Characterisation of the Egg Secretions of

Schistosoma mansoni

Peter Damian Ashton

Submitted for the Degree

of

Doctor of Philosophy

The University of York

Department of Biology

September 2001

Abstract

Schistosomiasis is a chronic debilitating disease caused by several species of parasitic worms of the genus *Schistosoma*. Pathology in the disease occurs mainly due to the presence of the eggs laid by the worm in the host bloodstream. The secretions of the egg are believed to be involved in two aspects of the disease. Firstly, it is assumed that they are essential in the process of egg escape from the host, which is clearly required for the life cycle of the parasite to continue, and has also been demonstrated to be dependant on the host's own immune response. Secondly, they are believed to elicit the formation of granulomas that occur around eggs that become trapped in the host's liver. It is the formation of these granulomas that leads to the major pathological consequences of infections. Although much work has been done on the proteins of the egg, its secretions have never been systematically studied and characterised and little attention has been paid to its basic physiology.

An ultrastructural examination of the egg identified that the syncytial envelope which surrounds the developing miracidium in many species of trematodes is a biosynthetically active tissue and it therefore capable of synthesising the secretions itself, without the involvement of the miracidium. Culturing eggs in serum free medium has enabled the secretions to be collected directly and systematically characterised. Zymography has shown that two of the proteins are proteases of different specificities. Proteomics techniques, including separation by two-dimensional electrophoresis, peptide mass fingerprinting and peptide sequencing using tandem mass spectrometry have enabled genes encoding several of the secreted proteins to be identified.

Table of Contents

Acknowledgments

I would like to extend my thanks to the following people:

Prof. A. Fitter in whose department this work was carried out.

The UNDP/World Bank/World Health Organisation Special Program for Research and Training in Tropical Diseases and the European Commission's BIOTEC program for providing the funding for the project.

Mr. Srdjan Ljubojevic for technical assistance during the early part of the project.

Dr. Sandie Baldauf, for explaining phylogenetic analysis.

Staff at both PE Biosystems and Micromass, especially Aaron Hudson(Micromass) and Amelia Jackson and Martin Hornshaw (PE), who have helped out with assistance and the use of equipment at times of stress.

Friends and colleagues in the Biology Department at the University of York.

I would like to extend special thanks to the following people:

Alan for his guidance, tolerance and most of all his endless enthusiasm.

Bindiya for all her help with the dark art called "molecular biology" and Rachel for having to put up with me more than most.

All the other members of the Schistosome Research Group, past and present.

Finally, Jan for coping with me for so long and Zoë, for brightening my life more than I could ever have imagined. It is to them that this thesis is dedicated.

Declaration

All of the work contained in this thesis is my own with the following exceptions:

Addendum to Chapter 2: The mouse surgery and Sepharose bead injections were performed by Prof. R.A. Wilson and Dr. P.S. Coulson.

The histological preparations were performed by Deborah Saxton.

The mouse infection was done by Clare Sadler.

Chapter 3: The 2-D gel of cercarial antigens, tryptic digests and MALDI-ToF identifications were performed by Rachel Curwen.

Chapter 1: Introduction

Background

Schistosomiasis is a disease caused by several species of parasitic trematode worms of the genus *Schistosoma*. According to estimates by the World Health Organisation, over 200 million people are infected worldwide, with 120 million suffering some symptoms, and 20 million suffering serious consequences of infection. In terms of economic and social impact, this puts schistosomiasis as the second most important parasitic disease after malaria. *Schistosoma mansoni* alone is prevalent in 54 countries in Sub-Saharan Africa, South America and the Caribbean [WHO, 1998].

Life Cycle

Schistosoma species are digenetic, and therefore require two hosts for completion of the life-cycle, a mammalian definitive host and an intermediate molluscan host, usually an aquatic snail. The intermediate host is infected by a free-swimming miracidium that emerges from eggs excreted in the faeces (for *S. mansoni* and *S. japonicum*) or urine (for *S. haematobium*) of the definitive host. After penetration, the miracidium transforms into a mother sporocyst and produces several daughter sporocysts. Each of these contains many germ balls, which develop into cercariae, resulting in the production of many thousand larvae for each infecting miracidium. These free-living cercariae emerge from the snail on stimulation by light and can survive in water for 16-30 hours [Wilson, 1988]. However, the cercariae do not feed and rely on a fixed supply of glycogen, and so they must locate a mammalian host to infect within this time. Infection is achieved by penetration of the host skin, which is digested by a number of enzymes released from two sets of glands whose contents are synthesised as the cercariae develop within the snail [Dorsey & Stirewalt, 1971]. After penetration, the cercariae transform into schistosomula, which reside in the skin for approximately four days and then enter the vasculature, possibly with

the use of further digestive enzymes, or simply by the mechanical disruption of the blood vessels [Crabtree & Wilson, 1985].

The schistosomulum migrates through the host bloodstream until it reaches the lungs, where it must elongate in order to proceed through the capillary beds. Eventually, possibly after several complete cycles through the host circulation [Wilson & Coulson, 1986], the worms reach the hepatic portal vasculature where they differentiate into a slender blood-feeding form. They then mature, pair up, migrate to the blood vessels of the gut wall and, approximately 5-6 weeks after initial infection, commence egg-laying. The deposited eggs must lodge against the endothelium of the blood vessels in which the adults reside, and then penetrate through the wall of the gut into the lumen, where they are voided with the faeces and hence continue the life cycle.

The Disease

Few or no pathological consequences arise from the presence of adult *S. mansoni* worms in the mesenteric veins of the hepatic portal system, except for an acute reaction, termed *Katayama fever*, which may occur around the onset of egg-laying, particularly in individuals who have had no previous exposure to schistosomiasis. Instead, most of the chronic pathology occurs due to the eggs, which are laid continuously throughout the infection, with each worm pair producing up to 300 eggs per day. Although a proportion of these migrate successfully through the gut wall and escape from the host, some are instead swept away by the portal circulation and lodge in the pre-sinusoidal vessels of the liver. For as long as the eggs survive, they continue to secrete the factors that are required for their escape, but in this location they provoke the infiltration of large numbers of macrophages, T-cells, granulocytes (mainly eosinophils) and B-cells, resulting in the formation of large granulomas around each of the eggs [von Lichtenberg, 1988]. These lesions are similar to the reactions observed in delayed-type (type IV) hypersensitivity responses [Warren, Domingo & Cowan, 1967]. During the initial stages of infection the developing worms stimulate mainly T-

helper 1 type responses, however, at the commencement of egg-laying, the egg antigens cause a switch to a T-helper 2 response. As the disease progresses and the egg load increases, in experimental infections, this Th2-response begins to dominate, leading to the modulation of the granulomatous response and, 8-20 weeks after initial infection, a decrease in the size of the egg granulomas [Stadecker, 1994; Chensue & Boros, 1979]. The granulomas become fibrotic as fibroblasts are recruited and lay down collagen, and in a small number of cases, this can lead to the formation of more generalised pre-sinusoidal inflammation and fibrosis, termed "pipe-stem fibrosis".

These immune reactions serve to encapsulate the entrapped eggs, and as a consequence block the blood vessels. As the number of eggs, and the proportion of the blocked vessels increases, the pressure in the hepatic portal vein rises [McHugh, Coulson & Wilson, 1987]. Eventually, this portal hypertension causes the formation of porta-caval anastamoses, to relieve the pressure. The leakage of oesophageal varices formed in this way can lead to haematemesis (in human patients, often the first sign of infection) and their rupture can be potentially fatal. The damage that occurs to the liver does not generally affect its function, but the attendant increase in arterial flow may result in the gross enlargement of the liver (hepatomegaly). The presence of the infection, with the active immune response, also results in the enlargement of the spleen.

It is clear from this account that the egg is central to the pathological processes of the disease, and much effort has been directed at understanding its role, and that of the host's immune response, in schistosomiasis.

The Role of the Host Immune Response in Pathology

Since the immune response, rather that the presence of the egg itself, was seen to be the cause of much of the pathology, the question of whether removing the immune response would also remove the

pathology was raised [Doenhoff *et al.*, 1981]. In experiments performed on mice that had been neonatally thymectomised and treated with rabbit anti-mouse thymocyte serum to eliminate any T-cell dependent immune responses, it was observed that although the granulomatous response was abrogated, there was instead evidence of damage to the hepatic tissue immediately surrounding the entrapped eggs. This led to the hypothesis that the role of the granuloma was host protective, effectively encapsulating the egg and preventing its secretions from destroying the host tissue. In addition to deficiencies in granuloma formation, these mice also failed to excrete as many eggs as normal mice with a similar worm burden, suggesting that the process of egg escape is dependant on the host's immune response. This raises the possibility that the reaction is deliberately provoked by the eggs in order to facilitate their egress [Doenhoff *et al.*, 1986]. Confirmation of this hypothesis was sought by studying the egg excretion rates in schistosomiasis patients who were co-infected with HIV [Karanja *et al.*, 1997]. Again, the severe reduction of an effective $CD4^+$ T-cell response caused by the HIV led to a corresponding reduction in the rates of egg excretion, even when corrected for the worm burden (as assessed by measuring circulating schistosome antigen levels in serum).

The Evidence for Egg Secretions

Although it has generally been assumed that the eggs actively secreted "factors" that would be important in both pathogenesis and translocation across the gut wall, they have never previously been directly collected and extensively characterised. However, there are several lines of evidence that support their existence. Firstly, and most obviously, liver granulomas (as well as experimentally induced lung granulomas [Peterson & von Lichtenberg, 1965]) form around live eggs. It is clear from the microscopic examination of liver sections containing eggs that the miracidium is undamaged and viable in most granulomas, although clearly a proportion of the granulomas will be around eggs that have died and degenerated subsequent to the formation of the

granuloma. In contrast, granulomas do not form as quickly, if at all, around heat-killed eggs, isolated shells or miracidia alone [von Lichtenberg & Raslavicius, 1967].

Secondly, a technique was developed for serodiagnosis, which relies on the recognition of actively secreted antigens from live eggs. In this "circumoval precipitation test" (COPT), live eggs are incubated for 24 hours in the serum of a patient with suspected schistosomiasis. When the eggs are subsequently examined, the presence of precipitated immune complexes around the egg is indicative of the previous exposure of the patient to egg antigens and therefore indicates that the individual must have been infected. For the precipitation to form around live eggs, the antigens must clearly be actively secreted from the eggs in order to react with the antibodies in the patient's serum.

Finally, experiments have been performed in which mice were injected with culture medium that had been used to maintain *S. mansoni* eggs *in vitro* [Hang, Warren and Boros, 1974]. These mice were challenged by the injection of viable eggs into the tail vein, to lodge, not in the liver, but in the lungs, where a similar granulomatous response would occur. It was shown that the intraperitoneal injection of the culture medium sensitised mice for pulmonary granuloma formation, with larger lesions forming more rapidly than in naïve mice. This clearly demonstrated that the eggs were producing material that was recognised during the process of granuloma formation and that this material could be collected from culture medium. Perhaps surprisingly, the authors of this study did not attempt to characterise these secretions further and, in subsequent work, concentrated instead on the proteins contained in a homogenate of whole eggs, known as "soluble egg antigens" (SEA).

Previous Work on the Schistosome Egg

Although the egg has always been regarded as central to pathology, very little attention has been paid to its basic biology. Whilst

the ultrastructure of the egg has been examined, most of the work has involved attempting to identify individual antigenic components of the "soluble egg antigen" preparation.

Egg Ultrastructure

The single major ultrastructural study of the egg [Neill *et al.*, 1988] reveals a number of features of the mature egg. Of particular relevance to this project is the structure Neill *et al.* referred to as "von Lichtenberg's envelope", which corresponds exactly with the so-called vitelline membrane present in other trematode eggs. This is a syncytial envelope completely surrounding the miracidium and would therefore present a barrier to simple passive diffusion of nutrients into the egg and of secretions from the miracidium out. Since the development of the eggs of other trematode species has been studied in great detail (e.g., *Fasciola hepatica* [Schubmann, 1905; Ortmann, 1908], *Fasciolopsis buski* [Ishii, 1934], *Paragonimus kellicoti* [Chen, 1937], *Parorchis acanthus* [Rees, 1940]), it is probably safe to assume that the envelope of *S. mansoni* arises in the same manner. In these species, several somatic cells from the embryonic miracidium separate from the main cell mass and move outwards. On reaching the shell, they flatten against the inner surface and spread out, eventually meeting and fusing to form the enveloping syncytium. Although the presence of the envelope could be discerned, the quality of the preservation of the sections the Neill study was not ideal, and the level of detail that could be distinguished was relatively poor. This was probably due to the slam freezing and freeze substitution which were used to circumvent problems with the infiltration of embedding resin through the eggshell. As a result, the fine structure of the envelope was not visualised clearly in any of the electron micrographs presented.

"Soluble Egg Antigens"

Several attempts have been made to identify components of the egg that are responsible for the pathology observed in schistosomiasis. To date, all of these have concentrated on the so-called "soluble egg

antigens" or SEA. This was first defined by Boros and Warren [1970] as that fraction of the total egg contents which was soluble in PBS after sonication. Since it was originally defined, several groups have identified proteins from this complex mixture using a variety of techniques. Most groups have attempted to identify antigenic proteins either by screening of cDNA expression libraries using sera from infected animals or by using such sera to purify the antigenic components of SEA. The proteins that have been cloned and sequenced include, most importantly, SmP40 [Nene *et al.*, 1986], a 40kDa protein, which is the most abundant component of SEA, comprising ~10% of the total protein. This protein appears to be a heat-shock protein, containing two HSP20-like domains. Other antigens that have been identified by library screening include SmE16 (a 16kDa calcium binding protein) and adenylate kinase. A number of other proteins have been characterised by different chromatographic or electrophoretic methods, most notably α_1 and ω_1 [Dunne *et al.*, 1991]; however, the sequences of these proteins have not been determined. None of the above proteins have been shown to be secreted from the egg.

Screening with antibodies has enabled a number of proteins to be characterised but, as mentioned above, the granuloma is a T-cell mediated DTH response [Boros & Warren, 1970] and, it is therefore more likely that screening for T-cell antigens would be a better approach. By separating the crude SEA mixture e.g., by SDS-PAGE, and culturing T-cells with individual fractions it is possible to identify components of the mixture which might be involved in granuloma formation. Two egg proteins that have recently been identified by this method are phosphoenolpyruvate carboxykinase [Asahi *et al.*, 2000] and thioredoxin peroxidase (TPx) [Williams *et al.*, 2001]. Antibodies raised against recombinant TPx were used to demonstrate secretion, by probing western blots of egg-conditioned culture medium. Additionally, immunolocalisation of the protein in the egg showed that it was present

predominantly in the envelope, rather than originating in the miracidium.

