations in this region, a growing number of cases has been reported by the US military. We were very fortunate to establish a collaboration with Dr Naomi Aronson from the Walter Reed Army Medical Center, who provided us with well-characterized serum samples from CL patients and control sera.
Literature General Introduction


CHAPTER 1:

USING SELDI-TOF-MS TO DISCOVER BIOMARKERS IN CUTANEOUS LEISHMANIASIS PATIENTS

Manfred Fussi\(^1\), Naomi Aronson\(^2\), Christine Straccini\(^1\), Bernard Gibbs\(^4\), Adam Rainczuk\(^1\), Greg Matlashewski\(^3\), Momar Ndao\(^1\) and Brian J. Ward\(^1\)

National Reference Centre for Parasitology, Montreal General Hospital, McGill University, Montreal, Quebec, Canada\(^1\)
Walter Reed Army Medical Center, USA\(^2\)
Department of Microbiology & Immunology, McGill University, Montreal, Quebec, Canada\(^3\)
Sheldon Biotechnology Centre, McGill University, Montreal, Quebec, Canada\(^4\)
(Paper in preparation)

Introduction

Leishmaniasis is endemic in 88 countries (66 in the Old World and 22 in the New World) with an estimated yearly incidence of 1-1.5 million cases of cutaneous leishmaniasis (CL). 350 million people are at risk of infection, the overall prevalence is 12 million (Desjeux, 1996). It has been estimated that 90% of all CL infections develop in Afghanistan, Pakistan, Syria, Saudi Arabia, Algeria, Iran, Brazil, and Peru (Desjeux, 2004).

CL is caused by a wide range of species, including *Leishmania major*, *L. aethiopica* and *L. tropica* in the Old World, and *L. mexicana*, *L. braziliensis*, *L. amazonensis*, *L. pifanoi*, *L. garnhami*, *L. venezuelensis*, *L. guyanensis*, *L. peruviana*, and *L. panamensis* in the New World. Mucocutaneous Leishmaniasis (MCL) is most commonly caused by the New World species, *L. braziliensis* (Renee *et al*., 2005, Wilson *et al*., 2005), though *L. aethiopica* has also been reported to cause this syndrome.

CL lesions may vary in severity (e.g., lesion size), clinical appearance (e.g., open ulcer versus flat plaques versus wart-like lesions), and duration [e.g., time of evolution, time to spontaneous cure (Reithinger & Dujardin, 2007)]. The broad clinical spectrum of leishmaniasis makes the diagnosis of present and past cases difficult. However, differential diagnosis is important because other diseases that produce lesions with a clinical appearance similar to those of leishmaniases (e.g., leprosy, skin cancers, and
tuberculosis for CL) are often present in areas of endemicity (Reithinger & Dujardin, 2007).

Because of their high specificity, microscopic visualization and culture remain the gold standard diagnostic tests in leishmaniasis (Herwaldt, 1999). Importantly, the sensitivity of microscopy and culture tends to be low and can be highly variable (Herwaldt, 1999), depending on the number and dispersion of parasites in biopsy samples, the sampling procedure, and the technical skills of the personnel. Polymerase chain reaction (PCR) applied to lesion material can be highly sensitive and specific (Singh, 2006) but usually requires a biopsy for optimal sensitivity and the expensive machinery and reagents are not feasible for most endemic areas. There is currently no serologic assay that can reliably diagnose subjects with CL.

Proteomic analysis has been successfully employed in the discovery of new biomarkers in different human diseases (Srinivas et al., 2002). Surface-enhanced laser desorption ionization-time of flight-mass spectrometry (SELDI-TOF-MS) is a promising new technology in the search for new biomarkers and was first introduced by Hutchens & Yip (1993). Since the study of Petricoin et al (2002a) on protein profiling to detect ovarian cancer, the use of SELDI protein profiling as a diagnostic tool has become an important subject of investigation (Petricoin et al., 2002b). To date, this approach has been applied to a number of different conditions, for example ovarian (Petricoin et al., 2002a; Kozak et al., 2003; Zhang et al., 2004; Vlahou et al., 2003; Ye et al., 2003; Rai et al., 2002), prostate (Banez et al., 2003; Adam et al., 2002; Petricoin c et al., 2002; Qu et al., 2002; Li et al., 2004) and lung cancer research (Zhukov et al., 2003), as well as inflammatory diseases (Poon et al., 2005; Zhu et al., 2004) and infectious diseases like human African trypanosomiasis (Agranoff et al., 2005).

This study is focused on the detection of potential serum biomarkers for the diagnosis of cutaneous leishmaniasis using SELDI-TOF-MS.
Materials and Methods

Patients and Samples

Samples consisted of pretreatment sera from 18 human patients (see table 1) diagnosed with CL [3 infected with New World CL (NWCL) and 15 with Old World CL (OWCL)], follow-up sera from most of these patients (17 sera taken after 14 days of treatment and 14 sera taken after 20 days of treatment). All patients were treated with sodium stibogluconate at 20 mg/kg/day). Sera from 4 healthy human controls were also available.

The controls were age-, sex- and race-matched samples based on the demographic distribution of the last 100 patients treated at the Walter Reed Army Medical Centre. They were >18 years old health care beneficiaries with no chronic medical history, no history of nitrate/nitrite containing medicines or recreational drug use, no prior or current *Leishmania* infection, no travel to *Leishmania* endemic areas and no history of working in a laboratory handling *Leishmania*.

All sera were obtained from US military personnel enrolled in the Walter Reed Army Hospital prospective study of CL (Principal investigator: Dr Naomi Aronson). Samples were aliquoted (30 µL/tube) and then stored at -20°C until further use. All fractionated serum samples had been thawed a maximum of 3 times between collection and analysis. The following clinical information (see table 2) was available from the patients: duration, location, number and size of the lesion(s), lymphadenopathy (yes or no) as well as information about tests already done (histopathology, culture and PCR). We also received information about the age, race and gender of the persons, origin of infection (country where the infection occurred) and *Leishmania* species (see table 1).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Age</th>
<th>Race</th>
<th>Gender</th>
<th>Species</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>W</td>
<td>M</td>
<td><em>L.major</em></td>
<td>Iraq</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>W</td>
<td>M</td>
<td>NWCL</td>
<td>Belize</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>A</td>
<td>M</td>
<td><em>L.major</em></td>
<td>Kuwait</td>
</tr>
<tr>
<td>4</td>
<td>n.a.</td>
<td>W</td>
<td>M</td>
<td>NWCL</td>
<td>Panama</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>B</td>
<td>M</td>
<td><em>L.major</em></td>
<td>Afghanistan</td>
</tr>
<tr>
<td>Sera</td>
<td>Duration (days)</td>
<td>Location</td>
<td>L. A.</td>
<td>No.</td>
<td>Size</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
<td>----------</td>
<td>-------</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>180</td>
<td>B arms</td>
<td>No</td>
<td>5</td>
<td>34x25; 36x26</td>
</tr>
<tr>
<td>2</td>
<td>120</td>
<td>L thigh</td>
<td>No</td>
<td>1</td>
<td>30 mm</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>R hand, forearm, upper arm, shoulder, flank, thigh</td>
<td>No</td>
<td>7</td>
<td>Various sizes, some large 30 mm</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>Face</td>
<td>No</td>
<td>1</td>
<td>21x10 mm</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>R neck, L earlobe</td>
<td>Yes</td>
<td>2</td>
<td>20x29mm, 7x6 mm</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>L knee/leg</td>
<td>No</td>
<td>2</td>
<td>12x11, 18x16</td>
</tr>
<tr>
<td>7</td>
<td>67</td>
<td>B arms, face</td>
<td>Yes</td>
<td>9</td>
<td>15x11, 6x6, 9x11, 14x7, 23x11, 13x8, 10x12, 6x6, 7x9</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>R arm, B legs</td>
<td>No</td>
<td>3</td>
<td>34x17, 27x18, 23x18</td>
</tr>
</tbody>
</table>

Table 1: Clinical data of the tested patients.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>No</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>150</td>
<td>L arm, B legs</td>
<td>No</td>
<td>3</td>
<td>8x8, 11x6, 8x8</td>
<td>No amastigotes</td>
<td>L.major</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>150</td>
<td>face</td>
<td>No</td>
<td>1</td>
<td>6x10</td>
<td>Amastigotes</td>
<td>L.major</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>90</td>
<td>R arm</td>
<td>No</td>
<td>1</td>
<td>13x14</td>
<td>Amastigotes</td>
<td>n.d.</td>
<td>n.d.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>75</td>
<td>elbow</td>
<td>No</td>
<td>multiple</td>
<td>5x5</td>
<td>Superficial and deep chronic granulomatus inflammation, no amastigotes</td>
<td>negative</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>75</td>
<td>R calf</td>
<td>Yes</td>
<td>1</td>
<td>60x35</td>
<td>Granulomatus inflammation, + amastigotes</td>
<td>L.major</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>60</td>
<td>L knee</td>
<td>Yes</td>
<td>3</td>
<td>“Large” approx 40mm</td>
<td>Amastigotes seen</td>
<td>L.major</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>63</td>
<td>R thigh, L ankle</td>
<td>No</td>
<td>5</td>
<td>8x7, 15x12, 11x10, 12x10, 18x11</td>
<td>Amastigotes</td>
<td>L.major</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>750</td>
<td>Face</td>
<td>No</td>
<td>1</td>
<td>10x10 mm</td>
<td>Granulomatus and lymphoplasmacytic infiltrate, no amastigotes</td>
<td>L.VB</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>30</td>
<td>R ankle</td>
<td>No</td>
<td>1</td>
<td>23x13</td>
<td>No parasites</td>
<td>negative</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>75</td>
<td>Both arms</td>
<td>No</td>
<td>7</td>
<td>3x1.5, 2x3, 4x2, 6x6, 17x10, 25x20, 18x10, 16x12, 6x8</td>
<td>Granulomatus inflammation, numerous amastigotes in giant cells</td>
<td>L.major</td>
<td>+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Clinical and laboratory data of the tested patients (L.A. = lymphadenopathy; R = right; L = left; W = white; B = black; A = asian; M = male; F = female; n.d. = not done).
Fractionation and Binding of Sera

Fractionation greatly increases the number of peptide and protein ion signals, when compared to unfractionated samples, increasing the chance of finding potential biomarkers (Linke et al., 2004). For fractionation, Cyphergen’s Expression Difference Mapping™ Kit-Serum Fractionation was used, allowing a high throughput processing on anion exchange beads in a 96-well microplate format. To increase the number of proteins visualized, sera were separated into six different fractions (pH 9+ flow through, pH7, pH5, pH4, pH3, and organic wash). This fractionation procedure significantly increases the number of peaks detectable from each sample (Fung & Enderwick, 2002). For the anion exchange fractionation, we used 20 µL of serum and 30 µL of U9 buffer as starting material according to the Ciphergen protocols.

To find the optimal array-type and fraction, each fraction of the 18 pretreatment sera and 4 controls was bound on 2 different array surfaces (ProteinChip™: Ciphergen Biosystems Inc.), the weak cation exchange chip (CM10) and immobilized metal affinity capture coupled with copper (IMAC-Cu²⁺) chip arrays. For the follow-up samples, only the fraction(s) that gave the most promising results in the pretreatment sera, was/were studied: primarily fraction 4 on CM10 and fraction 3 on IMAC30.

CM10 ProteinChip Arrays, which are weak cationic exchangers, have a negatively charged surface that makes them bind positively charged proteins at a given pH. IMAC30 ProteinChip Arrays bind proteins by interacting with specific amino acid residues (e.g. histidine) or post-translational modifications (e.g. phosphate groups) on their surface through metal ions (e.g. Cu²⁺).

For ProteinChip Array binding, the spots on the IMAC30 arrays were preloaded with 50 µL CuSO4 (0.1 M). The CM10 and IMAC30 arrays were equilibrated with 150 µL binding buffer. 10 µL of each fraction was diluted in 90 µL of specific binding buffer, added to each well of the bioprocessor and mixed for 30 minutes with vigorous shaking.
on a MicroMix. After removing the remaining sample, the arrays were washed and mixed three times with 150 µL binding buffer for 5 min followed by two short washes with 150 µL de-ionized water. All steps were performed at room temperature.

An energy absorbing matrix containing sinapinic acid (SPA) was prepared according to the recommendations of the manufacturer [5 mg/vial dissolved in 400 µL solution (200 µL Acetonitrile; 200 µL of 1% Trifluoroacetic Acid in HPLC grade H₂O)] and, after the arrays were dry, 1 µL SPA solution applied on each spot. This was repeated once and the ProteinChip arrays were stored at room temperature in the dark until used.

**Reading of Arrays**

The arrays were read on a Protein Biosystem IIc (PB IIc) instrument, which was calibrated externally using the Protein MW standards kit E100-0001 (Ciphergen Biosystems) loaded on an NP-20 array. Each array was run twice with two different laser intensities to achieve better resolution for low- and high-molecular-mass proteins.

**Data Analysis**

Serum protein profiles of CL patients and healthy controls were compared with each other to detect biomarkers for CL (this was done principally by comparing pretreatment samples against healthy controls, and the paired follow up samples against the same healthy controls). Peaks were first auto-detected using Ciphergen Express Data manager 2.1 (1st pass analysis) and each peak was visually inspected for the 2nd pass analysis.

Spectra were baseline subtracted and normalized to the total ion current from 2000 Da to 100 000 Da (low energy settings) and 10 000 Da to 200 000 Da (high energy settings). Spectra that were not within the “twice the average” QC rule were deleted. For peak cluster detection, EDM (expression difference mapping) analysis was performed. Peaks were auto-detected in a 0.3% cluster mass window for low energy settings and 2% cluster mass window for high energy settings. Biomarker Patterns Software (BPS; Ciphergen
Biosystems) was used to generate diagnostic classification trees (algorithms) by comparing pre-treatment sera with the controls and post-treatment sera with the same controls.

BPS uses the peak information generated by the training set of known samples to build a binary decision tree algorithm. The algorithm functions by assigning each sample in the data set into 1 of the 2 groups or nodes with a rule based on the intensity of a particular peak or splitter. Each sub node has a different rule that further divides the data set and this process continues until all cases are assigned into terminal nodes. This results in correct classification percentages of the so-called “learn set.” The software generates and tests the models, using a process of cross-validation by randomly picking 10% of the samples. The peaks that formed the main splitters of the tree with the highest prediction rates in the cross-validation analysis were then selected to make a final decision tree with the greatest possible predictive power and this results in correct classification percentages of the so-called “test set” (Bons et al., 2007).

**Identification**

50 µL of the fractionated sample that showed the highest concentration of targeted candidate biomarkers as well as 50 µL of the fractionated control sample that showed the lowest concentration of the same particular biomarker were desalted and 10 µL of each sample loaded on a 12% Bis-Tris gel. Each positive sample was compared to its control. Gels were fixed with fixer solution for 2 hours and afterwards stained with Coomassie overnight to visualize the bands. De-staining with distilled water was performed until the bands were clearly visible.

**Sequencing of the proteins**

Bands were cut from the gels and sequenced using LC-MS-MS followed by a Mascot database search (performed by Dr Bernard Gibbs at the Sheldon Biotechnology Centre).
The principle of LC-MS-MS lies in the separation of a mixture of peptides (resulting from protein digestion with trypsin) by one-, two- or three-dimensional LC and measurement of peptide masses by MS-MS. The advantage of that technique is the direct identification of several hundred to more than 1000 (for three-dimensional) proteins per sample by MS-MS of peptides. The disadvantages are that it is a low throughput (one sample per run) and time consuming method. Additionally there can be problems regarding the detection by MS-MS because often it is not comprehensive, thus complicating the comparison of different samples (Engwegen et al., 2006).

Results

The patient population consisted of 2 women and 16 men with ages between 20 and 44 years. The number of lesions ranged between 1 and 9 and the mean duration of the lesions was 124 days. The size of the lesions had a huge range between 4.5 and 2100 mm² (see table 3). None of the tested subjects had mucocutaneous leishmaniasis. Clinical cure (meaning complete epithelialization and no new lesions) was not achieved in the CL patients on treatment day 20. 11 patients were cured after 2 months, 4 patients after 6 months and 2 patients after 12 months of completing the treatment (these data were generally assessed by patient interview). For one subject follow up data was not available.

<table>
<thead>
<tr>
<th>Patients</th>
<th>Lesion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age (years)</td>
</tr>
<tr>
<td>Range</td>
<td>20-44</td>
</tr>
<tr>
<td>Average</td>
<td>29</td>
</tr>
</tbody>
</table>

Table 3: Characteristics of the tested CL patients.

Pre-treatment Samples

Table 4 shows the molecular weights of all 25 preliminary biomarkers found in pre-treatment samples of CL patients (selected versus the 4 control sera) after the 2nd pass analysis based on p- and ROC-values. Highlighted molecular weights indicate proteins that were the most useful biomarkers in the BPS analysis (for decision tree building).
Table 4: 25 biomarkers detected in pretreatment samples of CL patients (F = fraction; C = CM10; I = IMAC30; L = low energy settings; H = high energy settings).

The p-values of the 25 biomarkers ranged between 0.05 and 0.001. Fraction 4 bound on a CM10 array showed the highest number of expression differences (5 biomarkers detected under low energy and 9 biomarkers under high energy settings). The highest MW-range of detected biomarkers was also seen in F4CH (from 10 kDa to 161 kDa). Only one biomarker was detected in F1 and F2 bound on CM10 arrays, and in F2 bound on IMAC30 arrays.

**Cluster Plots**

The cluster plots illustrate the distribution of the intensities for each m/z value. With an increased number of significant m/z values (p<0.05) and superior cluster plots, the chance to detect a potential biomarker in a larger sample set is theoretically higher. A cluster plot is superior when there is minor or no overlap between the data points of the leishmaniasis patients compared to the control samples, which means that there is good discrimination between both groups (Bons *et al.*, 2007). Figure 1 shows a m/z scatter plot of the 56.6 kDa biomarker found in F4CH of pretreatment sera from CL cases. Each dot represents one serum sample.
The above shown m/z scatter plot of the biomarker with a MW of 56 669 Da successfully distinguishes between the positive (pre-treatment) and negative (control) samples (ROC value of 1.0). The ROC (Receiver Operator Characteristic) is a calculation used to indicate the value of the individual peak to distinguish between infected and non-infected samples (ROC values of 1 or -1 are the best possible results). The p-value of 0.002 is excellent given the small number of control samples available, however the intensity of this protein was low (<0.175) compared to the proteins from our VL study which had intensities up to 36 (Fussi et al., in preparation).

Figure 2 shows the m/z scatter plot of the 47.7 kDa biomarker found in F4CL of pre-treatment sera. The m/z scatter plot of the 47.7 kDa biomarker also successfully distinguishes between the positive and negative samples (ROC value of 1.0). The p-value is also very good (0.002) but again, the intensity of the protein is low (<0.065).
Fig 2: M/z scatter plot of the 47773 Da biomarker.

**Follow-up Sera**

Table 5 shows that only 4 preliminary biomarkers were detected with CiphergenExpress in samples from patients treated for 14 days with sodium stibogluconate (20 mg/kg/day). Highlighted molecular weights indicate proteins that were the most useful biomarkers in the BPS analysis (for decision tree building).

Table 5: 4 biomarkers detected in follow up sera from CL patients treated for 14 days with sodium stibogluconate.

Table 6 presents all 82 preliminary biomarkers which were found in sera from patients treated for 20 days. Proteins that were the most useful biomarkers in the BPS analysis (for decision tree building) are again highlighted.
Table 6: 82 biomarkers detected in follow up sera of CL patients treated with sodium stibogluconate (20 mg/kg/day) for 20 days.

82 biomarkers were found using CiphergenExpress, but only 2 of the 82 biomarkers were detected using BPS. 77 expression differences were detected in fraction 4 bound on a CM10 array read under low energy settings. The MWs of the biomarkers ranged between 4 kDa and 83 kDa in fraction 4, and 2.5 kDa and 72 kDa in fraction 3. None of the candidate disease-specific biomarkers found in pre-treatment sera (vs controls) were detected in the same persons on the same array-type after 14 or 20 days of treatment.