This strategy of looking for immunogenic components, in order to locate the egg secretions seems logical at first sight, since granuloma formation is clearly an immunological phenomenon. However, the eggs encapsulated by the granuloma will eventually die, and their contents will be released and become accessible to the immune system. The reaction that is then mounted against them will have little or nothing to do with the initiation of granuloma formation, or egg escape. It is therefore possible that the proteins identified by immunoscreening will not be directly relevant to either of these processes. To study the secreted proteins systematically, it is necessary to collect them directly and characterise those components of the culture medium that are responsible for the sensitisation of mice observed by Hang, Warren and Boros [1974]. In addition, studying the basic biology of the egg, in terms of both its ultrastructure and development, should provide clarification of the processes of escape from the host and hepatic granuloma formation.

Proteomics

In recent years, the field of "proteomics", defined as the study of the total protein complement expressed by a genome [Wasinger *et al.*, 1995], has emerged, largely due to the application of two-dimensional electrophoresis and mass spectrometry. These two techniques provide methods of profiling the expression of proteins directly and identifying the proteins present in individual gel spots. Although these techniques are well established, recent developments in both have enabled their reliable and reproducible application to complex biological samples.

Two-Dimensional Electrophoresis (2-DE)

Two-dimensional electrophoresis was first described some 25 years ago [O-Farrell, 1975; Klose, 1975]. As initially described, the technique consists of two main steps. Firstly, proteins are separated

according to their isoelectric point (p*I*), which is defined to be the pH at which the net charge present on the protein is zero. By placing the proteins on a pH gradient (created using a large number of small buffering molecules, termed carrier ampholytes) in a polyacrylamide tube gel and applying a large voltage across the gradient, the proteins will migrate towards either the positive or negative electrode. As the proteins move through the pH gradient, their overall charge will change, and once an individual molecule reaches the pH corresponding to its p*I*, its charge will become zero and it will not migrate further. In theory, if the voltage is maintained for long enough, all of the proteins will become confined to small zones in the pH gradient, corresponding to their p*I*. Secondly, the tube gel containing the focussed proteins is placed on top of a normal SDS-PAGE gel and run so that the proteins are separated orthogonally by size. After visualisation of the proteins by a suitable staining technique, a two dimensional map of the sample is produced.

This original technique suffers from two main limitations relating to the first dimension separation. Firstly, there are two competing constraints on the state of the proteins during the focusing: the proteins must remain soluble, preferably in a denatured state, and the native charge of the proteins must be maintained. The former requires that strong denaturing conditions must be used, but the latter precludes the use of any ionic reagents such as SDS. This limits the usefulness of the method for certain classes of proteins, such as membrane proteins, which contain large hydrophobic regions and are therefore not readily solubilised [Santoni, Molloy & Rabilloud, 2000]. Secondly, the use of small, soluble carrier ampholytes leads to the problems of cathodic drift, where the high pH end of the gradient progressively decays as the focussing proceeds, leading to the loss of the basic proteins from the end of the tube gel during extended separations. This also leads to a lack of reproducibility of the resulting gels.

The problems with cathodic drift have now effectively been eliminated by the introduction of immobilised pH gradient (IPG) strips [Bjellqvist *et al.*, 1982]. By using acrylamido derivatives of the carrier ampholytes, the buffering molecules can be covalently bonded into the gel matrix, so that it is impossible to for them to drift during the focussing run. This, combined with the mass production of such strips (with suitable quality control measures) has somewhat simplified 2-DE and led to a high degree of reproducibility between laboratories.

Although problems remain (e.g., with the separation of membrane proteins, or the under-representation of large proteins in gels), modern 2-DE allows the proteins present in a complex sample to be visualised and difference in protein expression between samples to be measured reliably. Whilst new techniques are emerging, such as multi-dimensional liquid chromatography or isotope-coded affinity tags, that may replace 2-DE in the long run [Haynes and Yates, 2000], none yet offers an equally rapid and direct means of observing the differences in expression between so many proteins at once.

Mass Spectrometry

Mass spectrometry (MS) has been used for decades to accurately measure the mass of molecules, but has only recently been widely applied to biological macromolecules. In any mass spectrometer, there are two steps to measuring the mass of a molecule: the molecule must be ionised in the gas phase (e.g., by chemical ionisation and vaporisation), accelerated in an electric field and then the mass to charge ratio determined in a suitable mass analyser. The original mass analysers used were magnetic sector instruments where the deflection of a beam of ions in a magnetic field was measured, the amount of deflection being directly proportional to the charge on the ion and inversely proportional to its mass. The problem with the MS analysis of large molecules has been in the use of a suitable ionisation technique that is capable of producing ions in the gas phase without disrupting the molecule at the same time. In recent years two suitable and complementary "soft"

ionisation techniques have been developed: matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI).

Both of these techniques can be used with proteins and peptides, however, measuring the mass of a protein will not, of itself, yield an identification. Instead, the techniques are normally applied to mixtures of peptides derived from a protein e.g., by tryptic digestion of a particular gel spot. MALDI is used with a time-of-flight (ToF) mass analyser to generate a "peptide mass fingerprint", which is simply a list of the masses of the peptides from a single protein in a gel spot. ESI is used with a hybrid quadrupole-ToF analyser to produce tandem mass spectra of individual peptides to generate peptide sequence tags.

MALDI-ToF/MS

In MALDI-ToF/MS [Karas & Hillenkamp, 1988], peptides (e.g., resulting from the tryptic digestion of a purified protein) are mixed with an acidified saturated solution of a small UV absorbing organic acid (e.g., α-cyano-4-hydroxycinammic acid) and allowed to crystallise on a metal target. When dry, the target is introduced into the vacuum of the mass spectrometer and is irradiated with short bursts of light from a UV laser. This results in desorption and ionisation of the peptides into the gas phase, predominantly as singly charged ions. These ions are then accelerated in an electric field, so that all of the ions have the same kinetic energy ($\frac{1}{2}mv^2$). The speed of a particular ion will therefore be dependent on its mass, with lighter peptides travelling faster down the flight tube. By measuring the time between the application of the electric field and the arrival of the molecules at a detector at the end of the tube, the mass of the peptides can be determined.

Since the method was introduced there have been two main refinements that have improved the accuracy of the mass measurements. Firstly, a short delay (of the order of 100ns) is employed between the firing of the laser and the application of the accelerating electric field allowing the ions to "cool" briefly, thereby reducing the spread of

velocities of a particular species and resulting in sharper peaks for each peptide [Vestal, Juhasz & Martin, 1995]. Secondly, it is possible to place an ion mirror (or "reflectron") at the end of the tube, which reverses the direction of the ions and directs them to a second detector. This has the effect of increasing the effective length of the flight tube (enabling molecules with similar masses to be resolved) and again of focussing the ions to produce sharper peaks [Chien, Michael & Lubman, 1993]. The combination of these refinements, with appropriate calibration, has made it possible to reliably measure the masses of peptides to an accuracy of better than 0.1Da.

The list of masses of the peptides from a particular protein (e.g., produced by the tryptic digestion of a gel spot) is usually referred to as a peptide mass fingerprint (PMF) [Cottrell, 1994]. As this name implies, the combination of these masses will usually be diagnostic of a particular protein, and it is therefore possible to identify proteins based on its PMF alone. Given a suitable database, it is relatively simple to compare the observed masses with those that would be produced by the theoretical digestion of each of the protein sequences in the database, and there are now many web sites available that allow such searches to be performed.

Electrospray Ionisation Tandem MS

Peptide mass fingerprinting allows a large number of proteins to be rapidly screened and, providing that the sequences of the proteins are present in the database, identified. However, if identification is not achieved by this method, or the identification is ambiguous, it is necessary to obtain more information about the protein. Fortunately, it is possible to use the same mixture of peptides to obtain peptide sequence tags [Mann & Wilm, 1994], or even the sequence of complete peptides, using tandem mass spectrometry, where two mass analysers are combined. Typically, these will be a "triple quadrupole" machine, or a quadrupole and a time-of-flight mass analyser [Morris *et al.* 1996]. In this case, the peptides are introduced into the machine using

electrospray ionisation, producing a continuous beam of ions which pass through the quadrupole. This is an arrangement of four parallel metal rods to which oscillating electric fields are applied. By controlling the frequencies of the fields, the quadrupole can be tuned to allow the passage of ions with a given charge to mass ratio (m/z), effectively selecting a single peptide from the mixture. The peptides then enter a collision cell which contains a small amount of an inert gas (e.g., argon) and undergo "collision-induced dissociation". The masses of the resulting fragments are then measured by the second mass analyser. Since this is usually a ToF analyser, the continuous beam of ions must be converted to distinct pulses, so that the time of flight can be measured. The flight tube is therefore placed orthogonally to the quadrupole, and a "pusher" is used to periodically accelerate the fragments along the flight tube. The fragmentation of the peptide that takes place is not random and occurs mainly along the peptide backbone, and each of the fragments will therefore contain a different number of amino acids from either the carboxy- or amino-terminus of the peptide. By calculating the mass differences between peaks, the sequence of the peptide can be easily determined. The peptide sequence can then be used to search online sequence databases, e.g., by using the NCBI BLAST program (http://www.ncbi.nlm.nih.gov/BLAST/). Whilst matching a single peptide with a particular sequence might well be sufficient to suggest an identification, matching several peptide sequences obtained from the same protein will give a very high level of confidence.

The available range of proteomics tools is now allowing biologists to identify proteins of interest directly without the time-consuming methods of Edman sequencing or antibody production and expression library screening.

Aims of the Project

The intention of this project is to obtain as much information as possible on the secretions of the *S. mansoni* egg which are involved in

the processes of egg escape and granuloma formation. The role of the secretions is assumed to be two-fold. Firstly, the proteins must contain factors which are capable of disrupting host tissue, in a similar manner to the proteases that are known to be secreted by the cercarial stage to digest the skin, and these factors presumably account for the hepatic damage observed in immunocompromised mice. Secondly, they must provoke an immune reaction by the host, which has been shown to be necessary for the successful escape of the eggs to continue the life cycle.

Rather than relying on indirect methods of identification, such as library screening, a range of approaches is taken to achieve the following aims:

- To examine of the ultrastructure of the egg by electron microscopy and localise of the site of production of the secretions by immunocytochemistry.

- To collect the secretions during *in vitro* culture, and characterise them by a range of electrophoretic techniques, including SDS-PAGE, 2-dimensional electrophoresis and 1-D and 2-D zymography.

- To measure the immune responses elicited by the egg secreted proteins both during the course of infection using ELISA and in an *in vivo* model of granuloma formation using antigen-coupled Sepharose beads.

- To establish the techniques of proteomics, including 2-D electrophoresis, MALDI-ToF/MS and tandem MS in the laboratory.

- To separate the egg secretions by 2-D electrophoresis and subject them to tryptic digestion and MALDI-ToF/MS and to use the resulting peptide mass fingerprints to search on-line sequence databases.

- To generate peptide sequence data from the secretions, and use them to search the dbEST division of the GenBank database in order to identify the genes that encode the ESP proteins.

Chapter 2: The Schistosome Egg: development and secretions

P.D. Ashton, R. Harrop, B. Shah and R.A. Wilson (2001). The Schistosome Egg: Development and Secretions. *Parasitology*. **122** (3): 329-338.

Introduction

This chapter is composed of a paper, as published in the journal *Parasitology* and an accompanying addendum, giving some further details of the characterisation of the immune responses to the egg secretions. The paper deals with the initial characterisation of the proteins that can be collected from the culture medium used to maintain *Schistosoma mansoni* eggs *in vitro* and demonstrates that a simple mixture of proteins are produced. Ultrastructural observations suggest that the envelope, which surrounds the miracidium in the mature egg, would be impermeable to the secretions of the larva itself, and immunocytochemical localisation confirms that is it this envelope, rather than the miracidium, which is their source. Use of the circumoval precipitation test (COPT) demonstrates that the collected proteins are actively secreted from the egg.

In addition to confirming that these proteins are truly secreted from the egg, if they are important in the process of granuloma formation, it must be shown that they are recognised by the immune system and that they are capable of inducing granulomas. These two points are demonstrated in the addendum to this chapter, in which the antibody responses to the secretions during infection and after intraperitoneal injection of eggs are measured by ELISA and their granuloma-provoking potential is demonstrated by the injection of ESP and SEA coupled beads into the superior mesenteric vein of primed mice.

The schistosome egg: development and secretions

P. D. ASHTON[1]*, R. HARROP[1]†, B. SHAH[1] *and* R. A. WILSON[1]

[1] *Department of Biology, University of York, PO Box 373, York YO10 5YW*

(*Received 7 May 2000; revised 7 September 2000; accepted 7 September 2000*)

SUMMARY

We have investigated the development of the schistosome egg and its secretions in order to understand how it migrates through gut tissues and also initiates pathology in the liver. We show by electron microscopy that the subshell envelope is absent in the newly deposited egg, but appears very early and differentiates as development progresses. In the mature egg, this nucleated envelope contains extensive endoplasmic reticulum, suggestive of a protein synthetic capacity. Furthermore, Reynolds' layer only appears between the envelope and the egg-shell in the mature egg and may represent its accumulated secretions. We have biosynthetically labelled and collected the secretions (ESP) released by mature but not immature eggs during culture. Their fractionation by SDS–PAGE reveals a simple pattern of 6 bands, differing markedly in composition from soluble egg antigen preparations. Electrophoresis in casein substrate gels demonstrates the presence of 2 distinct proteases in the egg secretions. By immunocytochemistry, ESP localized predominantly to the envelope of the mature egg, suggesting that this layer rather than the miracidium is the source of egg secretions.

Key words: schistosome egg, ultrastructure, development, secretion, protease.

INTRODUCTION

Adult worm pairs of the blood fluke *Schistosoma mansoni* reside in the hepatic portal system where the females deposit undeveloped eggs in the vessels of the gut wall. These eggs must undergo development, traverse the gut tissue to reach the lumen and be voided with the faeces to continue the life-cycle. Little is known about the secretions from eggs which facilitate their passage through the tissues but the process is known to be dependent on the immune response, in both rodents (Doenhoff *et al.* 1986) and humans (Karanja *et al.* 1997). From these observations it can be inferred that eggs release immunogens to mediate the inflammatory response which assists their tissue migration. Furthermore, it is plausible that such secretions from viable eggs embolized in the liver might initiate granuloma formation (Hang, Warren & Boros, 1974) which drives the pathogenic processes of schistosomiasis.

To date, there has been only 1 major ultrastructural study of the *S. mansoni* egg (Neill *et al.* 1988). The relative impermeability of the egg-shell is an obstacle to investigations, preventing the infiltration of embedding resins, and was circumvented by slam freezing and freeze substitution. This study revealed in the mature egg a subshell layer (Reynolds' layer) comprising microfibrils in a granular matrix. Beneath this and completely surrounding the miracidium and other structures was a thin envelope (von Lichtenberg's envelope) of squamous cellular origin. This is equivalent to the 'vitelline membrane' described in other trematode eggs (Wilson, 1967), and was believed by Neill *et al.* (1988) to represent a barrier to simple passive diffusion between the host extracellular fluids and the developing miracidium. They concluded that the envelope might regulate the transport of egg antigens and other embryonic products out of, and of nutrients into, the milieu surrounding the miracidium.

As part of a wider investigation of schistosome egg biology, we have examined the structures of both immature and mature schistosome eggs after conventional fixation and embedding procedures. We describe the development and differentiation of the subshell envelope and associated Reynold's layer. We have collected and characterized the egg secretions and shown that they are different in composition to the standard soluble egg antigen (SEA; Boros & Warren, 1970) preparation used by schistosome researchers and contain at least 2 proteases. Furthermore, the envelope possesses protein synthetic machinery and we demonstrate that egg secretions localise abundantly to this layer.

MATERIALS AND METHODS

Generation of parasite material

Groups of MF1 or CBA × C57Bl/6 (F1 cross) mice each were infected with 200 cercariae of *S. mansoni*. Seven weeks later they were sacrificed, their livers were removed, homogenized and digested with

* Corresponding author. Tel: +01904 432871. Fax: +01904 432884. E-mail: pda2@york.ac.uk
† Current address: Oxford BioMedica (UK) Ltd, Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA.

Parasitology (2001), **122**, 329–338. Printed in the United Kingdom

trypsin for 3 h at 37 °C. After incubation, the livers were sieved, the eggs collected by sedimentation and cleaned by washing 6 times in RPMI-1640 (Gibco). Mature and immature cleaned eggs were separated by layering on top of a mixture of 6 ml of Percoll (Pharmacia), 0·6 ml of 9% saline and 3·4 ml of RPMI-1640 and centrifuging for 15 min at 250 *g*. The mature eggs collected at the bottom of the tube with immature eggs and some tissue debris remaining at the interface. The separate egg fractions were then washed a further 6 times in RPMI-1640 to remove the Percoll.

Preparation of soluble egg antigens (SEA)

Cleaned eggs were resuspended in 1 ml of PBS and sonicated for 4 × 1 min on full power with 1 min on ice between each sonication, centrifuged at 100 000 *g* for 60 min and the supernatant taken as SEA. The protein concentration was measured using a BioRad protein assay kit based on the technique of Bradford (1974).

Preparation for electron microscopy

Mature and immature eggs were prepared separately. Primary fixation was performed in 2·5% glutaraldehyde/4% paraformaldehyde in 100 mM phosphate buffer, pH 7·2, at 4 °C overnight. Groups of eggs were washed thoroughly in PBS and placed into a Petri dish on a dissecting microscope. Individual eggs were hand-pricked with a fine entomology pin, to allow the penetration of resin into the shell, and collected into an Eppendorf tube. Secondary fixation was by incubation in a 1% solution of osmium tetroxide in 100 mM phosphate buffer, pH 7·2, for 1 h on ice. The eggs were then embedded in Spurr's resin, 0·075 μm sections were cut, stained with lead citrate and uranyl acetate and examined on a Jeol JEM-1200EX transmission electron microscope.

Culture of mature eggs

After separation and cleaning, approximately 500 000 mature eggs were placed into a tissue culture flask with 30 ml of RPMI-1640 culture medium containing penicillin (300 U/ml), streptomycin (300 μg/ml) and gentamycin (500 μg/ml), but in the absence of serum supplements. Eggs were maintained for 72 h at 37 °C in 5% CO_2 and the culture medium was collected and concentrated in an Amicon Ultrafree-15 centrifugal ultrafiltration cell (5 kDa molecular weight cut-off) to a final volume of approximately 1 ml to give the preparation referred to as egg secreted protein (ESP). Viability of the eggs after the culture period was better than 85% as assessed by flame cell activity or muscular contraction of the miracidia. The protein content of the ESP was measured as before. Some aliquots of eggs were biosynthetically labelled by adding 0·3 μl/ml (0·158 MBq/ml) of ^{35}S-labelled methionine and cysteine (ProMix, Amersham) to the medium during the culture period. The counts incorporated into egg proteins were determined by sonicating known numbers of eggs for 1 min in the presence of 100 μl of 1% bovine serum albumin as a carrier protein and 200 μl of 10% trichloroacetic acid (TCA). The amounts of labelled protein secreted by immature and mature eggs during the culture period were determined on TCA-precipitated aliquots of culture medium, and related to the number of eggs present.

Gel electrophoresis

ESP proteins were separated on a 4–12% Bis–Tris SDS–PAGE gel (Novex) according to the manufacturer's instructions, with and without the presence of reducing agents in the sample buffer and the resulting gel stained with Coomassie Blue. Molecular weights were calculated on a GelDoc system (FlowGen), using Precision Protein Standards (BioRad) as calibrants. Dried gels of radio-isotope labelled protein preparations were exposed to autoradiography film for 16 h at −80 °C.

Protease substrate gels

ESP proteins were separated on a 15% SDS-PAGE gel using a BioRad Mini-Protean II system, without reducing agent. The gel included 0·1% casein to act as a protease substrate (Lockwood *et al.* 1987). After electrophoresis, the gels were washed twice for 30 min in 2·5% Triton X-100 (Sigma), to remove the SDS and renature the proteins. The gel was then incubated in a 50 mM Tris, pH 7·5 buffer, containing 200 mM NaCl, 5 mM $CaCl_2$ and 5 μM $ZnSO_4$, at 37 °C overnight and stained with Coomassie Blue as normal.

Preparation of antisera

One rabbit was immunized with 200 μg ESP in TiterMax Gold adjuvant (Sigma), bled after 4 weeks and serum prepared by centrifugation after clotting. The titre (50% maximum OD) of the antiserum was 1 in 3000 as assessed by ELISA.

Immunocytochemistry

Pieces of mouse liver were removed 7 weeks post-infection and frozen in OCT compound (Histological Equipment Ltd, Nottingham) by immersion in thawing isopentane. Sections (10 μm thick) were cut on a cryostat, air dried and fixed in acetone for 30 min at −20 °C. After drying, the sections were blocked by incubation in 10% normal goat serum in

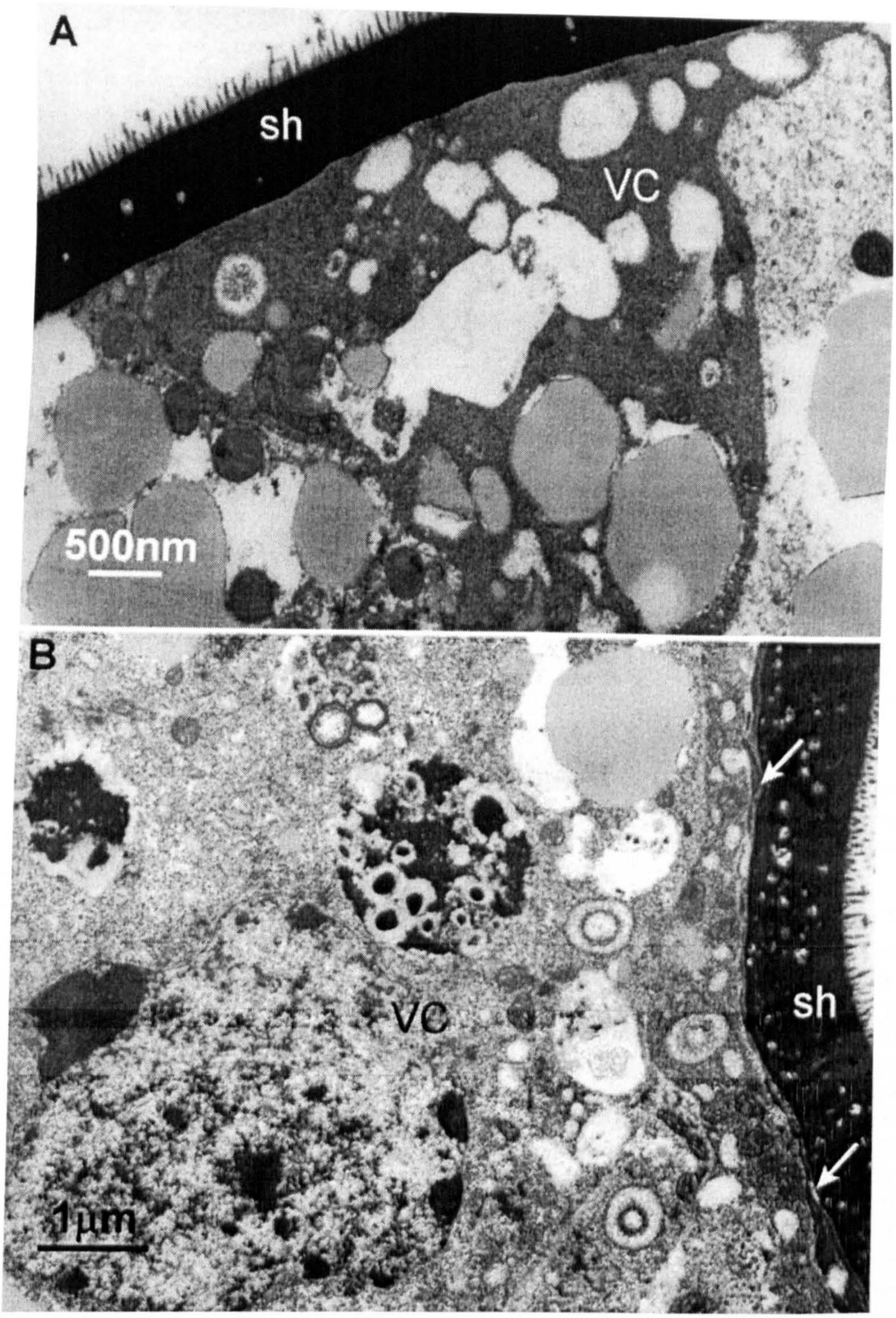


Fig. 1. (A) Transmission electron micrograph through the outer layers of an immature egg of *Schistosoma mansoni*. The only structures visible are the shell (sh) and the enclosed vitelline cells (VC). (B) Transmission electron micrograph through the outer layers of an intermediate egg, showing the presence of cytoplasmic plates (arrowed) interposed between the egg contents and the shell.

PBS–Tween for 30 min. The sections were then incubated in the anti-ESP antiserum diluted 1:300 with 10% normal goat serum in PBS–Tween for 90 min, washed 3 times in PBS–Tween and incubated for 30 min in an FITC-conjugated goat anti-rabbit IgG antibody (Vector Laboratories, Peterborough), diluted 1:100 in PBS–Tween with 10% normal goat serum. After a final wash, the sections were mounted in CitiFluor antifade mountant (CitiFluor, London). As a control, sections were incubated with a 1:300 dilution of the pre-immunization serum from the test rabbit, prior to incubation with secondary antibody, or with secondary antibody alone. To reduce the excessive

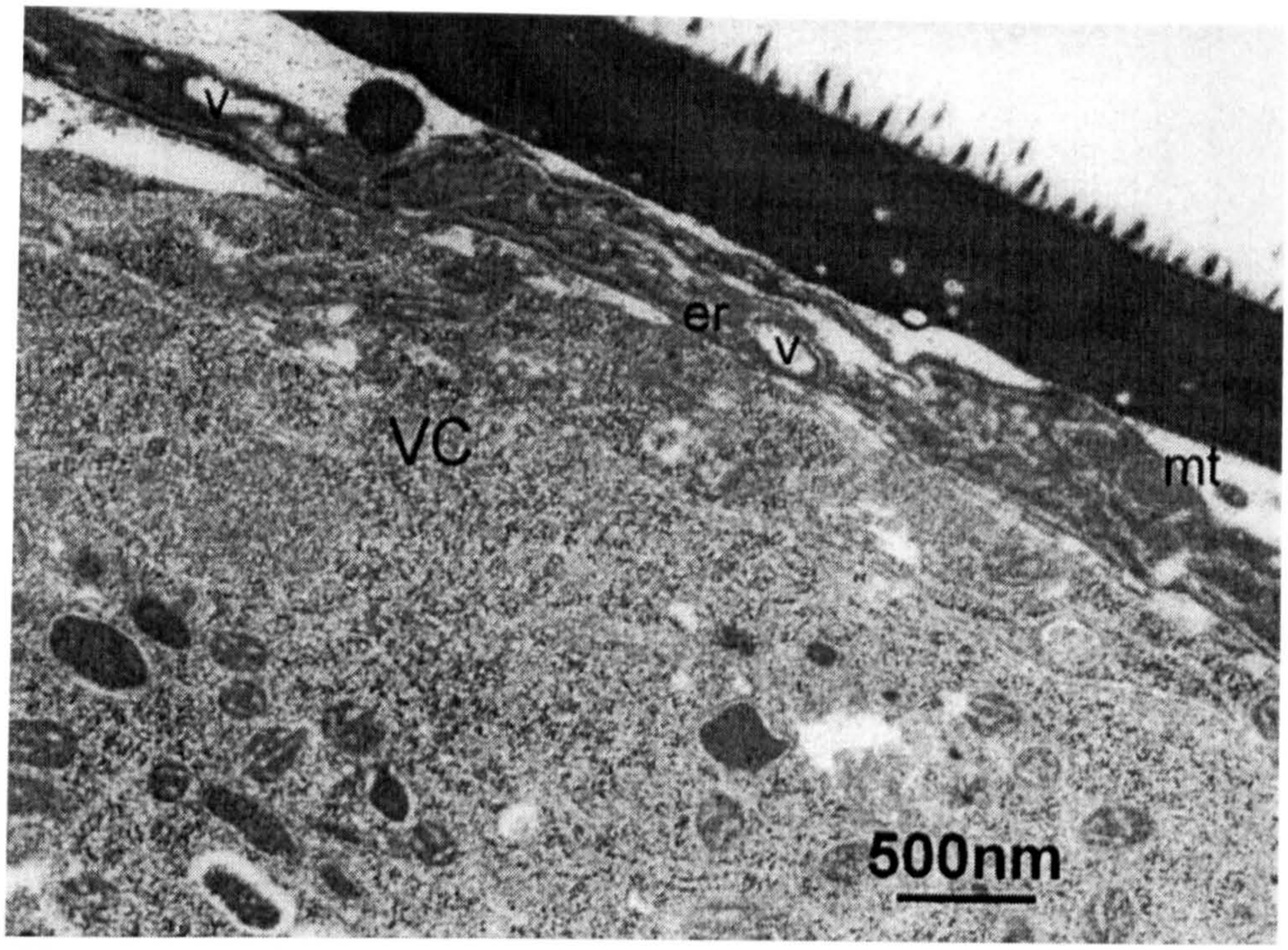


Fig. 2. Transmission electron micrograph through the outer layers of an intermediate egg. The vitelline cells (VC) are still present within the egg, but the envelope is now thicker and contains small mitochondria (mt), endoplasmic reticulum (er), ribosomes and vesicles (v).

yellow autofluorescence contributed by the egg-shell, sections were viewed on a Nikon Labophot fluorescence microscope fitted with a narrow band-pass filter for the green emissions.

Circumoval precipitation

Aliquots of mature and immature eggs were washed in PBS and 50 μl was combined with 50 μl of rabbit anti-ESP serum. This mixture was then pipetted into a chamber on a microscope slide, sealed with vaseline, and incubated for 24 h at 37 °C before examination and photography.