**Classification Trees (Biomarker Pattern Software BPS)**

**Pre-treatment Sera**

Figure 3 shows a classification tree that needs only one biomarker (47 776 Da) to correctly classify all 18 positive pre-treatment sera from the 4 controls. It shows that if the intensity of the 47.7 kDa protein is \( \leq 0.025 \) all 4 controls are correctly classified as negative, whereas if the protein intensity is \( > 0.025 \) all 18 pre-treatment ‘case’ sera are correctly classified as positive.
Fig 3: Decision tree using one biomarker with a MW of 47 776 Da to classify pre-treatment sera from controls.

Follow-up Sera (patients treated for 20 days)

In figure 4 we see a classification tree that needs also only one biomarker (5024 Da) to correctly classify all 14 positive sera from the 2 controls. If the intensity of the 5024 Da biomarker was <= 0.581 the 2 negative controls were correctly classified, if the protein intensity was > 0.581 all 14 sera from leishmaniasis patients treated for 20 days were classified as positive. Using the 5 kDa biomarker alone makes it possible to correctly classify all positive samples treated for 20 days and the negative controls.
Correlation between Peak Intensity and Clinical Data

In general, peak intensities for most of the candidate biomarkers were very low. There was no correlation between the clinical data (duration, location, appearance, number and size of the lesion(s), lymphadenopathy, age, race and gender) and the peak intensities of the biomarkers.

Pre-treatment versus Post-Treatment Samples

Because pre-treatment as well as post-treatment sera were compared to the same controls, different biomarkers detected in these actions show also the differences between the pre- and post-treatment sera (see also table 4, 5 and 6). For example, the 5024 Da protein in F4CL from sera treated for 20 days was not detected in F4CL of pre-treatment sera or patients treated for 14 days. Therefore this marker appeared de novo after 20 days of treatment. None of the biomarkers found in pre-treatment samples (see table 4) were detected in the samples taken on treatment day 14 (see table 5) or 20 (see table 6) on the same array type. A 43 kDa protein was detected in fraction 4 on IMAC30 arrays in pre-treatment sera (see table 4). In sera treated for 20 days a 42.9 kDa protein was detected in fraction 4 but on CM10 arrays (see table 6).
Identification

SDS-PAGE was used for isolation and purification of the potential biomarkers and LC-MS-MS for their identification. We attempted to identify all of the biomarkers (see table 7) detected in the 2nd pass analysis (using CiphergenExpress) and BPS (9 biomarkers in pre-treatment samples, 3 biomarkers after 14 days of treatment and 2 biomarkers after 20 days of treatment).

<table>
<thead>
<tr>
<th>Pretreatment-sera</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1CL</td>
</tr>
<tr>
<td>10281</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sera treatment 14 d</th>
</tr>
</thead>
<tbody>
<tr>
<td>F3IL</td>
</tr>
<tr>
<td>2875</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sera treat. 20 d</th>
</tr>
</thead>
<tbody>
<tr>
<td>F3IL</td>
</tr>
<tr>
<td>2535</td>
</tr>
</tbody>
</table>

Table 7: Targeted and visible biomarkers on gels (the numbers represent the MWs of the proteins; highlighted numbers indicate the visible markers).

Bands at the appropriate molecular weight were visible on the gels for only 3 of these 14 biomarkers. The highlighted MWs in table 7 indicate these proteins (one at 43 kDa in pre-treatment sera, a 13.9 kDa band in sera treated for 14 days, and a 5 kDa band in sera treated for 20 days).
In figure 5 the numbers underneath the gel-lanes indicate the MWs of the biomarkers we were trying to visualize in that particular lane. C indicates lanes loaded with the fraction of the specific control sample. Gel 1 shows 4 targeted biomarkers (120 901 Da, 43 005 Da, 94 130 Da and 13 905 Da) with control sera (C) in the adjacent lane. Only 2 biomarkers (indicated with the boxes on the gel) were detectable.
Fig. 6: Biomarkers visible on gel 2.

Gel 2 shows that 1 biomarker (indicated with the boxes) with a MW of 5 kDa was detected on the gel (a faint band was visible on the original gel; unfortunately the photo doesn’t show the band that well), whereas the other 5 targeted biomarkers (47773 Da, 2535 Da, 56669 Da, 28185 Da and 2875 Da) were not visible on the gel.

Microsequence analysis of the excised 43 kDa band (shown in fig. 5) followed by a Mascot database search resulted in the list of potential proteins shown in table 8.

<table>
<thead>
<tr>
<th>Mascot search results for the 43 kDa protein found in CL</th>
<th>Coverage (%)</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complement C3 precursor</td>
<td>26</td>
<td>Human</td>
</tr>
<tr>
<td>Serum albumin</td>
<td>26</td>
<td>Human</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td></td>
<td><em>Macaca fascicularis</em></td>
</tr>
<tr>
<td>C4B1 (complement component C4B)</td>
<td>21</td>
<td><em>Homo sapiens</em></td>
</tr>
<tr>
<td>Complement factor H precursor, short</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mascot search results for the 43 kDa protein found in CL-pre-treatment sera.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------</td>
<td>----------------</td>
</tr>
<tr>
<td>Coverage (%)</td>
<td>Origin</td>
<td></td>
</tr>
<tr>
<td>Serum albumin, chain A</td>
<td>38</td>
<td>Human</td>
</tr>
<tr>
<td>Transferrin precursor</td>
<td>28</td>
<td>Human</td>
</tr>
<tr>
<td>Haptoglobin precursor, allele 1</td>
<td>28</td>
<td>Human</td>
</tr>
<tr>
<td>Hypothetical protein DKFZp459H0229</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haptoglobin precursor, allele 2</td>
<td>23</td>
<td>Human</td>
</tr>
<tr>
<td>Brain cDNA, clone: QccE-13766</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum albumin precursor</td>
<td></td>
<td>Cat</td>
</tr>
<tr>
<td>Serum albumin precursor</td>
<td></td>
<td>Bovine</td>
</tr>
<tr>
<td>MUSHAPTGB NID</td>
<td>Mus saxicola</td>
<td></td>
</tr>
<tr>
<td>Hemoglobin beta chain</td>
<td>Gelada baboon</td>
<td></td>
</tr>
</tbody>
</table>
Table 9: Mascot search results for the 13.9 kDa protein found in CL follow up sera treated for 14 days.

The 13.9 kDa protein was up-regulated in the positive samples. Our literature search showed a possible connection between leishmaniasis and haptoglobin precursor (highlighted in table 9).

Discussion

The diagnosis of CL is still a challenge. New non-invasive diagnostic tests with high sensitivity and specificity are needed.

SELDI-TOF-MS represents an emerging proteomic technology in biomarker discovery that allows the rapid and sensitive analysis of complex protein mixtures (Hutchens & Yip, 1993; Merchant & Weinberger, 2000; Davies, 2000; Issaq et al., 2002; Verma et al., 2001). This relatively limited study demonstrates the potential SELDI-TOF-MS has in searching for new biomarkers in the serum of CL patients.

With proteomics defined as the study of all proteins in a biological system, it is obvious that much is demanded from the analytical technologies used. More than 500,000 proteins have been estimated to comprise the human proteome (Banks et al., 2000). The bulk of the easily accessible serum proteome consists of only a small number of abundant proteins such as albumin and immunoglobulins. The presence of these proteins complicates the detection of the many low-abundance proteins. Thus, the major tasks in any biomarker discovery program are to separate and analyze the whole proteome and to process the massive amount of data into meaningful results using statistics and bioinformatics (Engwegen et al., 2006). Protein profiling increases the chance of finding proteins with altered expression levels during disease development, progression or treatment. Additionally, protein profiling makes it possible to combine several of these proteins into a discriminative protein pattern (Engwegen et al., 2006).
The potential for SELDI-based proteomics studies has been explored in various medical disciplines [e.g. infectious diseases (Bernhard et al., 2005), Alzheimer’s disease (Ho et al., 2005) and cardiovascular diseases (Anderson, 2005)].

The current study shows that SELDI-TOF-MS can be successfully applied to the search for biomarkers in serum from CL patients. When only CiphergenExpress was used, the largest number of biomarkers was detected in sera from patients treated for 20 days, whereas using both software programs the greatest number was found in pre-treatment sera. We were surprised to get m/z scatter plots with a ROC of 1 or close to it with only 4 control sera at our disposal. These biomarkers were able to distinguish successfully between positive and negative serum samples. In each serum-class (pre-treatment, treated for 14 days and 20 days) it was possible to successfully detect one biomarker on gel.

Because none of the biomarkers found in pre-treatment sera were detected in the same persons after 14 or 20 days of treatment on the same array type it seems that these proteins could be used as pre-treatment markers for CL infections. It is unclear if the markers found in the samples from treated people are usable as markers for treatment progress because lesions were present in the patients after treatment ended [None of the subjects achieved clinical cure by treatment day 20. 11 patients were reported as cured after 2 months of completing the treatment, 4 patients after 6 months and 2 patients after 12 months (cure means complete epithelialization and no new lesions)] and the duration of the lesions showed no correlation to the peak intensity.

Sequencing with LC-MS-MS followed by a Mascot search revealed more than 1 protein for the 43 kDa band found in pre-treatment sera (see table 8). Ceruloplasmin was one of the proteins found (see table 8). It is an acute phase protein and according to Cooper (1990), acute phase proteins are specific indicators of tissue damage. The acute phase response is part of the innate defense system of an animal against trauma, inflammation and infection (Gabay & Kushner, 1999). During this response, an increase in the synthesis rate and release of certain plasma proteins such as haptoglobin, C-reactice protein and
ceruloplasmin can be seen. At the same time a decrease in the production of other plasma proteins such as albumin can be detected (Eckersall & Conner, 1998; Toussaint et al., 1995; Eckersall, 2000). Martinez-Subiela et al. (2002) report in their study that concentrations of acute phase proteins were significantly higher in dogs with leishmaniasis than in control dogs. In our study the 43 kDa protein was down regulated in the positive samples. Therefore, if the 43 kDa protein represents ceruloplasmin, our findings show the opposite from the ones found in canine leishmaniasis by Martinez-Subiela et al. (2002), whereas if this 43 kDa band represents albumin, our findings support the ones from Eckersall & Conner (1998).