RESULTS

Newly deposited eggs lack complex subshell structures

Viewed by light microscope the newly deposited egg is small ($\bar{x} = 111{\cdot}9 \pm$ S.E. $2{\cdot}2 \times 44{\cdot}0 \pm 0{\cdot}5$ μm, $n = 10$) with highly granular contents and no visible sign of the complex arrangement of membranes and vesicles which surrounds the miracidium in the mature egg ($\bar{x} = 148{\cdot}8 \pm$ S.E. $1{\cdot}4 \times 68{\cdot}7 \pm 1{\cdot}4$ μm, $n = 10$). This lack of structures is confirmed by electron microscopy on undeveloped eggs (Fig. 1A). They are each bounded by the cross-linked protein shell bearing a dense array of microspines on its outer surface. The bulk of the egg contents comprises nucleated vitelline cells with cytoplasm rich in mitochondria, lipid droplets, vesicles of various kinds and granular inclusions, possibly glycogen. The plasma membranes of the vitelline cells lie in direct contact with the inner surface of the egg-shell with no sign of intervening material.

Subshell structures appear as eggs develop

It is difficult to classify precisely how far an individual egg has progressed in development, but by examining a number of eggs it is possible to construct a plausible sequence of events. At an early stage, very thin plates of cytoplasm become interposed between the egg-shell and the intact, metabolically active vitelline cells (Fig. 1B). These plates, bounded on both surfaces by plasma membranes, are presumably extensions from 1 or more parent cells (see Fig. 2) and in thicker parts, internal structure is apparent. As development proceeds, the plates become a continuous layer of cytoplasm 200–300 nm in thickness, separating the vitelline cells from the shell and completely enclosing the egg contents. As such, the layer must represent the proto-envelope. At this stage, the layer has a smooth inner and outer surface; its granular cytoplasm contains small mitochondria and lamellate membranous structures extending laterally throughout. A small amount of flocculent material is apparent in the space between the layer and the shell.

The subshell envelope is a synthetically active tissue

In the fully developed egg, the envelope attains a thickness of 500–600 nm, and has a highly differen-

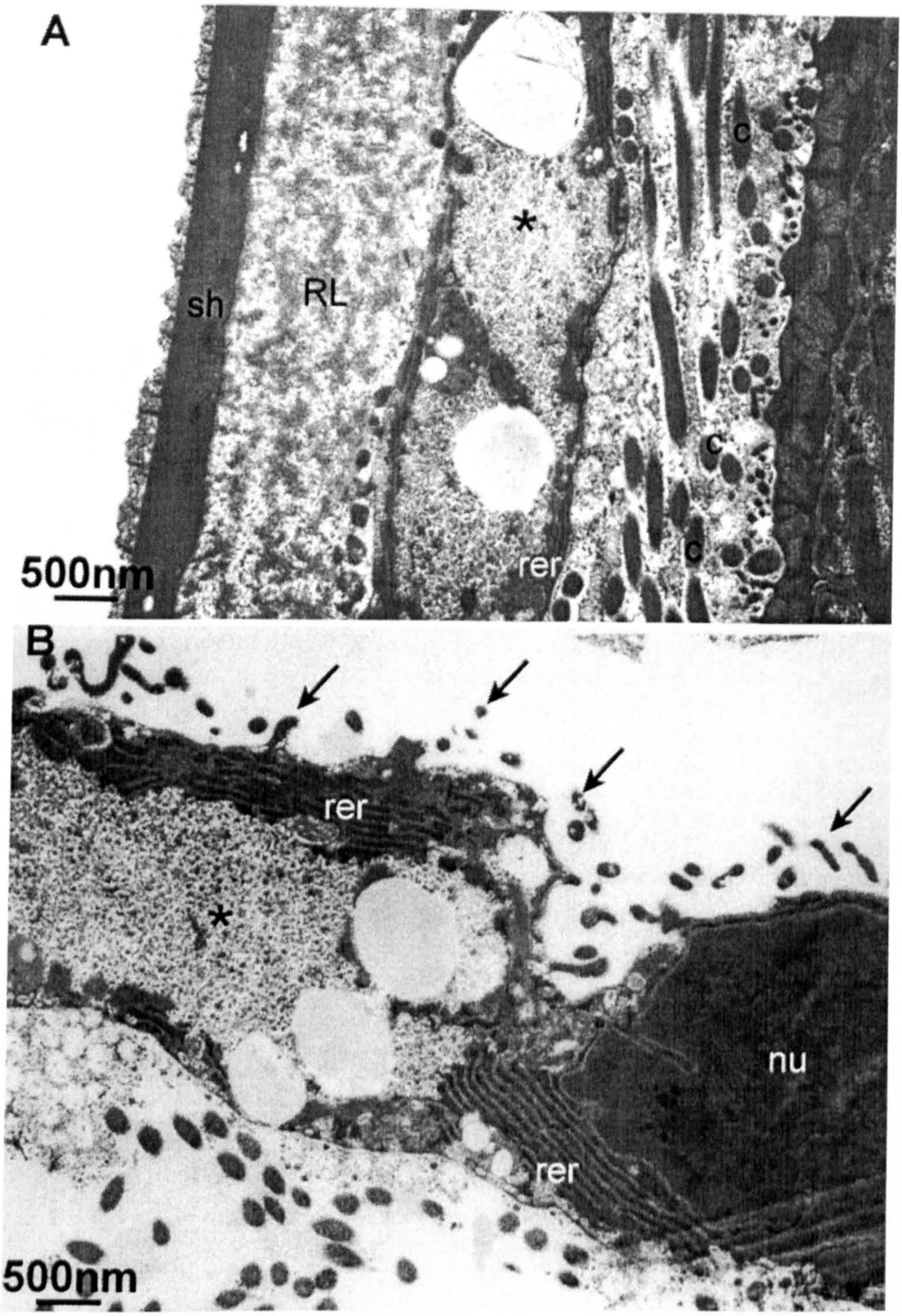


Fig. 3. Transmission electron micrographs through the outer layers of mature eggs. (A) The fully developed envelope lies between the shell (sh) and Reynolds' layer (RL) on the left and the cilia (c) of the miracidium on the right. The presence of large expanses of rough endoplasmic reticulum (rer) is indicative of a protein synthetic role for the envelope. (B) The contents have pulled away from the shell during preparation, highlighting the presence of the projections from the outer surface of the envelope (arrowed). This image also contains one of the nuclei (nu) of the envelope. Also notable in both images is the presence of large expanses of granular material within the envelope (*)

tiated cytoplasm (Fig. 3A, B). Nuclei can clearly be identified confirming the cellular (syncytial?) nature of the layer. The external plasma membrane is drawn out into numerous squat projections reminiscent of microvilli, whilst the inner membrane is smooth. The cytoplasm is divided into distinct regions. Large vesicles are present, probably for lipid storage, whilst the greatest area is occupied by micro-aggregates of granular material. The other most noticeable feature is the expanse of regular,

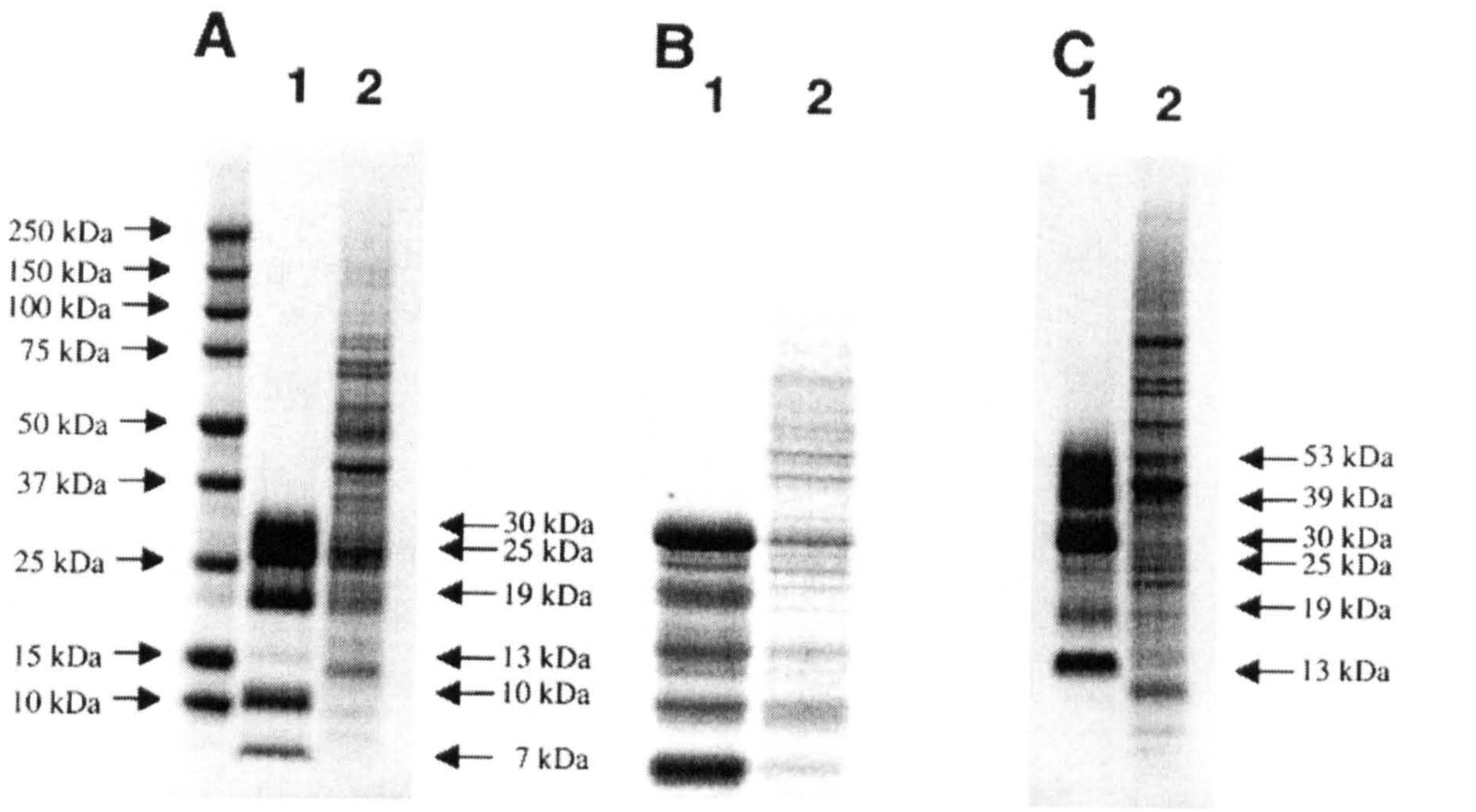


Fig. 4. (A) Coomassie-stained SDS–PAGE gel of ESP (lane 1) and SEA (lane 2). (B) Autoradiograph of an SDS–PAGE gel of ESP and SEA, showing the biosynthetic labelling of ESP during the culture period. (C) SDS–PAGE gel of ESP and SEA run without reducing agents in the sample buffer. Two bands are present in ESP which have higher molecular weights than any band in the reduced sample.

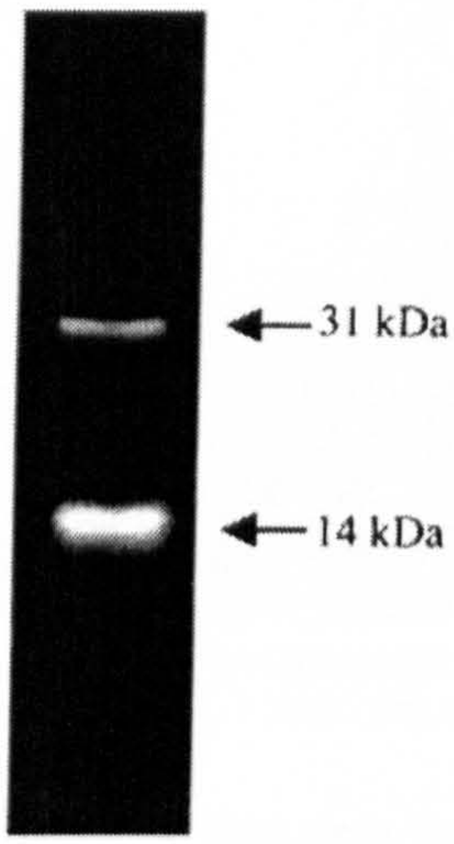


Fig. 5. Coomassie-stained protease substrate gel of ESP. The 2 clear bands indicate the presence of active proteases in the gel.

tightly packed, rough endoplasmic reticulum (ER), indicative of protein synthetic activity. The cytoplasm between the ER cisternae is very dense, and in places contains aggregates of small vesicles. It is possible that the extensive granular micro-aggregates lie in ballooned cisternae of the ER (reminiscent of mammalian plasma cells), rather than free in the cytosol. In the mature egg, the space between the envelope and the shell (Reynolds' layer) is extensive (1 μm wide) and filled with heterogeneous granular material. The microspines on the outer surface of the shell entrap a layer of material, very similar in appearance to the contents of Reynold's layer. The miracidium lies adjacent to the envelope with some of its cilia in direct contact, and is bathed in a milieu of free, dense material, different in texture from either the envelope contents or Reynolds' layer.

Cultured eggs synthesize and release a simple mixture of proteins, containing 2 proteases

Mature eggs isolated from livers and intestines, and cultured *in vitro* for a 3-day period in RPMI-1640, released significant amounts of protein (2·63 μg/1000 eggs) whereas immature eggs were much less active (0·32 μg/1000 eggs). When the filter-concentrated culture supernatants from mature eggs were fractionated by SDS–PAGE a simple pattern of 6 bands (M_r 7, 10, 13, 19, 25 and 30 kDa) was revealed, quite distinct from the very complex pattern characteristic of SEA (Fig. 4A). Indeed, the highly enriched released proteins are either not detectable or present at low abundance in SEA. Mature and immature eggs were cultured in RPMI containing ^{35}S-labelled amino acids, and the TCA-precipitable counts determined. In a typical experiment ($n = 3$) 1000 mature eggs incorporated 86752 cpm during culture whilst immature eggs incorporated a roughly similar 69252 cpm. During the culture period, 1000 mature eggs released 13299 cpm of TCA precipitable material whilst the same number of immature eggs released only 1122 cpm. When the material released by mature eggs was characterized by electrophoresis and detected by autoradiography, a similar pattern of

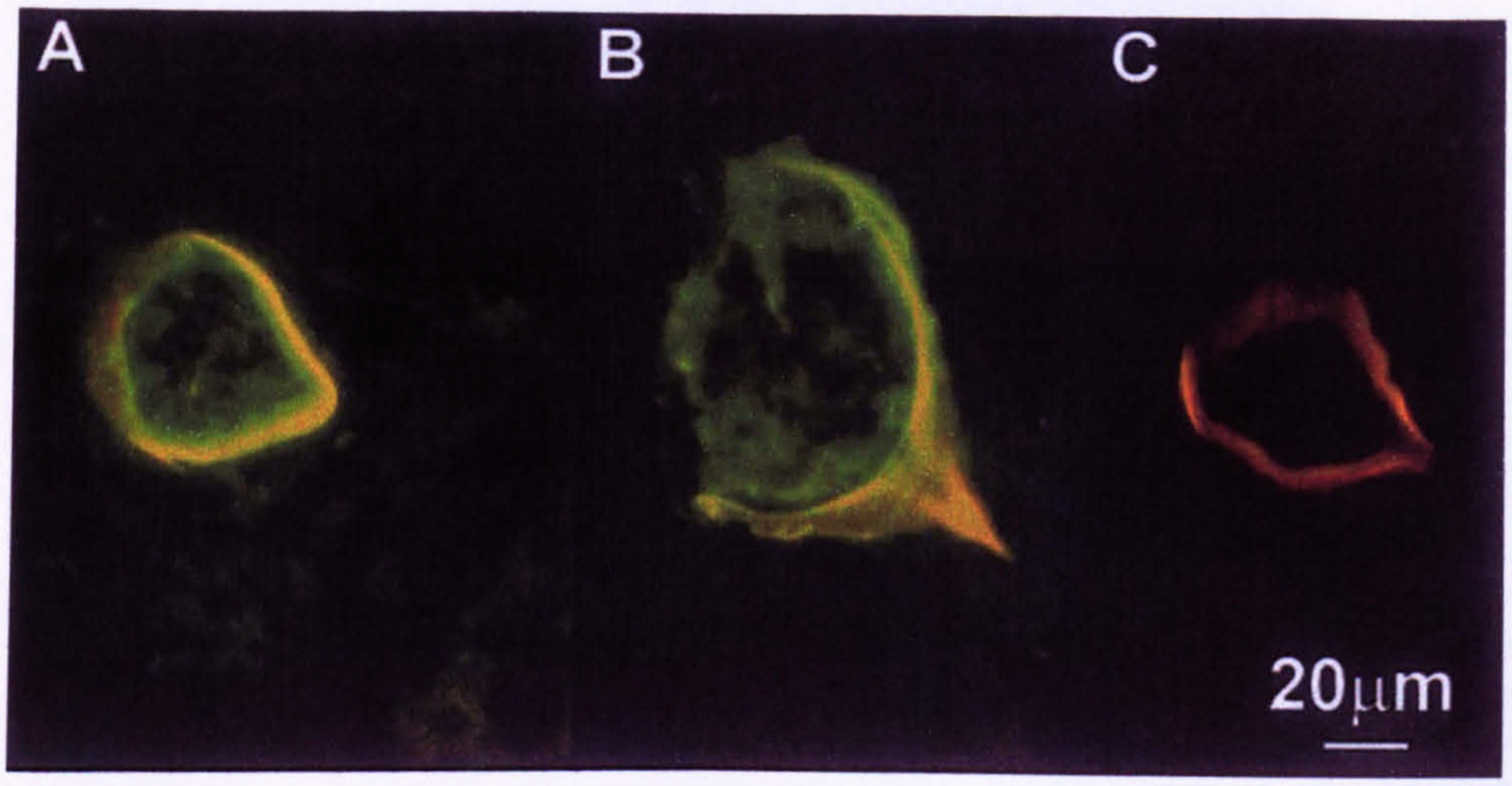


Fig. 6. Immunocytochemical localization using antisera raised against ESP. In addition to the yellow autofluorescence of the shell, the most intense positive (green) staining can be seen immediately beneath the shell, corresponding to the position of the envelope (A) or beneath the shell, and around the outer surface suggestive of release from the egg (B). Sections probed with pre-immunization rabbit serum show no positive reaction (C).

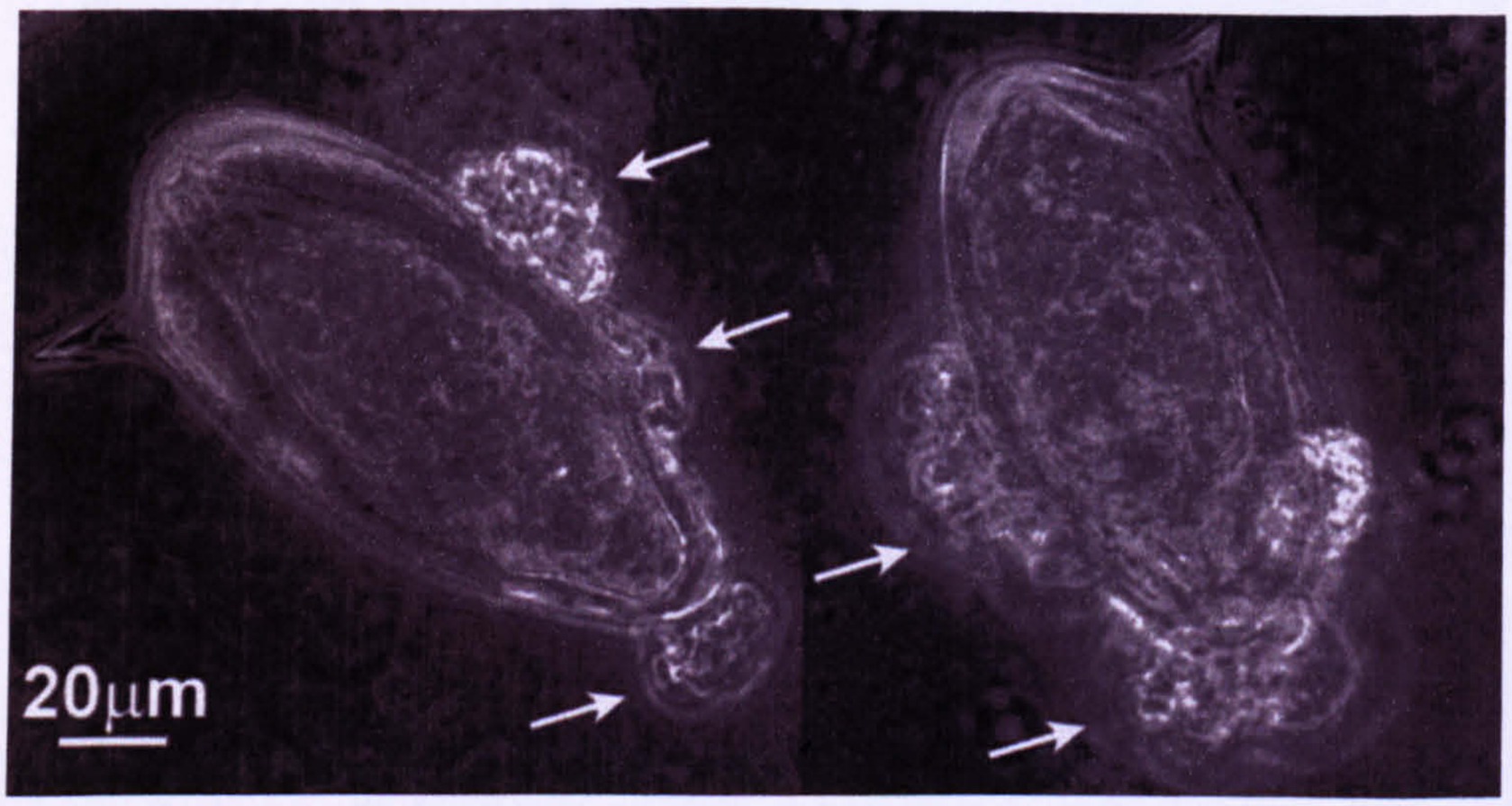


Fig. 7. Circumoval precipitation (arrowed) of ESP proteins by a rabbit anti-ESP antiserum demonstrates the active secretion of ESP from live mature eggs.

6 bands to that seen by Coomassie staining was observed (Fig. 4B), but with differing intensities. This demonstrates the *de novo* synthesis and release of proteins during the 3-day culture period. Furthermore, the pattern of labelled released proteins is closer to that obtained by fractionation of labelled SEA made from cultured eggs, suggesting that in mature eggs the major synthetic activity is the production of protein for secretion.

If the ESP proteins are separated without the presence of a reducing agent in the sample buffer (Fig. 4C), the pattern of bands is slightly different, with the 2 lowest molecular weight bands (7, 10 kDa) absent, and an additional 2 bands present at higher molecular weight (39, 53 kDa). This indicates that at least some of the proteins in the mixture exist as disulphide-bonded oligomers. Inclusion of casein as a protease substrate in the SDS–PAGE gel prior to electrophoresis allowed the detection of proteolytic activity in the ESP mixture (Fig. 5). The 2 distinct clear bands (molecular weights, 14 and 31 kDa, corresponding approximately with the bands at 13 and 30 kDa), where the substrate has been digested, demonstrate that there are at least 2 separate proteases secreted by the egg.

Immunocytochemistry reveals that egg secretions originate in the subshell envelope

Immunocytochemistry on cryostat sections of infected mouse liver revealed that, among all the subshell structures, the most intense staining was in the envelope (Fig. 6). Positive reactions were seen immediately outside the egg-shell and could also be discerned in the surrounding granulomatous tissue, suggestive of release and diffusion away of the

secreted proteins. No reactivity was observed with either pre-immunization serum or with the secondary antibody alone.

Circumoval precipitation demonstrates secretion by live eggs

Incubation of live mature eggs for 24 h in the presence of rabbit anti-ESP serum resulted in the formation of characteristic circumoval precipitations (Fig. 7). These were localized to particular regions, rather than being uniformly distributed around the circumference of the egg. This implies variations in shell permeability, but with no obvious pattern between eggs. No such precipitations were observed around immature eggs or when normal rabbit serum was used as a control.

DISCUSSION

In view of the importance of the schistosome egg in pathogenesis, it is remarkable that its basic 'physiology' has received so little attention. The newly laid egg of *S. mansoni* shows a relatively simple organization, with a tanned protein shell surrounding a single ovum and 20–30 vitelline cells. The lack of a subshell envelope at this stage (so evident in mature eggs) is not perhaps surprising since its presence would impede the release of granules from the vitelline cells and their coalescence to form the shell. Thus, the plasma membranes of the numerous vitelline cells lie in direct contact with the egg-shell. When such eggs are cultured *in vitro* in defined medium plus a serum supplement they will develop to maturity in 7 days (Michaels & Prata, 1968). The absence of development in simple medium strongly suggests a need for an external source of nutrients and growth factors, in complete contrast to the egg of *Fasciola hepatica* which is self-contained and will develop to maturity in distilled water (Wilson, 1967). This need is confirmed by the uptake and incorporation by schistosome eggs of labelled amino acids into newly synthesized proteins, and by the increase in egg size during development. Comparison of the volumes of newly laid and mature eggs, using the formula for a prolate spheroid ($\frac{4}{3}\pi ab^2$), reveals a 3·24-fold increase during development. Our method for separating undeveloped and mature eggs, using Percoll density gradients relies on densities of $<1{\cdot}06$ and $>1{\cdot}09$ g/ml respectively. This implies that a schistosome egg increases in mass by a factor of 3·33 during development, i.e. more than 2/3 of the constituents of the fully mature egg are derived from external nutrients. The increased size also reveals that the egg-shell must have considerable elasticity.

We observed that early in development, cytoplasmic plates interpose between the vitelline cells and the shell. The subsequent growth and differentiation of this layer and the identification of nuclei within it, reveals its cellular origin. Indeed, in this respect, *S. mansoni* does not appear to differ from other digenetic trematodes, where this envelope is referred to as a vitelline membrane (strictly inaccurate, since it is not the direct secretion of a fertilized egg). Morphological studies on digeneans as diverse as *Fasciola* (Schubmann, 1905; Ortmann, 1908), *Paragonimus* (Chen, 1937) and *Parorchis* (Rees, 1940), showed that the envelope is derived early in development (after the first 7 or 8 divisions of the ovum) from cells which separate from the embryonic miracidium. These cells migrate to the inner surface of the shell where they flatten and extend to form the continuous lining. In our present study, this phase is represented by the partially differentiated egg, where mitochondria are evident within the envelope, implying metabolic activity; the first signs of an endomembrane system are also apparent. From the continuous and cellular nature of this layer, it is clear that it must exert significant control over the transport of materials into and out of the egg.

Our observations on the mature envelope reveal a more organized structure than reported by Neill *et al.* (1988). We can ascribe this in part to our use of glutaraldehyde as a primary fixative on isolated eggs, together with the puncture of the egg-shell to allow ingress of resin, compared with osmium fixation after slam freezing and freeze substitution of intestinal tissue samples in the earlier study. The presence of extensive rough ER indicates high levels of protein synthetic activity. The occurrence of extensive aggregates of granular material in the mature envelope may imply storage of the newly synthesized protein prior to exocytosis from the layer. An analogous situation is found in the egg of *Fasciola* where the microfibrils of the viscous cushion involved in egg hatching are laid down entirely within the confines of the equivalent vitelline membrane layer (Wilson, 1967). In the *S. mansoni* egg, since the vesicles in the envelope associated with the ER lie close to both the inner and outer bounding membranes, we cannot make any deductions from the morphology of this layer about the direction of export. The presence of microvillar-like extensions on the outer surface of the mature (but not immature) envelope supports an inward transport function for this layer, presumably of the large quantities of nutrients required to explain the increase in mass. The maximum thickness of the envelope and its full capacity for protein synthesis are only apparent in the mature egg.

The occurrence of macromolecular material between the egg-shell and the envelope is slight in the partially developed egg, but becomes very evident as the egg matures and is sufficiently prominent to be termed Reynolds' layer by Neill *et al.* (1988). The origin of this layer from outside the egg seems unlikely, since the shell is inert and so would not operate differentially in partially and fully developed

eggs (unless the stretching of the shell as egg volume increased opened pores to allow the ingress of host proteins). A more likely explanation is that the material in Reynolds' layer originates either within the envelope itself, or from the internal milieu surrounding the miracidium. An origin within the envelope accords with the large expanses of granular material in the cytoplasm. Since the envelope and Reynolds' layer reach their maximum thickness and complexity as the egg matures, it is tempting to conclude that their formation is related to the process of egg transit through the intestinal tissues. Cultured mature eggs secreted much more protein than immature eggs, an observation confirmed by the relative amounts of biosynthetically labelled protein released. We suggest that this is a device to prevent premature expulsion of the eggs from the tissue environment which is clearly conducive to their development. In this context, the accumulation of Reynolds' layer and the slow diffusion of its constituents through the egg-shell pores (Race *et al.* 1969), suggest that the egg acts as a slow release capsule. Furthermore, the presence of similar granular material entrapped by the surface microspines may represent the sequestration of secretions in the immediate vicinity of the egg, thus focussing the host inflammatory response. We raised antibodies against ESP to investigate the relationship between egg development and secretion using the circumoval precipitation reaction as an indicator. Our observations confirm the abundant egress of proteins from the mature, but not the immature, egg.