Another interesting human protein found by the Mascot search is complement C3 precursor. The complement system helps to clear pathogens from the organism by disrupting the target cell’s plasma membrane. Metacyclic Leishmania promastigotes can activate complement in both the classical and the alternative pathways (Dominguez et al., 2003). Complement activation results in binding of C3bi (and other complement factors) to the parasite surface. This process is called ‘opsonization’. Leishmania uses this opsonization to escape from the hostile environment by promoting phagocytosis via complement receptors (CR). Parasites coated with C3bi bind to CR1 on erythrocytes (immune adherence). When the parasites encounter macrophages they bind to CR3 on the macrophage surface which facilitates the uptake of the parasites into their major host cell. Within these phagocytes Leishmania can transform into the obligate intracellular life form, the amastigote (von Stebut, 2007). Because the 43 kDa protein was down-regulated in our tested pre-treatment sera, it could mean that complement-precursor was activated in our patients and therefore is probably used in its active form for the ‘opsonization’ process.

Of all possible proteins revealed by the Mascot search another plausible connection in the literature exists between ApoA-I and CL. Therefore this band could show a host protein called apolipoprotein a-I lipid-binding domain [ApoA-I is an apolipoprotein and the major protein component of high density lipoprotein (HDL) in plasma. The protein promotes cholesterol efflux from tissues to the liver for excretion. It is a cofactor for
lecithin cholesterolacyltransferase (LCAT) which is responsible for the formation of most plasma cholesteryl esters. ApoA-I was also isolated as a prostacyclin (PGI2) stabilizing factor, and thus may have an anticoagulant effect (Yui et al., 1988).

Apolipoprotein was down-regulated in the pre-treatment sera compared to healthy controls. This would make sense because Shamshiev et al (2007) found in mice that dyslipidemia (which is a disruption in the amount of lipids in the blood) inhibited Toll-like receptor (TLR)-induced production of proinflammatory cytokines, including interleukin (IL)-12, IL-6, and tumor necrosis factor-α, as well as up-regulation of costimulatory molecules by CD8α DCs, but not by CD8α⁺ DCs, in vivo. Decreased DC activation profoundly influenced T helper (Th) cell responses, leading to impaired Th1 and enhanced Th2 responses. As a consequence of this immune modulation, host resistance to Leishmania major was compromised (Shamshiev et al., 2007). In the literature nothing was found published about clinical evidence for dyslipidemia in human CL cases, whereas in VL reports about patients have been made. Erdeve et al (2004) describe 2 children who had hypertriglyceridaemia, but no family history of hyperlipidemia. Elnour et al (2001) also described hypertriglyceridaemia as a new observation in VL infected Omani children. In our pre-treatment sera, a band of 43 kDa (detected on IMAC30) was down-regulated. If this band represents apolipoprotein, our findings could show that in our CL cases dyslipidemia was also present. It might be that patients developed dyslipidemia because of the CL-infection. The patients were cured between 2 and 12 months after completing treatment. That means on treatment day 20 all the patients still had lesions and were not cured. To verify if dyslipidemia was caused by the parasite in these patients it would help to investigate if sera taken after the lesions healed still show the down-regulated 43 kDa band.

In patients treated for 20 days a protein with 42.9 kDa was detected, but on a different chip type (CM10). This protein was not present in pre-treatment sera on that array type. Therefore to exclude the possibility that the 42.9 kDa protein represents ApoA-I it has to be isolated, purified and identified in the future. There is no evidence in the literature that treatment with sodium stibogluconate influences ApoA-I levels in humans.
Shamshiev et al (2007) mention also that perturbation of lipoprotein homeostasis by environmental or genetic factors leads to dyslipidemia (for instance an excessive consumption of lipid-laden foods). Since the positive and negative serum samples were drawn from people exposed to the same environment (US-army) we assume that they had the same food available. This means it might be less likely that dyslipidemia was caused by different nutrition in our tested persons.

When we looked at the up-regulated 13.9 kDa protein found in patients treated for 14 days, haptoglobin precursor was a potential candidate found in the Mascot database search. If the 13.9 kDa protein represents haptoglobin precursor, it could mean that less haptoglobin was present in the patient, because its inactive form was upregulated. This finding would also be contrary to the results of the canine leishmaniasis study from Martinez-Subiela et al (2002). However it is also possible that any other human protein from the lists (table 8 and 9) could serve as a biomarker for leishmaniasis.

In comparison to the results we obtained from our VL study, far fewer biomarkers were detected in patients with CL (Fussi et al, in preparation). This may be due to the fact that CL (in contrast to the visceral form) is not a systemic disease.

Because only a small number (4) of control sera were available to us, more negative serum samples will be needed in future studies. However, even with this limited number of control sera we still found biomarkers of very good quality with low p-values. Furthermore, our analysis of pre-treatment samples with the paired follow-up samples (presumably ‘controls’) demonstrated that none of the biomarkers found in pre-treatment samples (against the controls) were detected (against the same controls) during or after treatment in the same persons on the same chip-type.

Additional studies are required to assess the broader sensitivity and specificity of the biomarkers we have discovered (e.g.: CL samples from other geographic areas should be tested, as well as sera from patients with other diseases). Because the biomarkers seem to
be of host origin, it will be necessary to determine what other pathological or non-pathological conditions modulate the level of these proteins in the sera. Also, simpler assays will be needed to detect the biomarkers (e.g.: ELISA) in the future if this knowledge is to be applied to CL diagnosis in low resource settings.

Our preliminary study suggests that SELDI-TOF-MS may be of great potential use in the search for new biomarkers in serum samples from CL patients. Nevertheless, it will be necessary to test sera from more patients from other geographic regions and people with other diseases to validate our findings and to check for specificity. If validated in more comprehensive studies, this discovery will have a major impact on future (cutaneous) leishmaniasis diagnosis and treatment monitoring.
Literature Article 1


Contributions of Authors

Naomi Aronson provided the sera and information regarding the subjects studied. Christine Straccini ran the gels and helped cutting the bands. Bernard Gibbs sequenced the proteins and did the Mascot search. Adam Rainczuk assisted with conducting the practical tests and analyzing the results. Greg Matlashewski and Brian Ward were the supervisors of Manfred Fussi, who did the practical and theoretical work described in this article during his M.Sc. studies. Momar Ndao supported Manfred Fussi with all aspects of the project.
Connecting Statement

The discovery of biomarkers using SELDI-TOF-MS in serum from CL patients made us question if it would be possible to find biomarkers in VL patients using the same technology. Since the visceral form of the leishmaniasis disease complex is the most severe and diagnosis still has its limitations, the discovery of serum biomarkers could be of tremendous importance.

Unfortunately well-characterized and carefully-handled samples from VL cases are very hard to obtain. Therefore we were very lucky to establish a collaboration with Dr Nirmal Baral from the Koirala Institute in Nepal, who provided us with serum samples from VL patients and control sera from the same geographical area.
CHAPTER 2:
USING SELDI-TOF-MS TO DISCOVER BIOMARKERS IN VISCERAL LEISHMANIASIS PATIENTS

Manfred Fussi¹, Christine Straccini¹, Bernard Gibbs⁴, Nirmal Baral², Greg Matlashewski³, Momar Ndao¹ and Brian J. Ward¹

National Reference Centre for Parasitology, Montreal General Hospital, McGill University, Montreal, Quebec, Canada¹
B.P.Koirala Institute, Nepal²
Department of Microbiology & Immunology, McGill University, Montreal, Quebec, Canada³
Sheldon Biotechnology Centre, McGill University, Montreal, Quebec, Canada⁴
(Paper in Preparation)

Introduction

Visceral leishmaniasis (VL), commonly known as “kala-azar”, is caused by *L. donovani* and *L. infantum* in the Old World and *L. chagasi* in the New World. Desjeux (1996) reports that VL has an estimated yearly incidence of 500,000 cases. More than 90% of the reported VL occurs in India, Nepal, Sudan and Brazil (Desjeux, 2004). VL is also a serious problem in HIV-positive individuals. It is the fourth most common opportunistic parasitic disease in people with HIV in Spain (Paredes *et al.*, 2003). In Ethiopia, Sudan and Southern Europe as many as 70 per cent of adults with VL are also HIV positive (Desjeux, 2001). VL is characterized by fever, cachexia, hepatosplenomegaly, and hematologic cytopaenias, and is usually fatal without specific chemotherapy (Saha *et al.*, 2006). The name “kala azar” is thought to have originated from India, meaning “black fever”, which refers to the hyperpigmentation of skin during the course of disease. Alternatively, the term might be derived from the word “kal” meaning “death”, which signifies the fatality of the disease (Brahmachari UN, 1928). Rare patients, cured of VL in Sudan and India, can develop post-kala azar dermal leishmaniasis (PKDL), which appears as a dermatotropic form of *L. donovani* infection (Saha *et al.*, 2006).

A recently-recognized problem associated with the visceral form of leishmaniasis is the fact that it can be transmitted via blood. Transfusion-transmitted leishmaniasis has been reported from many countries including India (Singh *et al.*, 1996) in association with
visceral disease but not from CL. For such transmission to occur the parasites must be present in the peripheral blood of the donor (who is almost always asymptomatic) and these parasites must survive processing and storage in the blood bank to infect the recipient. Asymptomatic infection with VL generally does not persist for more than one year, but rare individuals can remain asymptomatic for decades (Guevara et al., 1993).

Direct visualization of the parasites is still the gold standard in Leishmania diagnostics due to the high specificity of this approach (Herwaldt, 1999). However, the sensitivity of microscopy and culture can be low and highly variable (Herwaldt, 1999), depending on the sampling procedure, number and dispersion of parasites in biopsy samples, and the technical skills of the personnel.

Kala azar may be confused with other similar conditions such as tropical splenomegaly, malaria, schistosomiasis or cirrhosis with portal hypertension, milliary tuberculosis, brucellosis, African trypanosomiasis, typhoid fever, bacterial endocarditis, histoplasmosis, malnutrition, leukaemia and lymphoma (Singh & Sivakumar, 2003). To make the diagnosis of VL in immunocompetent patients, the best samples for aspiration are from the spleen where a sensitivity > 94 percent can be achieved (Singh, 2006). This procedure is risky however, and even in the most experienced hands, the risk of fatal bleeding cannot be eliminated (Bryceson, 1987). Another option is the demonstration of the parasites in liver aspirates and biopsies, but the sensitivity can be as low as 40 per cent. Additionally liver aspiration also has to be performed very carefully to minimize the bleeding risks for the patient (Singh, 2006). A much safer but painful method is the examination of bone marrow obtained from iliac or sternal crest puncture (Singh, 2006). In VL, the hallmark is hyperimmunoglobulinaemia. Tests to measure serum VL-specific antibodies have been developed but lack sensitivity and/or specificity (Singh & Sivakumar, 2003; Singh et al., 2005; Boelaert et al., 2004).