The pattern of proteins secreted by mature eggs (ESP) was markedly different from the total egg homogenate represented by the standard SEA preparation. No constituent of ESP had an M_r of >40 kDa whereas constituents of SEA ranged up to >200 kDa. This could imply size selection by pores in the egg-shell, or partial degradation of proteins after release. The distinct patterns indicate a compartmentalization of egg proteins which could presage a difference in host responses to the products of live and dead eggs, with obvious implications for granuloma formation. The ability to label the subset of released proteins biosynthetically confirms that they are actively secreted molecules, not simply the leakage products of dead eggs. In contrast to the disparity between ESP and SEA, there were similarities in composition between the ^{35}S-labelled proteins in the 2 fractions from cultured eggs. We interpret this to mean that the synthetic activity of mature eggs is largely centred on the secreted subset of proteins. This accords with the putative function of these proteins in the process of egg migration through the tissues and our identification of 2 proteases secreted by the mature egg is particularly pertinent in this context. Their presence supports the possibility that secreted proteins are partially degraded during the culture period. It is difficult to relate our observations to previous studies which have characterized major egg antigens from SEA. When Western blots of ESP and SEA were probed with monoclonal antibodies against p40 (Nene *et al.* 1986), kindly donated by Dr M. Stadecker, a single band was recognized in SEA, but no reactivity was seen in ESP, suggesting that p40 is not present. Probing with polyclonal antisera raised against other native SEA antigens (e.g. α_1 and ω_1, Dunne *et al.* 1981) resulted in the recognition of multiple bands in both ESP and SEA, suggesting that the antibodies recognized common glycan epitopes present on several proteins.

Immunocytochemistry with an antiserum directed against the secretory proteins strongly suggests that they localize within the envelope, not the miracidium or its surrounding milieu. Thus, the presence of protein synthetic machinery in the mature envelope, the likelihood that it is a barrier to passive diffusion, the coincidence of its differentiation with this onset of ESP secretion and the immunolocalization of ESP all lead us to conclude that the envelope is the source of the macromolecules which mediate egg transit through the gut tissues. We are currently undertaking the detailed characterization of the egg secretions, especially the 2 proteases, and their roles in egg excretion or as initiators of hepatic pathology.

This work was funded by grants from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases and European Commission's BIOTEC programme. B.S. was funded by a BBSRC CASE Studentship with the Natural History Museum, London. We would like to acknowledge the technical assistance of Mr Srdjan Ljubojevic with the recovery and culture of the eggs.

REFERENCES

BOROS, D. L. & WARREN, K. S. (1970). Delayed hypersensitivity-type granuloma formation and dermal reaction induced and elicited by a soluble factor isolated from *Schistosoma mansoni* eggs. *Journal of Experimental Medicine* **132**, 488–507.

BRADFORD, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye-binding. *Analytical Chemistry* **72**, 248–254.

CHEN, P.-D. (1937). The germ cell cycle in the trematode *Paragonimus kellicotti* (Ward). *Transactions of the American Microscopical Society* **56**, 208–236.

DOENHOFF, M. J., HASSOUNAH, O., MURARE, H., BAIN, J. & LUCAS, S. (1986). The schistosome egg granuloma: immunopathology in the cause of host protection or parasite survival? *Transactions of the Royal Society of Tropical Medicine and Hygiene* **80**, 503–514.

DUNNE, D. W., LUCAS, S., BICKLE, Q., PEARSON, S., MADGWICK, L., BAIN, J. & DOENHOFF, M. J. (1981). Identification and partial purification of an antigen (ω_1) from *Schistosoma mansoni* eggs which is putatively hepatotoxic in T-cell deprived mice. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **75**, 54–71.

HANG, L. M., WARREN, K. S. & BOROS, D. L. (1974). *Schistosoma mansoni*: antigenic secretions and the etiology of egg granulomas in mice. *Experimental Parasitology* **35**, 288–298.

KARANJA, D. M., COLLEY, D. G., NAHLEN, B. L., OUMA, J. H. & SECOR, W. E. (1997). Studies on schistosomiasis in western Kenya: I. Evidence for immune-facilitated excretion of schistosome eggs from patients with *Schistosoma mansoni* and human immunodeficiency virus coinfections. *American Journal of Tropical Medicine and Hygiene* **56**, 515–521.

LOCKWOOD, B. C., NORTH, M. J., SCOTT, K. I., BREMNER, A. F. & COOMBS, G. H. (1987). The use of a highly sensitive electrophoretic method to compare the proteinases of trichomonads. *Molecular and Biochemical Parasitology* **24**, 89–95.

MICHAELS, R. M. & PRATA, A. (1968). Evolution and characteristics of *Schistosoma mansoni* eggs laid *in vitro*. *Journal of Parasitology* **54**, 921–930.

NEILL, P. J., SMITH, J. H., DOUGHTY, B. L & KEMP, M. (1988). The ultrastructure of the *Schistosoma mansoni* egg. *American Journal of Tropical Medicine and Hygiene* **39**, 52–65.

NENE, V., DUNNE, D. W., JOHNSON, K. S., TAYLOR, D. W. & CORDINGLEY, J. S. (1986). Sequence and expression of a major egg antigen from *Schistosoma mansoni* – homologies to heat-shock proteins and alpha-crystallins. *Molecular and Biochemical Parasitology* **21**, 179–188.

ORTMANN, W. (1908). Zur Embryonalentwicklung des Leberegels (*Fasciola hepatica*). *Zoologische Jahrbücher, Abteilung für Anatomie und Ontogenie* **26**, 255–292.

RACE, G. J., MICHAELS, R. M., MARTIN, J. H., LARSH, J. E. Jr. & MATTHEWS, J. L. (1969). *Schistosoma mansoni* eggs: an electron microscopic study of shell pores and microbarbs. *Proceeding of the Society of Experimental Biology and Medicine* **130**, 990–992.

REES, G. (1940). Studies on the germ cell cycle of the digenetic trematode *Parorchis acanthus* (Nicoll). *Parasitology* **32**, 372–391.

SCHUBMANN, W. (1905). Über die Eibildung und Embryonalentwicklung von *Fasciola hepatica* L. (*Distomum hepaticum* Retz.). *Zoologische Jahrbücher, Abteilung für Anatomie und Ontogenie* **21**, 571–606.

WILSON, R. A. (1967). The structure and permeability of the shell and vitelline membrane of the egg of *Fasciola hepatica*. *Parasitology* **57**, 47–58.

Addendum to Chapter 2: Immune Responses to ESP

Introduction

Chapter two established, through the use of the circumoval precipitation test, that the proteins present in the ESP preparation were secreted from live eggs and that they were recognised by the immune response during infection. As such, they are candidates for the proteins which initiate granuloma formation and in order to confirm whether they play a role in this process, it is necessary to examine the immune responses to both ESP and SEA. Clearly, both will be seen by the immune system in the course of an infection, but the hypothesis is that ESP is the deliberate secretion of the egg, and is therefore required, either directly or via the immune response that it stimulates, for the processes of egg escape and (coincidentally) initiation of granuloma formation. In contrast, I suggest that immune recognition of SEA results only when the eggs die within the host.

Several experimental methods have been used to study the process of granuloma formation and these approaches vary in two ways: the material used for the challenge, and the method of challenge. In order to mimic the situation in a natural infection as closely as possible, it would be desirable to introduce viable eggs into an animal so that they lodged in the liver. This would enable the responses to the eggs to be studied without any confounding factors, such as the responses induced by larval and adult worm antigens. Additionally, the formation of granulomas around the eggs would be synchronous, with all the granulomas being at the same stage, unlike during an infection, where eggs are laid continuously, and granulomas of many different stages of development can be observed at any particular time. However, to do this requires that surgery be performed on the mouse and the eggs injected into one of the systemic veins immediately upstream of the liver, e.g., the superior mesenteric vein. An alternative approach is to inject the eggs e.g., into the tail vein, from where they will be carried

through the circulation and lodge in the capillary beds of the lungs [Peterson and von Lichtenberg, 1965]. This has the obvious advantage of convenience, since the mice do not need to undergo surgery to perform the injection, but is open to the criticism that the site of granuloma formation is different from that during a natural infection.

Whichever method is used to introduce the eggs, there is still the problem that it will not be clear whether the responses that are measured will be directed towards the egg secretions (i.e., ESP) or other components of the egg (SEA). To resolve this, it is necessary to introduce particles that are similar in size to eggs, but which deliver material of a defined composition. To do this, a method was developed for covalently coupling antigens onto cyanogen bromide-activated Sepharose beads [van Marck *et al.*, 1980], which are comparable in size (~100μm) to the schistosome egg.

In this chapter, the responses were measured in two ways. ESP- and SEA-coupled Sepharose beads were injected into the superior mesenteric vein of mice to lodge in the liver and the granulomas so induced were examined by histology. It was observed that there were differences in the cellular composition and the degree of fibrosis between the granulomas around each type of bead. In addition, antibody responses to ESP and SEA were measured by ELISA after injection of eggs into the peritoneal cavity of naïve mice and during a natural infection and demonstrate that, in this situation, the responses to ESP and SEA are distinct and have little cross-reactivity.

Materials and Methods

Parasite Material and Priming of Mice

ESP and SEA were prepared as described in chapter 2. Six C57BL/6 mice were primed by intraperitoneal injection with ~3000 mature eggs and maintained for three weeks. A second group of 20 mice were injected with eggs i.p. and maintained for up to five weeks. Each

week four mice were anaesthetised, terminally bled by cardiac puncture and serum prepared by clotting and centrifugation. A third group of 24 mice were infected with 30-50 cercariae each and four mice were sacrificed at weeks 2, 4, 6, 8, 10 and 12 post-infection, bled and serum prepared.

Coupling of ESP and SEA to Sepharose Beads

Cyanogen bromide (CNBr)-activated Sepharose beads were purchased from Amersham Pharmacia Biotech and were re-swelled according to the manufacturer's instructions. To perform the coupling, 100μg of each preparation was added to a tube containing 100μl of reswelled beads and incubated for one hour at room temperature. Uncoupled reactive sites on the beads were then blocked by incubation for 2 hours with 100mM glycine. Control beads were incubated with glycine alone.

Intravenous Injection of Sepharose Beads

The six primed mice were anaesthetised and, after laparotomy, 30μl of the bead suspension was injected into the superior mesenteric vein, so that the beads would lodge in the pre-sinusoidal vessels of the livers, analogous to eggs during infection. Two mice were injected with SEA-coupled beads, two with ESP-coupled beads and two with control beads. At days four and eight post-injection, one mouse from each group was sacrificed and the liver removed. Pieces of liver were fixed in 4% formaldehyde before embedding in paraffin wax, sectioning and staining with haematoxylin and eosin.

ELISAs of Antisera against ESP and SEA

SEA and ESP (prepared as described in Chapter 2) were coated onto ELISA plates (Nunc) in PBS (0.1μg/ml, 100μl/well) overnight. The plates were washed four times in PBS containing 0.05% Tween-20 (Sigma), and incubated with doubling dilutions of the primary antibody. The plates were again washed in PBS/Tween-20 before incubation with a horseradish peroxidase-conjugated anti-mouse IgG secondary antibody

(1:10000 dilution). After a third series of washes, the plates were incubated with TMB peroxidase substrate (Dynex) and left to develop. At intervals of ten minutes, the plates were read on a microplate reader (Dynex), using Revelation software, at 630nm. Once the colour in the plates had developed sufficiently, the reaction was stopped by the addition of 50µl of 12.5% sulphuric acid and the plate re-read at 450nm.

Results

Intraperitoneal Injection of Eggs Reveals Distinct Responses to ESP and SEA

If the roles of ESP and SEA differ in granuloma formation, so should the antibody responses that are directed against them. In order to investigate this, responses were measured during the course of a natural infection. Figure 3-1A shows the antibody response to ESP and SEA over a 12-week period. No response is observed up to and including week 6, (the time at which egg deposition begins), and both responses subsequently rise together, with no significant difference between the profiles at any time point. This suggests that there is no cross-reactivity between the egg proteins and any larval or adult antigens that are accessible to the immune system before the onset of egg-laying.

However, after the i.p. injection of eggs, the responses do differ. Figure 3-1B shows that the response to ESP rises immediately after injection followed by the SEA response after a lag of ~10 days. This strongly suggests that the ESP is seen by the immune system before SEA, which is presumably only released once the eggs die.