The successful discovery of new biomarkers in different human diseases has been realized in recent years using a range of proteomic approaches. Specifically, SELDI-TOF-MS has been used in the discovery of biomarkers not only in ovarian (Kozak et al.,...
If biomarkers for VL can be measured in an accessible body fluid such as serum that not
requires a tissue biopsy, this would have a major impact on future VL diagnosis and
treatment monitoring. This study focused on the discovery of potential serum biomarkers
for the diagnosis of VL using SELDI-TOF-MS.

**Materials and Methods**

**Sera**

A total of 64 sera were available for this study which were split into set 1 (consisting of
21 pretreatment VL sera and 19 healthy controls from the same geographic region) used
to find preliminary biomarkers and set 2 (24 pretreatment sera) as a validation set. The
control sera were from the same geographic region as the patients sera. They were
negative for malaria and tuberculosis and the control subjects were free of fever during
sample collection. These subjects were not specifically ‘matched’ for demographic
characteristics with the cases. The control samples were sex and age matched to the pre-
treatment VL sera. The material was obtained from Dr Nirmal Baral, Koirala Institute,
Nepal.

The clinical information available for the positive samples included: duration of fever and
cough, hepatomegaly (size), splenomegaly (size), BM aspirate, albumin level, bilirubin
level and Hb.

Samples were aliquoted (30 µL/tube) and stored at -20°C until further use. All
fractionated serum samples were thawed a maximum of 3 times between collection and
analysis.


**Fractionation and Binding of Sera**

This was done using Cyphergen’s Expression Difference Mapping ™ Kit-Serum Fractionation and handled in the same way as described in the CL-study on page 37.

Pre-treatment sera and controls were bound on 2 different array surfaces (ProteinChip™: Ciphergen Biosystems Inc.), the weak cation exchange chip (CM10) and immobilized metal affinity capture coupled with copper (IMAC-Cu$^{2+}$) chip arrays (for a description of the arrays please see page 37).

To find the optimal fractions for each array type (CM10 and IMAC30), all 6 fractions from 8 randomly chosen sera (pre-treatment and controls) were loaded on both array surfaces (ProteinChip™: Ciphergen Biosystems Inc.). The best 3 fractions (that produced the highest number of peaks with good intensity) for each array type were chosen to be bound from all samples (on CM10 fraction 1, 3 and 4; on IMAC30 fraction 2, 5 and 6). ProteinChip Array binding was executed in the way as described on page 37 of the CL study.

**Reading of the Arrays**

The same procedure was followed as described in the CL-study on page 38.

**Data Analysis**

Serum protein profiles of VL patients and healthy controls were compared with each other to detect biomarkers for VL. Peaks were first auto detected using CiphergenExpress Data manager 2.1 [1st pass analysis calculates p-values (p < 0.05) and ROC (Receiver Operator Characteristic: calculation used to indicate how good the peak is to distinguish between infected and non-infected samples) from autodetected peaks] and each peak visually inspected for the 2nd pass analysis (p-values < 0.001, recalculates p-values and ROC from the peaks manually relabeled).
Spectra were baseline subtracted and normalized as performed in the CL-study on page 38. Peak cluster detection and Biomarker Pattern Software was handled in the same way as in our previously described CL study (see page 38).

**Identification**

80 µL of the fractioned sample that showed the highest concentration of targeted candidate biomarkers as well as 80 µL of the fractionated control sample that showed the lowest concentration of the same particular biomarker were desalted and 10 µL of each sample loaded on a 12% SDS-PAGE. Each positive sample was compared to its control. Gels were fixed with fixer solution for 2 hours and afterwards stained with Coomassie overnight to visualize the bands. De-staining with distilled water was performed until the bands were clearly visible.

**Sequencing of the Proteins**

The cut gel bands were sequenced from Dr Bernard Gibbs (Sheldon Biotechnology Centre) using LC-MS-MS. A Mascot database search followed.

**Results**

The tested patients consisted of 25 men and 20 women between 10 and 50 years of age. Their liver measured between 2 and 10 cm and the spleen between 3 and 25 cm. Fever was reported between 9 and 270 days, whereas cough was noticeable up to 120 days (see table 1). The controls consisted of sera from 6 women and 13 men between 10 and 44 years of age.

<table>
<thead>
<tr>
<th>Patients</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age (years)</td>
</tr>
<tr>
<td>Range</td>
<td>10-50</td>
</tr>
<tr>
<td>Average</td>
<td>27.5</td>
</tr>
</tbody>
</table>

Table 1: Characteristics of the tested VL patients.
Table 2 shows the molecular weights of all 212 biomarkers found in pre-treatment samples of VL patients after the 2\textsuperscript{nd} pass analysis. Highlighted molecular weights indicate proteins that were the most useful biomarkers in the BPS analysis (for decision tree building).

Table 2: 212 biomarkers detected in pre-treatment samples of CL patients after 2\textsuperscript{nd} pass analysis (F = fraction; C = CM10; I = IMAC30; L = low energy; H = high energy).

<table>
<thead>
<tr>
<th>F1CL</th>
<th>F1CH</th>
<th>F3CL</th>
<th>F3CH</th>
<th>F4CL</th>
<th>F4CH</th>
<th>F2IL</th>
<th>F2IL</th>
<th>F3IL</th>
<th>F3IL</th>
<th>F4IL</th>
<th>F4IL</th>
<th>F5IL</th>
<th>F5IL</th>
<th>F6IL</th>
<th>F6IL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2926</td>
<td>10406</td>
<td>3269</td>
<td>15211</td>
<td>3371</td>
<td>14850</td>
<td>2502</td>
<td>4745</td>
<td>12299</td>
<td>3162</td>
<td>10559</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2560</td>
<td>10452</td>
<td>3539</td>
<td>20474</td>
<td>3379</td>
<td>14708</td>
<td>2726</td>
<td>4509</td>
<td>12333</td>
<td>3375</td>
<td>11053</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3248</td>
<td>10456</td>
<td>3562</td>
<td>51327</td>
<td>4139</td>
<td>14670</td>
<td>3536</td>
<td>9478</td>
<td>12610</td>
<td>3552</td>
<td>11074</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3372</td>
<td>10448</td>
<td>4088</td>
<td>4031</td>
<td>16626</td>
<td>3549</td>
<td>13917</td>
<td>12678</td>
<td>3599</td>
<td>11128</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3516</td>
<td>10507</td>
<td>4196</td>
<td>6018</td>
<td>16736</td>
<td>3562</td>
<td>33646</td>
<td>12735</td>
<td>4261</td>
<td>11161</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3634</td>
<td>10556</td>
<td>4302</td>
<td>13501</td>
<td>16277</td>
<td>3743</td>
<td>65962</td>
<td>14586</td>
<td>4253</td>
<td>11730</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3853</td>
<td>11797</td>
<td>4847</td>
<td>16431</td>
<td>17393</td>
<td>3746</td>
<td>71916</td>
<td>14703</td>
<td>4363</td>
<td>12238</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4301</td>
<td>12324</td>
<td>4970</td>
<td>28238</td>
<td>17482</td>
<td>4266</td>
<td>15675</td>
<td>4474</td>
<td>12307</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4968</td>
<td>12457</td>
<td>5585</td>
<td>51351</td>
<td>17591</td>
<td>4303</td>
<td>26333</td>
<td>4544</td>
<td>12393</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4525</td>
<td>12506</td>
<td>5785</td>
<td>51774</td>
<td>19605</td>
<td>4817</td>
<td>49192</td>
<td>4631</td>
<td>12483</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4703</td>
<td>12753</td>
<td>5912</td>
<td>52003</td>
<td>26371</td>
<td>4852</td>
<td>49092</td>
<td>4765</td>
<td>12584</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4806</td>
<td>15914</td>
<td>81850</td>
<td>28535</td>
<td>5103</td>
<td>50581</td>
<td>5073</td>
<td>12625</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4826</td>
<td>16022</td>
<td>29387</td>
<td>51111</td>
<td>6200</td>
<td>12666</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4851</td>
<td>16139</td>
<td>50020</td>
<td>51595</td>
<td>6405</td>
<td>12786</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5176</td>
<td>16626</td>
<td>53167</td>
<td>52076</td>
<td>7295</td>
<td>12851</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5192</td>
<td>16850</td>
<td>56942</td>
<td>52721</td>
<td>7629</td>
<td>12931</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5264</td>
<td>27903</td>
<td>80097</td>
<td>54017</td>
<td>7580</td>
<td>13004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6594</td>
<td>29246</td>
<td>91845</td>
<td>66179</td>
<td>11551</td>
<td>13076</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16152</td>
<td>32326</td>
<td>117761</td>
<td>108685</td>
<td>13727</td>
<td>13590</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16550</td>
<td>33802</td>
<td>138627</td>
<td>133027</td>
<td>13907</td>
<td>14711</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17762</td>
<td>38039</td>
<td>14905</td>
<td>16767</td>
<td>16113</td>
<td>16825</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17854</td>
<td>40979</td>
<td>150222</td>
<td>17317</td>
<td>28213</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18584</td>
<td>41774</td>
<td>21821</td>
<td>26843</td>
<td>28104</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20952</td>
<td>44554</td>
<td>29218</td>
<td>29326</td>
<td>29004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44347</td>
<td>55196</td>
<td>50985</td>
<td>50252</td>
<td>50985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58900</td>
<td>60649</td>
<td>51350</td>
<td>50627</td>
<td>51350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66549</td>
<td>70207</td>
<td>53784</td>
<td>51218</td>
<td>53784</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73027</td>
<td>124504</td>
<td>60791</td>
<td>59346</td>
<td>60791</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>157018</td>
<td>62393</td>
<td>61861</td>
<td>61861</td>
<td>61861</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91819</td>
<td>94758</td>
<td>69621</td>
<td>69458</td>
<td>69458</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Altogether, 212 preliminary biomarkers (with p-values < 0.001) were detected after the 2\textsuperscript{nd} pass analysis using CiphergenExpress. Molecular weights ranged from 2.5 kDa to 157 kDa. The greatest number of biomarkers was found in fraction 6 bound on an IMAC30 array read under high energy settings. 15 of the 212 biomarkers were also detected using BPS (indicated with highlighted MW in table 2).
Cluster Plots

The cluster plots shown here illustrate how well some of the candidate biomarkers distinguished between the positive and negative samples. Fig. 1 shows a m/z scatter plot of the 3.3 kDa biomarker found in F6IL of pre-treatment sera from VL cases. Each dot represents one serum sample.

Fig. 1: M/z scatter plot of the 3375 Da biomarker.

The m/z scatter plot in fig. 1 of the biomarker with a MW of 3375 Da distinguished successfully between the positive (pre-treatment) and negative (control) samples (ROC value of 0.98). The ROC (Receiver Operator Characteristic) is a calculation used to indicate how well an individual peak distinguishes between infected and non-infected samples (ROC values of 1 or -1 are the best possible result). The p-value is very low, the intensity of the protein is very high (this biomarker had the highest intensity observed in our VL/CL studies).

Another very impressive but different view of that biomarker is presented in fig. 2. It clearly shows the difference between the pre-treatment samples and controls. Both gel view and peak view demonstrate the ability of the biomarker to distinguish between the 2 tested groups of sera.
Fig. 2: Peak and gel view showing the intensity of the 3375 Da biomarker in 4 pre-treatment sera and 4 controls.
Another biomarker found in pre-treatment sera is shown in figure 3. The m/z scatter plot of the biomarker with a MW of 12 463 Da found in F6IH also distinguished successfully between the positive (pre-treatment) and negative (control) samples (ROC value of 0.98). The p-value for this biomarker was also very low, and the intensity of the protein was high (in some samples over 2.5; see fig.3).

![M/z scatter plot of the 12 463 Da biomarker.](image)

**Classification Trees (Biomarker Patterns Software BPS)**

Biomarker Patterns™ Software is a Windows-based package for supervised classification of SELDI mass spectral data sets derived from the Ciphergen ProteinChip® platform. It identifies proteins that are relevant to a particular disease and summarizes and displays the results in a clear way. With this program, it is possible to discover protein multi-markers and to translate them into assays with high predictive accuracy. Protein multi-markers are combinations of single biomarkers that cannot distinguish between positive and negative samples on their own but can achieve near 100% separation when combined.

In figure 4 a classification tree is shown that combines three biomarkers (51 351 Da, 28 238 Da, 3378 Da) to correctly classify all 21 positive sera from the 19 controls.
Fig. 4: Decision tree using 3 biomarkers (51 351 Da, 28 238 Da and 3378 Da) found in F4CL to classify correctly between pre-treatment sera from VL patients and controls.

With the first biomarker in figure 4 (MW of 51 351 Da), 17 out of 21 positive samples were classified correctly. The second biomarker (28 238 Da) allowed us to classify a further 3 positive samples (out of the remaining 4) correctly, and finally with the 3rd biomarker (3378 Da) the remaining positive sample could be classified correctly. This meant that by combining the 3 biomarkers it was possible to successfully distinguish between all 21 positive and 19 negative samples.
A decision tree that only needed one biomarker to distinguish between positive and negative samples is shown in fig.5. If the intensity of the 12.4 kDa biomarker in figure 5 is <= 0.155 all 18 control sera are correctly classified, if the protein intensity is > 0.155 all 20 sera from VL patients are classified positive. This meant that by using only the 12.4 kDa biomarker it was possible to correctly classify all positive samples from the negative controls.

![Decision tree diagram](image)

Fig.5: Decision tree using one biomarker with a MW of 12.4 kDa (found in F6IH) to classify correctly between pre-treatment sera and controls.

**Correlation between Peak Intensity and Clinical Data**

No correlation between the clinical data (duration of fever, cough, hepatomegaly, splenomegaly, BM aspirate, albumin, bilirubin, Hb) and the peak intensities of any biomarker was detectable.

**Identification**

The potential biomarkers were isolated and purified using SDS-PAGE and identified with LC-MS-MS. We attempted to identify all of the 15 biomarkers (see table 3) detected in the 2nd pass analysis (using CiphergenExpress) and BPS.
Table 3: Targeted and visible biomarkers on gels (the numbers represent the MWs of the proteins; highlighted numbers indicate the visible markers).

Four of the 15 biomarkers targeted for identification were visible on the gels (2 down-regulated proteins: 28.2 kDa and 14.5 kDa, and 2 up-regulated proteins: 51.1 kDa and 12.4 kDa, see highlighted numbers table 3).

Fig.6: Biomarkers visible on gel 1.
The numbers underneath the gel-lanes indicate the MWs of the biomarkers we were trying to visualize in that particular lane. C indicates lanes loaded with the fraction of the specific control sample. Gel 1 shows 5 targeted biomarkers (27 603 Da, 51 327 Da, 28 238 Da, 51 351 Da and 14 580 Da) with their controls (C) next to them. Only 2 biomarkers (indicated with the boxes on the gel) were visible on the gel.

**Fig.7: Biomarkers visible on gel 2.**

Gel 2 shows 5 targeted biomarkers (9309 Da, 9478 Da, 12 610 Da, 51 111 Da and 12 463 Da) with their controls (C) next to them. Here as well only 2 biomarkers (indicated with the boxes and up-regulated in the positive sera) were detected on the gel. The band in the 12.4 kDa range of F6IH was faint but clearly visible on the original gel.
Even though the paired samples used to visualize the biomarkers in figure 6 and figure 7 showed several other more obvious differences, we focused on the candidate biomarkers based on molecular weight predictions, because these had been found in the entire sample set. On the gels, we compared fractionated samples at the extremes: i.e.: comparing the highest concentration of the particular biomarker (case or control) with the lowest concentration of the same biomarker (that means the other differences seen on the gel exist only in 1 patient sample and 1 control).

The next 4 tables (4-7) show the Mascot search results for the 4 visible biomarkers on gel.

<table>
<thead>
<tr>
<th>Mascot search results for the 51 kDa protein found in VL</th>
<th>Coverage (%)</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain, Human Serum Albumin</td>
<td>34</td>
<td>Human</td>
</tr>
<tr>
<td>Serpin peptidase inhibitor, clade A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unnamed protein product</td>
<td>26</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Unnamed protein product</td>
<td>23</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Anti-rabies SOJA immunoglobulin heavy chain</td>
<td>19</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Ig G1 H Nie</td>
<td>18</td>
<td>Human</td>
</tr>
<tr>
<td>Unnamed protein product</td>
<td>18</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Ig Alpha 1 Bur</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kininogen-1 precursor (Alpha-2-thiol proteinase inhibitor)</td>
<td>14</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Immunoglobulin heavy chain</td>
<td>11</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Unnamed protein product</td>
<td>10</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Immunoglobulin heavy chain</td>
<td>10</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Chain H, Immunoglobulin G1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haptoglobin</td>
<td>9</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Chain A, crystal structure of a human Fcg Receptor</td>
<td>9</td>
<td>Human</td>
</tr>
<tr>
<td>Albumin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum albumin precursor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preprohaptoglobin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha-1-acid glycoprotein 1 precursor</td>
<td></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Immunoglobulin heavy chain</td>
<td></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Immunoglobulin heavy chain variable region</td>
<td></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Immunoglobulin alpha heavy chain constant region</td>
<td></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Antithrombin III</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unnamed protein product</td>
<td></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Ig kappa chain C region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamin D-binding protein precursor (DBP)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4: Mascot search results for the 51.1 kDa protein found in VL sera.

The identification of the 51.1 kDa candidate biomarker revealed a list of potential proteins (see table 4). A literature search for links between VL and the human proteins from table 4 showed a possible connection to kininogen 1 precursor and immunoglobulins (proteins highlighted in table 4).

Table 5: Mascot search results for the 28 kDa protein found in VL sera.

For the 28 kDa protein a literature search for links between VL and the human proteins from table 5 showed a possible connection to immunoglobulins (highlighted in table 5).
Table 6: Mascot search results for the 14.5 kDa protein found in VL sera.

The literature search for the 14.5 kDa protein showed a link between leishmaniasis and hemoglobin (highlighted in table 6).

Table 7: Mascot search results for the 12.4 kDa protein found in VL sera.

Searching the literature for possible links between the proteins in table 7 and VL showed an interesting connection to proapolipoprotein and lipoprotein CIII.
Validation of the Biomarkers

To validate the 15 candidate biomarkers (found in sample set 1 using 2nd pass analysis/BPS) we tested the remaining 24 pre-treatment samples (under the same settings as set 1) and got across all fractions a maximum Dalton shift of ± 1% when the second set was run on a different day (see table 8).

<table>
<thead>
<tr>
<th></th>
<th>F1CL</th>
<th></th>
<th>F1CH</th>
<th></th>
<th>F3CL</th>
<th></th>
<th>F3CH</th>
<th></th>
<th>F4CL</th>
<th></th>
<th>F4CH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3534</td>
<td>1</td>
<td>3538</td>
<td>1</td>
<td>2760</td>
<td>1</td>
<td>2806</td>
<td>1</td>
<td>3562</td>
<td>1</td>
<td>3564</td>
</tr>
<tr>
<td>2</td>
<td>2862</td>
<td>2</td>
<td>8349</td>
<td>2</td>
<td>5132</td>
<td>2</td>
<td>5157</td>
<td>2</td>
<td>3379</td>
<td>2</td>
<td>3375</td>
</tr>
<tr>
<td>3</td>
<td>2283</td>
<td>3</td>
<td>2823</td>
<td>3</td>
<td>5135</td>
<td>3</td>
<td>5149</td>
<td>3</td>
<td>5151</td>
<td>3</td>
<td>5161</td>
</tr>
</tbody>
</table>

Table 8: Molecular weights of the biomarkers detected in sample set 1 and 2 (validation).