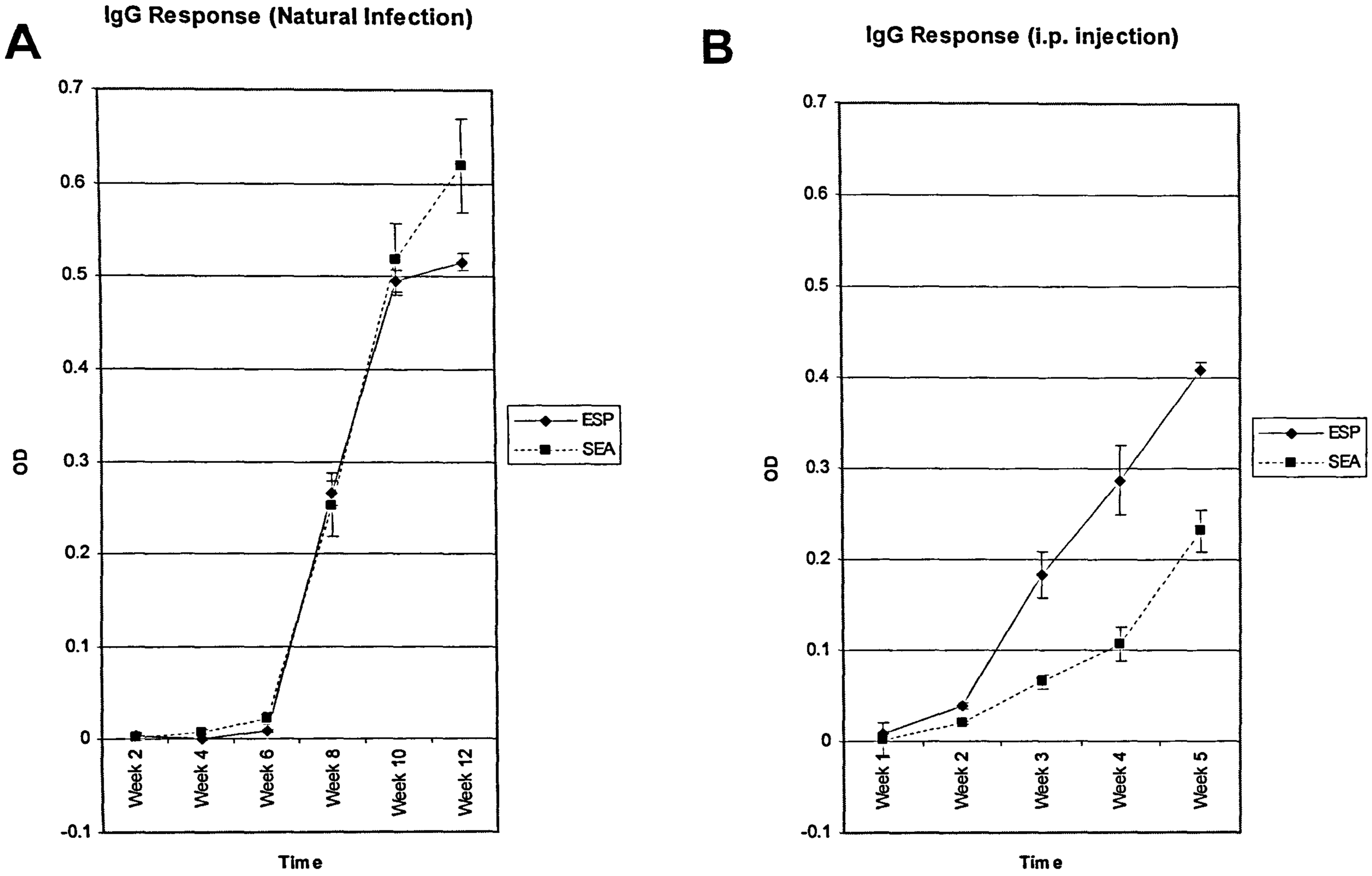
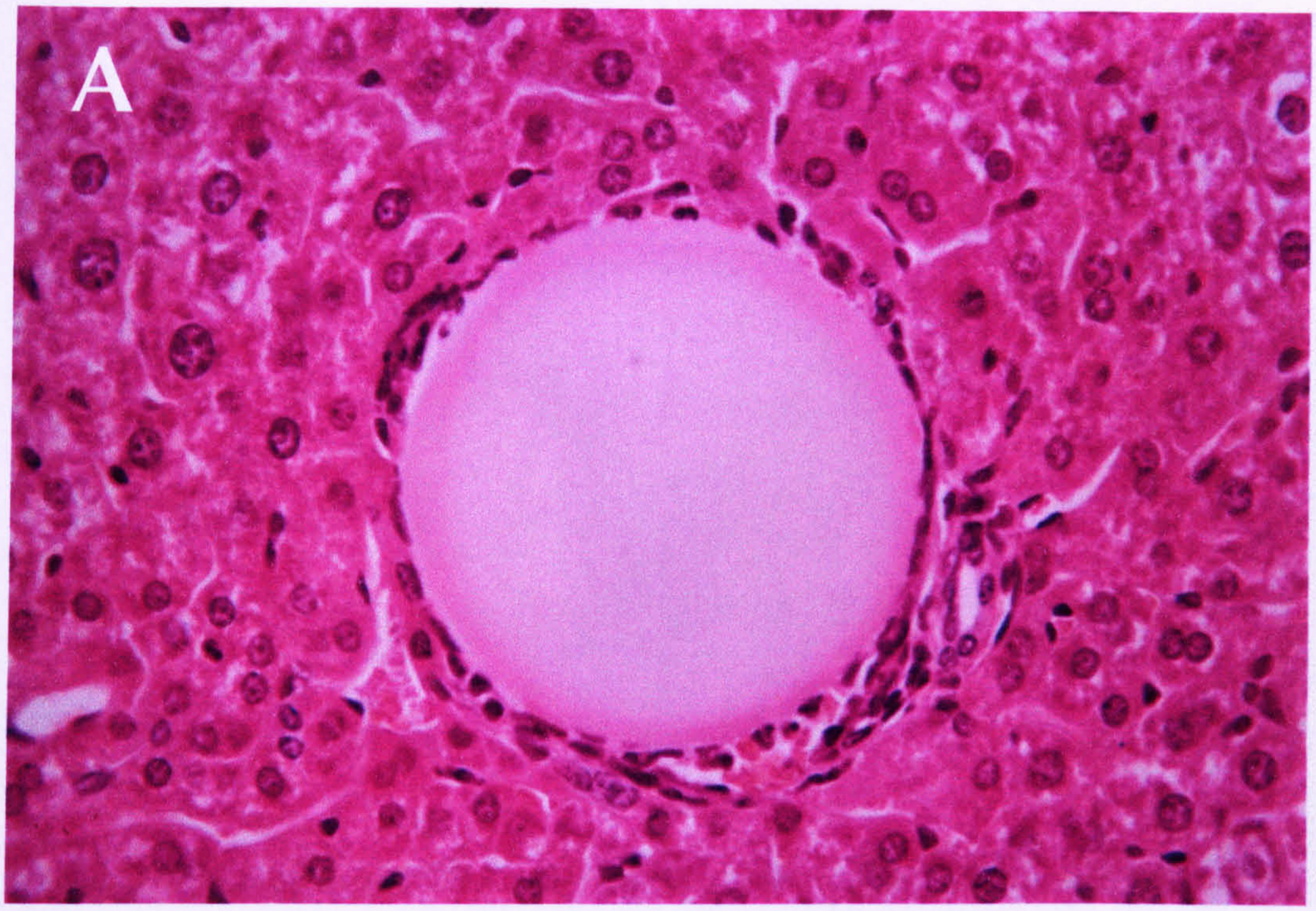


Figure 3-1: IgG antibody responses against ESP and SEA (A) during a natural infection and (B) after intraperitoneal injection of viable, mature eggs. Data are plotted as the means of the responses of four mice ± SEM.

Injection of Antigen-Coupled Beads induces Granuloma Formation

Demonstration of an antibody response to both SEA and ESP does not provide direct evidence that either is involved in the process of granuloma formation. In order to establish whether any differences could be discerned in the responses to the two preparations, ESP- and SEA-coupled Sepharose beads were injected into the superior mesenteric vein of primed mice and the livers examined four and eight days after injection. When control beads are injected (Figure 3-2), no response is seen at either time-point and the endothelium of the blood vessel can be seen in both cases, surrounded by normal hepatic tissue.

However, granulomas do form around beads which have antigen coupled on their surface. Even by four days post-injection, large lesions have found around both type of beads, resembling the granulomas observed around eggs in a normal infection (Figure 3-3). Although the granulomas are superficially similar, comprising a sharply circumscribed accumulation of cells 10-20 deep around each bead, some differences can be discerned. In particular, it can be seen that cells present around the SEA-coupled bead are mostly mononuclear cells, whereas those around the ESP-coupled bead contain many more polymorphonuclear cells (arrows), most probably eosinophils. By day eight (Fig 3-4), the differences are more pronounced, with the granuloma surrounding the SEA-coupled beads apparently showing more fibrosis than that around the ESP-coupled beads, with whorls of what is probably collagen present (*).



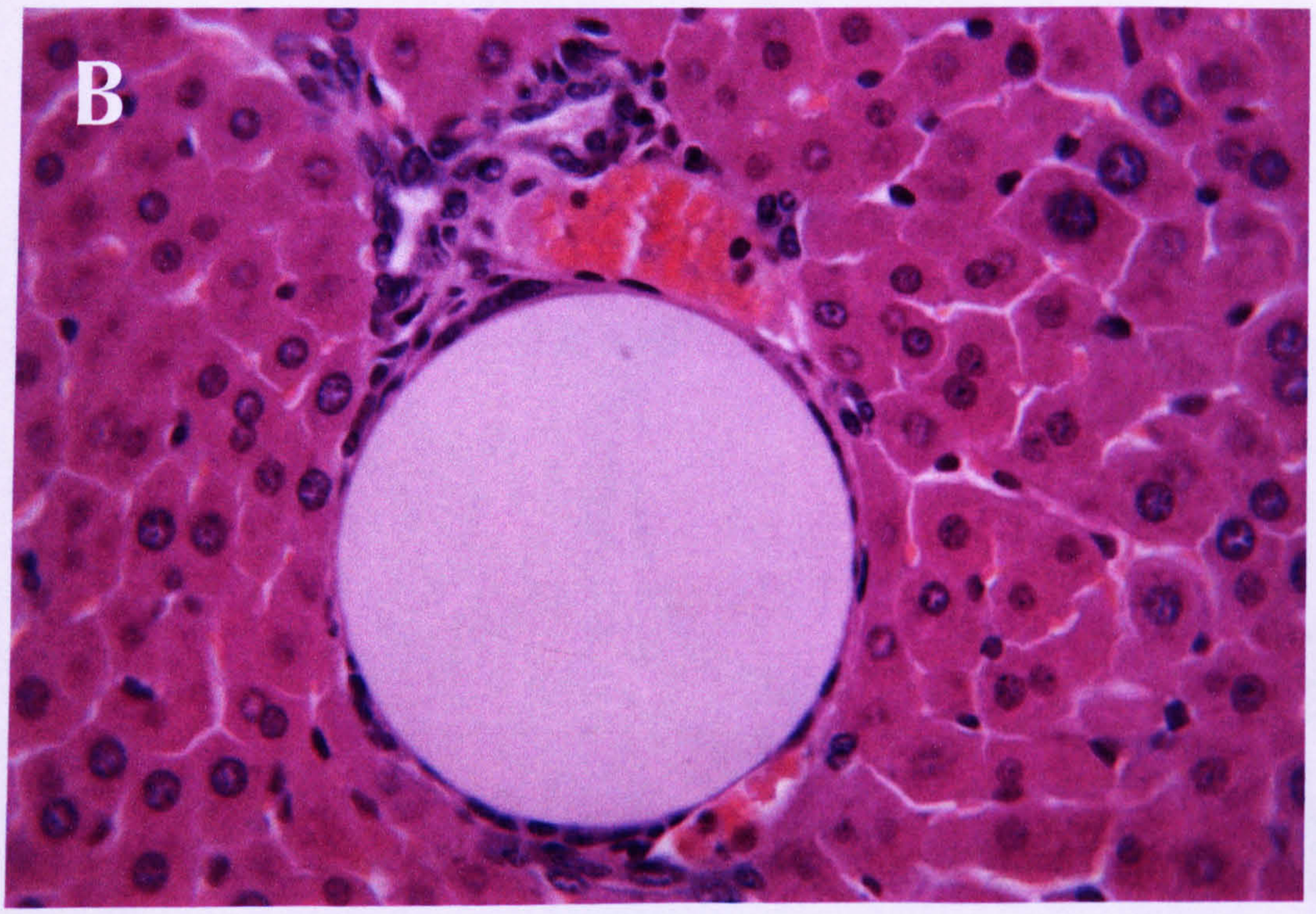


Fig 3-2: Sections of liver containing Sepharose beads blocked with glycine, four days (A) and eight days (B) after injection, stained with haematoxylin and eosin.

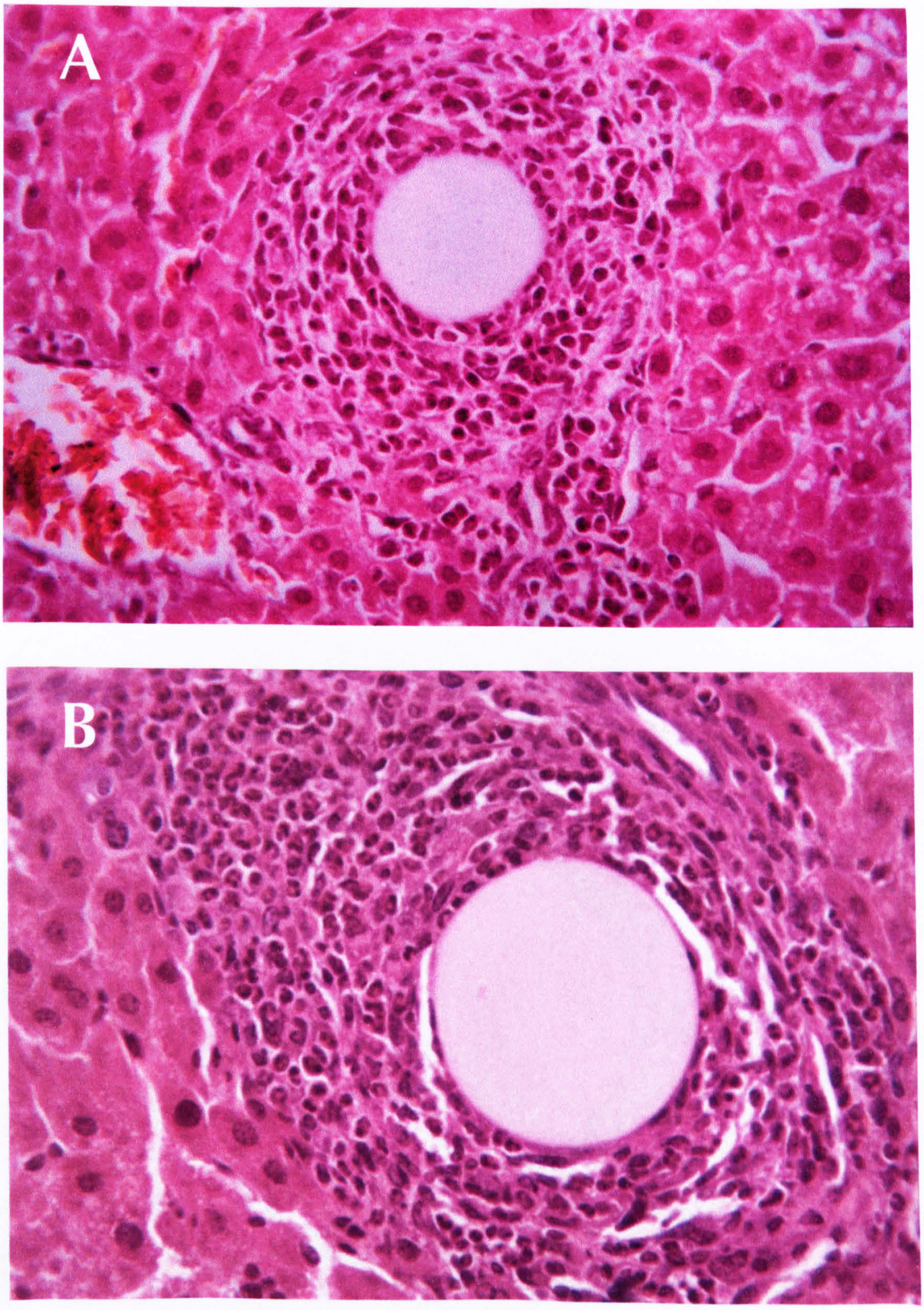
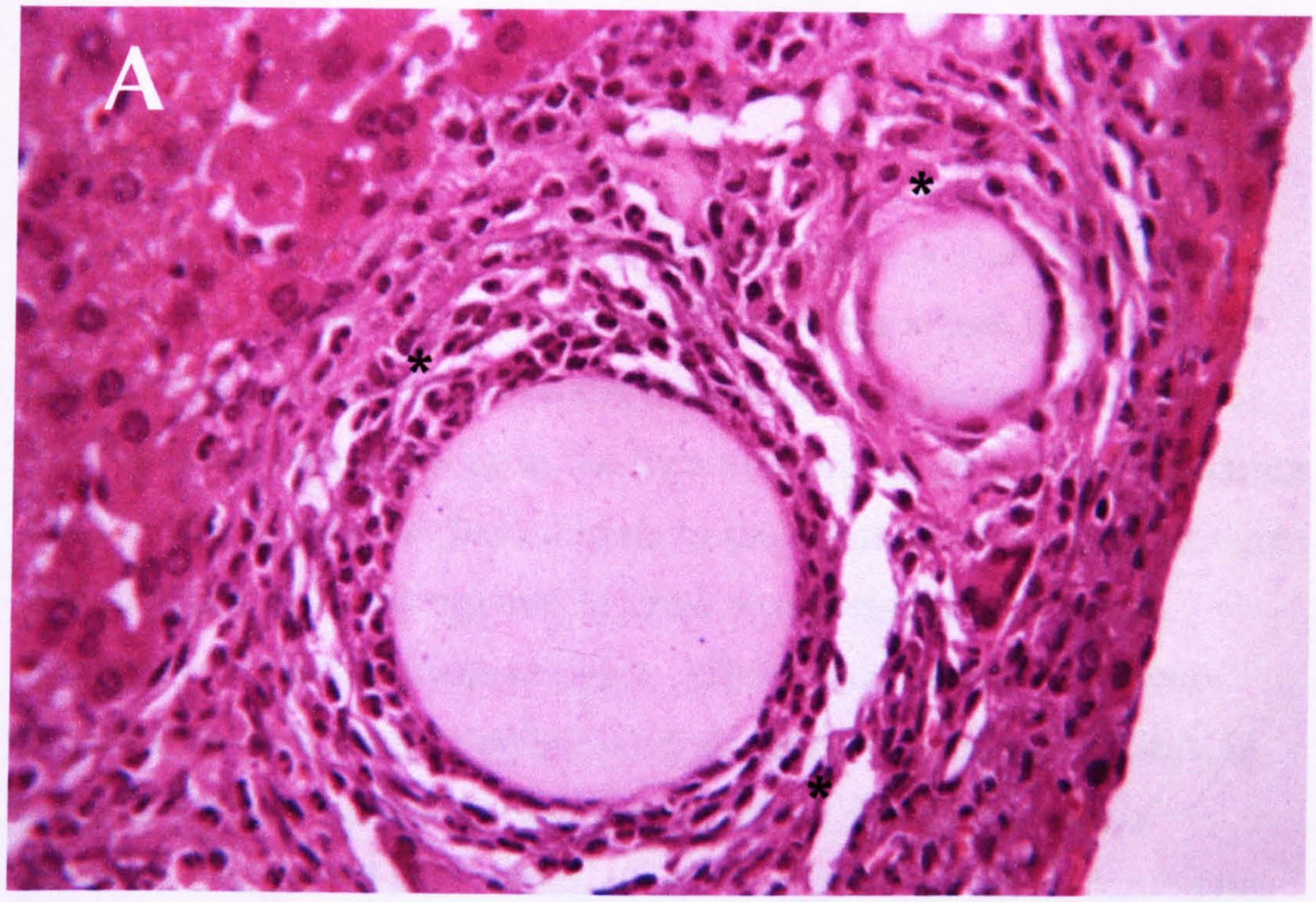