1: MW of biomarkers found in sample set 1
2: MW of biomarkers found in sample set 2 (validation)
n.d.: not detected

Fourteen of 15 preliminary biomarkers in set 1 were also detected in sample set 2. Only one (down-regulated) protein with a molecular weight of 9478 Da (found in set 1 in F5IL) was not detectable in the 2nd sample set. The second set was only used to verify if the biomarkers found in set 1 were also detected in set 2. We did not investigate new biomarkers revealed only in the second set, because their value was uncertain.
**Discussion**

VL is still not easy to diagnose. Therefore new diagnostic tests that are not invasive but high in sensitivity and specificity are needed.

A promising new proteomic technology that can be used in the discovery of biomarkers is SELDI-TOF-MS. With this technology complex protein mixtures can be fast and sensitively analyzed (Hutchens & Yip, 1993; Merchant & Weinberger, 2000; Davies, 2000; Issaq *et al.*, 2002; Verma *et al.*, 2001).

In this study we show that SELDI-TOF-MS can be successfully used to detect candidate biomarkers in VL patients.

Using CiphergenExpress on its own, the largest number of biomarkers was detected in fraction 6 bound on IMAC30 arrays, whereas using both software programs the largest number was found in fraction 4 on CM10 arrays. The ROCs in the m/z scatter plots show that the biomarkers were able to distinguish successfully between positive and negative serum samples.

In the past concerns about the reproducibility of SELDI-data were mentioned by several groups. Generation of reproducible SELDI-TOF spectra requires prompt separation of serum and storage at -20°C or below, and because repeated freeze-thaw cycles result in degradation of spectra (Papadopoulos *et al.*, 2004), it is necessary to use a robust cold chain (Agranoff *et al.*, 2005).

Ransohoff (2005) mentions that lessons may be learned from how the question of reproducibility has been effectively addressed in other “-omics” research. Rosenwald *et al* (2002) showed in a genomics study that by splitting the samples in a training set (to derive a pattern-recognition model) and an independent validation set, chance or over-fitting does not explain the results. They demonstrated reproducibility in different patients in a setting in which technical features were held constant.
In our study we targeted the reproducibility problem in the same way and successfully validated our biomarkers in a second sample set (consisting of pre-treatment sera from VL patients which were compared to the same controls as used for set 1) under the same settings. The intensities of the biomarkers found in our validation set were very similar to the ones in our original sample set. One biomarker with 9478 Da found in sample set 1 (fraction 5 bound on an IMAC30 array) was down-regulated in the positive samples. In the validation set, this protein was not detected as a biomarker at all (positive samples compared to controls). Therefore, its value as a marker can be questioned.

The small molecular weight differences are most likely caused by the fact that we processed the validation set with the sensitive SELDI technology on a different day using different arrays (even with the same settings on the same SELDI machine in the same laboratory using the same array type small molecular weight differences can be visible). This problem is discussed by Gretzer et al (2003). They mention that one limitation of SELDI is the lot-to-lot reproducibility of the chip surface chemistry. To maintain the reproducibility of mass-produced chips they ask for improved manufacturing methods.

In comparison to the results of our CL project more biomarkers were found in VL patients. This could be due to the fact that VL is a systemic disease and CL, which normally produces local skin lesions only, isn’t. Interestingly and to our surprise, no correlation between the clinical data and the peak intensities of any biomarker was detectable. Identification of markers that distinguish between VL and CL might help to unravel the pathophysiological differences between the 2 forms.

Agranoff et al (2005) report that the biological complexity of most disease states means that single biomarkers, with a few notable exceptions such as Troponin I or Troponin T as indicators of myocardial injury, have extremely limited diagnostic sensitivities and specificities. They mention that analyzing combinations of multiple biomarkers offers the possibility of greatly enhanced diagnostic accuracy. But in fact there are already existing
examples of highly sensitive and specific single protein ‘biomarkers’ (e.g.: hepatitis B surface antigen, circulating malaria antigens).

In this study single biomarkers which distinguished correctly between infected patients and controls were detected. Each candidate biomarker shown in fig.1 and fig.3 fulfills that task very well. It could be that because of the high sensitivity of SELDI-TOF-MS it is easier to find single biomarkers which distinguish correctly between positive and negative samples.

It was possible to successfully visualize 4 biomarkers on gel. Sequencing of the 51 kDa band detected in pre-treatment sera using LC-MS-MS followed by a Mascot search revealed a whole list of potential proteins. A literature investigation done on all proteins revealed links between the chosen proteins and the disease. Therefore it is possible that the 51 kDa band represents kininogen-1 precursor. High and low molecular weight kininogen are strong inhibitors of cysteine proteinases. Responsible for this activity are three related domains on its heavy chain (Higashiyama et al., 1986).

Since kininogen-1 precursor was up-regulated in pre-treatment samples and Leishmania parasites have cysteine proteases, a theoretical possibility could be that an infection with Leishmania up-regulates cysteine protease inhibitors in the host. Maybe kininogen is not in its active form because the parasite blocks its activation, which would be in favor of Leishmania.

High molecular weight kininogen plays a role in blood coagulation. It is also a precursor of bradykinin; this vasodilator substance is released through positive feedback by kallikrein.

It could be possible that kininogen is not in its active form because in that way blood doesn’t coagulate, which could be in favor of the sandfly trying to get a bloodmeal. In this way the chance of the parasite to get transmitted into other people is enhanced as well.

It is not surprising that so many proteins suggested in the Mascot database search were immunoglobulins (see table 4 and 5), because antileishmanial antibodies belonging primarily to the immunoglobulin G (IgG) class were already demonstrated in kala azar (Ghose et al., 1980) and post kala azar dermal leishmaniasis sera (Haldar et al., 1981). Ghosh et al (1995) showed in their study that antileishmanial reactivities in IgG ELISA followed the order IgG1 > IgG2 > IgG3, with very little IgG4. Interestingly, the 28 kDa protein (see table 5), which was down regulated in our tested patients, also included immunoglobulins as a result in the Mascot database search for possible identified proteins. Therefore, when compared to the literature, it is quite unlikely that the 28 kDa protein represents an immunoglobulin because hypergammaglobulinemia is present in the sera of most patients with visceral leishmaniasis (Carvalho et al., 1983; Pearson et al., 1983).

Anemia is almost always present in VL cases and may be severe. It appears to be due to a combination of factors, including hemolysis, marrow replacement with *Leishmania*-infected macrophages, hemorrhage, splenic sequestration of erythrocytes, hemodilution, and effects of cytokines such as TNF-α (Knight et al., 1967; Woodruff et al., 1972; Pippard et al., 1986). Therefore it could be possible that the 14.5 kDa biomarker, which was down-regulated in sera from patients, represents hemoglobin (see table 6).

An interesting link between VL and the up-regulated 12.4 kDa band was revealed in the literature. Other groups already reported cases of hyperlipidemia in VL cases (Erdeve et al., 2004; Elnour et al., 2001). Therefore it could be that this 12.4 kDa band represents
lipoprotein. This finding would be in agreement with the results from the previously mentioned groups.

As Agranoff et al (2005) mentioned, identification of the individual biomarkers will enable the design of immunologically based antigen-detection tests that could be implemented in dipstick or cassette formats. Furthermore, insights into disease pathophysiology afforded by knowledge of changes in circulating proteins could, ultimately, help to direct the search for novel drug targets. Because immunodetection of proteins is quantitatively more precise than MS, diagnostic performance of these downstream applications might surpass that of the original high-tech discovery-phase ‘test’.

This preliminary study suggests that SELDI-TOF-MS could be useful to detect new biomarkers in serum samples from VL patients. But before these biomarkers can be widely used, more tests have to be done, for instance additional patients have to be tested from other geographic regions, and people with other diseases have to be included to check for specificity. If more comprehensive studies can validate our results, it is quite likely that our findings will influence future VL diagnosis.
Literature Article 2


Leishmania braziliensis in blood 30 years after cure. Lancet, 341: 1341.


MS approach to proteomics: protein profiling and biomarker identification. Biochem 


proteinchips: potential use in diagnosis and prognosis. Proc Natl Acad Sci USA, 100: 
12343-8.

laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis, 21: 
1164-1177.

Veterinary Record, 144: 169-71.


Contributions of Authors

Christine Straccini assisted running the gels and cutting the bands. Bernard Gibbs sequenced the bands and did the Mascot search. Nirmal Baral provided all sera and information regarding the tested subjects. Greg Matlashewski and Brian J. Ward were the supervisors of Manfred Fussi, who did the practical and theoretical work described in this article during his M.Sc. studies. Momar Ndao supported Manfred Fussi with all aspects of the project.
CONCLUSION AND SUMMARY

Leishmaniasis is caused by the hemoflagellate *Leishmania* spp. and is one of the most important infectious diseases of the developing world, in terms of the number of people infected and the potential harm it can cause the patient.

The disease is prevalent in at least 88 countries. Depending on the causative species, it can manifest as cutaneous leishmaniasis (CL), mucocutaneous leishmaniasis, diffused CL, or visceral leishmaniasis (VL). Every year, an estimated 1.5-2 million children and adults develop symptomatic disease (cutaneous 1-1.5 million; visceral 0.5 million), and the incidence of infection is substantial when subclinical infections are included (Desjeux, 2004). The geographic hotspots (where ninety percent of the cases occur) for CL are Afghanistan, Algeria, Brazil, Iran, Peru, Saudi Arabia and Syria, while the main regions for VL are Bangladesh, India, Nepal, Sudan and Brazil (Desjeux, 2004). Leishmaniasis is of concern in several other countries too because of military activities (e.g. US missions in Iraq) and adventure tourism. Depending on the type of leishmaniasis, symptoms may occur for a few months to several years and range from skin lesions (as in the case of CL) to death (as in the case of VL if left untreated).

Currently, there are significant difficulties with the diagnosis of leishmaniasis which makes treatment difficult and extends patient suffering. The clinical signs and symptoms of VL and CL are not pathognomonic. VL may be confused with other similar conditions such as malaria, tropical splenomegaly, schistosomiasis or cirrhosis with portal hypertension, African trypanosomiasis, milliary tuberculosis, brucellosis, typhoid fever, bacterial endocarditis, histoplasmosis, malnutrition, lymphoma, and leukaemia (Singh & Sivakumar, 2003). However unlike malaria, there is no early dramatic fever to announce the arrival of VL; the presentation is insidious, with symptoms appearing over a period of weeks or even months. VL affects not only the weakest in the community, such as children and those weakened by other diseases such as HIV and tuberculosis, but also healthy adults and economically productive social groups (Hailu et al., 2005). Most VL infections occur in remote geographical areas where health facilities are not well
established and where the infection often co-exists with malaria and other debilitating parasitic infections. Under these circumstances, the disease usually presents a diagnostic dilemma. If VL is left undiagnosed and therefore untreated, it kills almost all patients (Guerin et al., 2003). If it is detected, the drugs used to treat it are so toxic that patients can die from the treatment. CL may be confused with tropical ulcers, impetigo, infected insect bites, leprosy, lupus vulgaris, tertiary syphilis, yaws, blastomycosis, skin cancer, etc. (Herwaldt, 1999; Lainson & Shaw, 1987; Ashford & Bates, 1998; Singh et al., 2005; Singh & Sivakumar, 2003). As no adequate screening tests for CL and VL exist on the market that can provide 100% sensitivity and specificity for all cases, in a suspected case of leishmaniasis only laboratory diagnosis can give a final answer (Singh, 2006).

Given the widespread nature of leishmaniasis and the serious implications that it can cause, particularly when it is not properly diagnosed and treated as is so often the case, it is clear that simple diagnostic tools, affordable and easy-to-administer drugs, and vaccines are necessary to win the fight against this disease.

The purpose of this research was to search for biomarkers in the serum of CL and VL patients using a highly sensitive mass spectrometric technology called Surface-Enhanced Laser Desorption/Ionization-Time of Flight- Mass Spectrometry (SELDI-TOF-MS) that detects proteins. If such biomarkers, or indicators of the disease, could be found they could be used to assist in the development of potential tests, drugs and vaccines.

In order to conduct this research, serum samples from subjects with VL were obtained from Nepal and samples from subjects with CL from countries in both the Old and New World. Samples were obtained in the pre-treatment (VL, CL) and treatment (CL) phase. They were compared with control samples from non-infected subjects from the same geographic regions. Serum and control samples were fractionated and each of the 6 fractions (pH 9, 7, 5, 4, 3 and organic phase) were bound on CM10 (weak cationic exchanger) and IMAC30 (immobilized metal affinity capture) ProteinChip Arrays. Proteins bound to chip surfaces were detected with the ProteinChip Reader using high
and low laser intensity. The results were analyzed using CiphergenExpress software and Biomarker Pattern Software.

In the following paragraphs we will present the key findings from our research, as elaborated on in chapter 1 in the case of CL and chapter 2 in the case of VL, before the impacts of these findings for the future are discussed.

**CL and VL Research Results**

Several differences in the protein spectra were detected by comparing pre-treatment samples from CL- and VL-patients with their controls. These comparisons led to the discovery of potential preliminary biomarkers (for VL cases with a molecular weight between 2.5 – 157 kDa, p-value < 0.001; and for CL from 2.6 – 167 kDa, p-value < 0.05). More biomarkers were found in VL samples compared to CL patients. To our surprise, the intensity of the individual biomarkers found in CL and VL cases did not correlate with the clinical data.

For the CL samples, we sought to identify all biomarkers detected in the 2\textsuperscript{nd} pass analysis (using CiphergenExpress) and BPS (9 biomarkers in pre-treatment samples, 3 biomarkers after 14 days of treatment and 2 biomarkers after 20 days of treatment). We were able to successfully detect 3 from 14 biomarkers on the gels. Sequencing of a 43 kDa band found in pre-treatment sera and a 13.9 kDa band from sera treated for 14 days using LC-MS-MS followed by a Mascot database and literature search (to find links between the proteins and the disease) revealed that there is a possible connection between ApoA-I, ceruloplasmin, complement C3 precursor, albumin, haptoglobin and CL. Therefore, our target proteins could represent one of these 5 proteins.

For the VL samples, it was possible to successfully detect 4 from 15 biomarkers on the gels (2 down-regulated proteins with 28.2 kDa and 14.5 kDa, and 2 up-regulated proteins with 51.1 kDa and 12.4 kDa). We sequenced (using LC-MS-MS) bands with 51 kDa, 28.2 kDa, 14.5 kDa and 12.4 kDa, all detected in pre-treatment sera followed by a Mascot
and literature search which revealed a possible link between kininogen, immunoglobulin, hemoglobin, lipoprotein and the disease. Therefore our sequenced bands could represent these proteins. The 15 preliminary biomarkers found in our sample set 1 were validated in a second sample set consisting of pre-treatment sera from VL patients. All biomarkers were detected with similar (Dalton shift +1%) or identical molecular weights except one down-regulated biomarker that was not detected in the 2nd sample set at all.

We can conclude that our study has shown the successful application of Ciphergen ProteinChip Array technology in the search for biomarkers in sera of CL and VL patients by identifying potential biomarkers.

**Future Direction**

Our research shows that SELDI-TOF-MS can be used to search for biomarkers in sera from human leishmaniasis patients. Although other groups have used SELDI-TOF-MS in the search for biomarkers in parasitic diseases like human African trypanosomiasis (Agranoff *et al.*, 2005), this research is first in applying SELDI-TOF-MS to search for biomarkers in CL and VL cases. Future studies can expand on this research, using greater sample sizes with a larger cohort of well-defined CL and VL patients to test the validity of the biomarkers detected or find other potential biomarkers for fighting CL and VL.

Future research samples should contain serum from patients treated with different medications as well, in order to see whether the use of medication has affected the CL results where we used both pre- and treatment phase samples. To test for the prognostic value of the biomarkers from our CL and VL study samples taken during and after treatment should be investigated too.

Using a greater sample size with a larger cohort of well-defined CL and VL patients would help to address two limitations of our study, namely the small sample size and the use of serum from patients during treatment where the same medication was used, and thereby better clarify the value of the biomarkers found. CL and VL serum samples from
different geographical regions have to be tested too in order to clarify whether our findings can be applied to a bigger geographical region, particularly in the case of VL. Furthermore, while the biomarkers for the visceral leishmaniasis study were validated by testing a second sample set, future research should do the same for CL.

Another means in which future research could validate the usefulness of these biomarkers in terms of their specificity would be to compare our results with the ones from human sera infected with other parasitic diseases (like malaria or trypanosomiasis) to see whether the same biomarkers will occur in these samples.

Our research also detected other biomarkers in the BPS analysis, which were not identified yet. Those marked as the most useful should be identified through future research to see whether they could also be useful in fighting leishmaniasis. In the case of CL, this should include the 8 remaining biomarkers in pretreatment sera, 2 in sera from patients treated for 14 days, and 2 in sera from patients treated for 20 days that we detected. In the case of VL, this should include the 11 biomarkers found in pre-treatment sera of human patients. The fact that we found more biomarkers in VL than in CL cases is also interesting and could be due to the fact that VL is a systemic disease.

Future research could determine whether a combination of biomarkers can enhance the sensitivity and specificity for diagnosis purposes. The underlying mechanism of diseases is quite complicated in that often multiple dysregulated proteins are involved. It is for this reason that recent hypotheses suggest that the detection of panels of biomarkers may provide higher sensitivities and specificities for disease diagnosis than is afforded with single markers (Xiao et al., 2005). Because we discovered more than one biomarker in each (CL and VL) study, future research could verify whether a combination of discovered biomarkers in each study would bring about a higher diagnostic value.

Also, as the biomarkers we identified in this research are of host origin, it would be valuable to investigate if any newly identified biomarkers consist of parasite proteins. Where this is the case, research should study the function and role that these proteins play
because it might reveal more about the parasite biology and the complicated host-parasite relations.

Once future research has confirmed whether the proteins identified are likely to be valuable in fighting leishmaniasis, they should be used to assist in the development of a cheap, easy to perform and sensitive test (e.g.: ELISA) which could be used in laboratories throughout the world. Such cheap tests would avoid difficulties associated with using SELDI-TOF-MS, in particular the significant costs, skills and specific conditions required to operate this technology successfully which make it unlikely to prove useful in most of the nations where leishmaniasis is currently the most prevalent. Another goal should be to produce substances which target the proteins contained in the most promising biomarkers. Such substances could be used to create effective drugs or an efficient vaccine to fight the disease.

Future studies may also use SELDI-TOF-MS not only to identify biomarkers from serum as we have done in our research, but also for other liquids that may also provide insights into fighting leishmaniasis. For example, the medium of a cell culture where macrophages are infected with *Leishmania* could be tested to give new insights into parasite biology. As host-parasite interactions are very complex and SELDI-TOF-MS is a very sensitive technology, it may provide new insights that have not yet been seen.

In summary, our research reported here is a first step towards verifying the suitability of SELDI-TOF-MS in the fight against leishmaniasis. It has contributed primarily to research by identifying six potential biomarkers, and detecting several other biomarkers, that could eventually prove useful in detecting, treating or immunizing against the disease. We certainly hope that our research will be built upon and eventually contribute towards assisting millions of people who suffer from this terrible disease around the world each day.
ACKNOWLEDGEMENTS

Many people supported me throughout my thesis. Those individuals to whom I am especially grateful are my supervisors Dr Brian Ward and Dr Greg Matlashewski for their guidance, help and assistance throughout the thesis. I am also indebted to Dr Momar Ndao for his good advice, time and humour whenever needed. Dr Naomi Aronson (Walter Reed Army Hospital, USA) and Dr Nirmal Baral (Koirala Institute, Nepal) were invaluable in providing the serum samples for this research. Christine Straccini’s positive attitude was always a great help, as were her technical skills with the SELDI and gels. Dr Gibbs was very helpful in sequencing the protein bands. The entire Ward lab (Carey, Claire, Marcia, Fabio, Milli, Angela, Dana, Katie, Sazini, Romina, John, Stephanie, Nathalie, Amanda, Adam, Cesar and Veerle) supported me in many ways through insightful discussions and a common appreciation for culinary delights. The Matlashewski lab was very helpful too, in particular Wen Wei Zhang, as well as the CIHR for their funding. My parents, who have provided outstanding support and love throughout my life, made it possible to come this far. Finally I would like to say Vielen Dank to my wonderful Madl Annelies for her great love and fantastic support.