Fig 3-3: Liver sections containing Sepharose beads coupled with (A) SEA and (B) ESP, four days after injection, stained with haematoxylin and eosin.



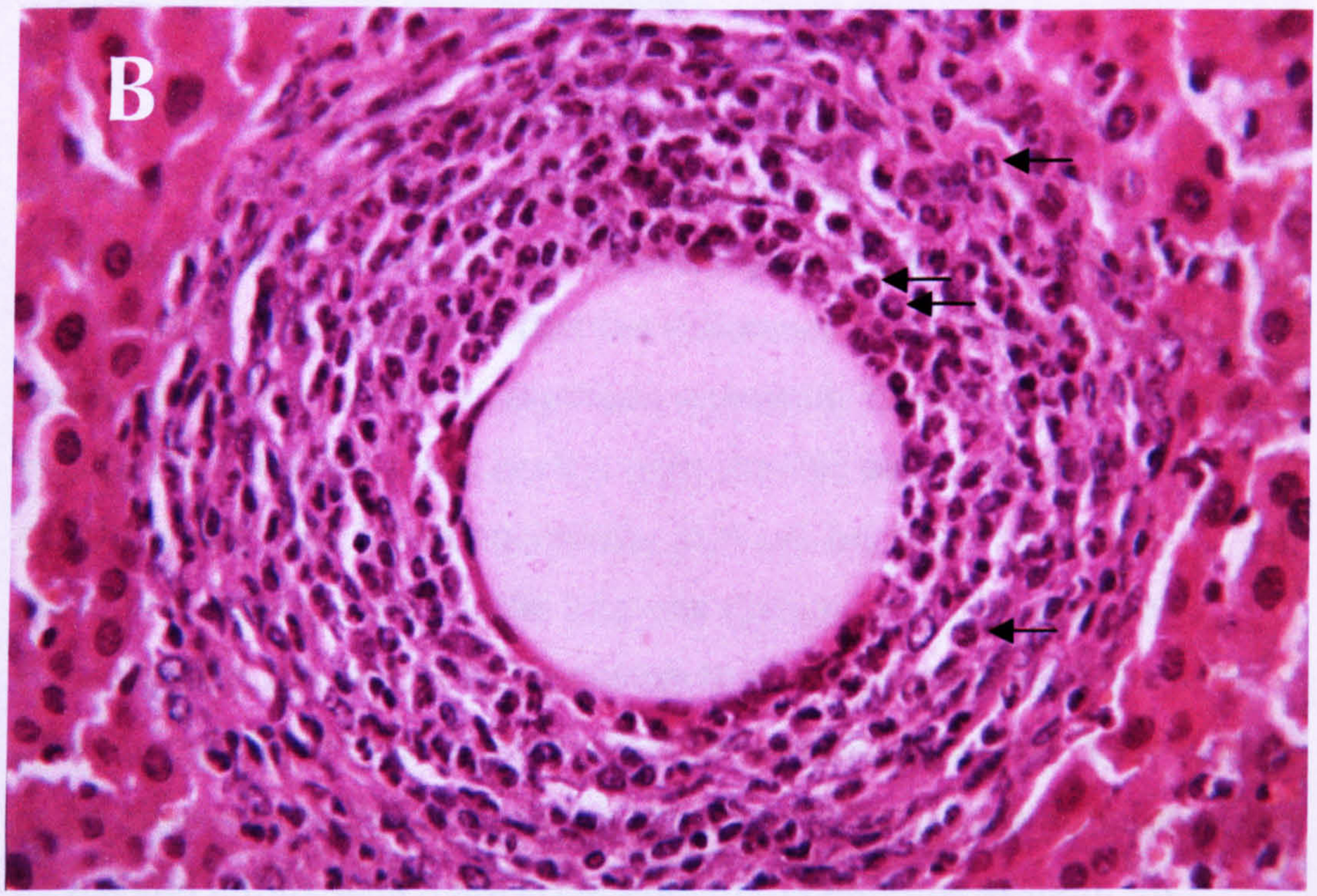


Fig 3-4: Liver sections containing Sepharose beads coupled with (A) SEA and (B) ESP, eight days after injection, stained with haematoxylin and eosin.

Discussion

When eggs are deposited by female worms in the mesenteric vessels, they suffer one of two fates. They either attach to the endothelium, migrate through the tissues and escape from the host or they are swept away, lodge in the liver and induce granuloma formation. Both of these processes have a basis in the immune response to proteins which originate in the egg and in order to understand them, it is necessary to identify these proteins. Whilst most efforts have been directed towards examining the immunogenic components of SEA, in chapter 2 it was shown that the eggs secrete a simple mixture of proteins, designated ESP, which must be regarded as prime candidates for the key elements in both processes.

Examining the antibody responses to SEA and ESP after intraperitoneal injection of viable, mature eggs reveals that the responses are distinct, and that ESP components are recognised well before SEA. It is suggested that the difference in time (approximately 10 days) between equivalent responses represents the life-time of the egg within the intraperitoneal cavity, with the ESP being secreted continuously after injection and the SEA not being released, and therefore not accessible to the immune system, until the eggs die. Additionally, the lack of IgG response to either preparation before the onset of egg deposition in a natural infection indicates that there is little or no cross-reactivity between the egg proteins and those proteins of the larvae and adults which are accessible to the immune system. In looking for the proteins involved in egg escape, it is clear that ESP must be more important than SEA, since it is the live egg which must be released from the host in order for the life cycle to continue.

Similarly, when looking for the factors that initiate granuloma formation, it is the proteins which are released from the live egg which must be examined. This is demonstrated by the formation of

granulomas around ESP- and SEA-coupled beads. In both cases, the lesions resemble those observed around live eggs during an infection. However, the differences in cellular composition are revealing. According to von Lichtenberg [1988], the cells comprising granulomas are eosinophils, mononuclear phagocytes, fibroblasts, lymphocytes, neutrophils, plasma cells and sporadic mast cells (in decreasing order of their proportion in the granuloma). It can be seen from Figs 3-3 and 3-4, that only the granulomas around the ESP-coupled beads contain significant numbers of granulocytes (arrows; probably eosinophils), with the SEA granulomas containing predominantly mononuclear cells. This suggests that ESP mimics the proteins involved in granuloma formation more closely than does SEA. Also worthy of note is the fact that that the fibrosis typical of egg granulomas is more pronounced with SEA than with ESP (marked by stars in fig 3.4), and may indicate that the fibrogenic response occurs only after the death of the egg and the release of SEA.

Whilst the results of these experiments are suggestive that the egg secretions, rather than components of SEA, should be characterised in order to understand the processes of egg escape and granuloma formation, further work could be undertaken to confirm the differences observed. For example, it would be possible to quantify the degree of fibrosis present in the granulomas around the different beads using different histochemical stains, such as Masson's trichrome, or Picro-sirius Red [Waldrop & Puchtler, 1982]. It might also prove worthwhile to examine the profiles of cytokine production elicited in the two different cases, e.g., by using the Cincinnati Cytokine Capture Assay [Finkelman & Morris, 1999] to measure the levels of interleukin-4 and interferon-γ after challenge with the beads, to determine whether the ESP and SEA induced different T-helper cell responses.

Chapter 3: From Proteome to Genome: How to Identify Parasite Proteins

P.D. Ashton, R.S. Curwen and R.A. Wilson (2001) From Proteome to Genome: How to Identify Parasite Proteins. *Trends in Parasitology* **17**(4): 198-202.

The text of this chapter has been reformatted to fit within the guidelines for thesis submission, but remains unchanged from the published article.

Introduction

The following chapter contains a review article describing the basic techniques that are referred to as proteomics, concentrating on those that are involved in the identification of protein spots from 2-D gels. The purpose of this chapter is to illustrate that these techniques were established in the laboratory by me during the course of this project, although most of the examples used were taken from the work of Rachel Curwen. Since this article was written there have been a number of developments of the various technologies, both in the process of selecting spots from a 2-D gel to go forward into the identification process and in the mass spectrometric methods used for this identification.

The only selection criterion used in this project was the presence of a spot on a 2-D gel of ESP, but in more typical experiments, it is differences between preparations that are important. Whilst 2-D gels are now highly reproducible, minute variations in gel composition or in running conditions mean that these differences can only be accurately assessed using sophisticated image analysis software. These usually work by identifying spots on images of two different gels, attempt to quantify the amount of staining in the spots and then match (with user intervention) corresponding spots between the gels. Two different systems have recently been developed that simplify this process. Firstly, a novel program called Z3 (Compugen, Tel Aviv, Israel) has been released which performs the matching of the images prior to spot detection and quantitation. This greatly simplifies the analysis process and enables comparisons to be performed in a matter of minutes, rather than the hours required by other systems. Secondly, a technique called difference gel electrophoresis (DIGE) has been developed, where two protein samples are labelled with different fluorophores (derivatives of Cy3 and Cy5), and are then run on the same gel (Unlu, Morgan & Minden, 1997). The samples are then visualised separately, by using different excitation frequencies. Since the samples were run on the

same gel (and therefore necessarily under identical conditions), spots corresponding to the same protein will occupy identical positions on the gel, and this again simplifies the analysis procedure.

The main advances in mass spectrometry relate to the sequencing of peptides from the tryptic digest of gel spots. Most of the work involved has been attempting to simplify the procedures used. Electrospray mass spectrometers now routinely employ an on-line capillary LC system to separate the peptides in the mixture before they are sprayed into the machine, rather than spraying the whole mixture from a capillary needle. This has effectively increased the sensitivity of these machines by concentrating the individual peptides as they enter the MS, and since the peptides are separated, this also makes it less likely that less abundant peptides are hidden by more dominant ones. However, these systems are still complex to operate, and it would be generally agreed that a more important advance would be the development of MALDI-based peptide sequencing techniques. One way of doing this is to use "charge-site directed fragmentation", a method by which the peptides are derivatised in such a way that the fragmentation spectrum seen during post-source decay experiments is simplified (e.g., Keough *et al.*, 1999). Whilst this is promising, the chemistry involved requires anhydrous conditions, and is therefore beyond the abilities of most biological laboratories. One more promising method has involved the development of a new type of mass spectrometer (a MALDI-ToF-ToF system; Medzihradszky *et al.*, 2000) which allows fragments to be generated by collision, rather than by post-source decay. This again simplifies the fragment ion spectra, and appears to hold the most promise for a fully automatable, high-throughput system for peptide sequencing.

Summary:

Parasite genome projects are generating a veritable avalanche of sequence data. If this resource it to be effectively exploited for drug and vaccine design, there is an urgent need to make the link between these DNA sequences and the functional proteins of the parasite, which they encode. Here, Peter Ashton, Rachel Curwen and Alan Wilson seek to demystify the revolutionary advances in protein identification based on mass spectrometry.

In spite of decades of intensive research on vaccines and chemotherapeutics, infectious agents remain a major cause of morbidity and mortality in humans and domestic livestock. The current hope is that genome initiatives, established to provide complete genetic blueprints for numerous pathogens, will yield new drug targets and vaccine candidates to aid disease control. Database mining reveals essential, conserved genes which may act as drug targets across several species[1], while subtractive analysis aims to highlight pathogen-specific sequences such as virulence genes[2]. Vaccine candidates are likely to prove harder to pinpoint in the genome, since knowing the function of a pathogen protein in no way guarantees its protective potential.

Genomes, Transcriptomes and Proteomes

Sequencing of microbial genomes has been rapid because of their relatively small size, and more than 20 have been completed[3, 4], whereas progress with parasites lags somewhat behind due to their large genome size. Nevertheless, sequencing of protozoan genomes[5] is advanced, with several chromosomes of *Plasmodium* and *Leishmania* completed and others under way[6], while further protozoan genomes are in the pipeline. For parasitic helminths, even larger genomes and the amount of repetitive DNA are obstacles precluding comprehensive sequencing with present resources[7]. For these reasons the approach taken with helminths has been to identify genes via expressed sequence tags[8] (ESTs). There are currently (in July 2000) 22,121 ESTs for *Brugia*

malayi and 12,675 ESTs for *Schistosoma mansoni* on dbEST (www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html).

Given that DNA sequences alone tell us little about the dynamic processes within the parasite, the task facing parasitologists is to extract useful information from the rapidly accumulating genome data. Although definition of the total mRNA transcription pattern (the so-called transcriptome), using DNA array technology, provides information about differential gene expression, there is often a poor correlation with the levels of translated proteins[9, 10]. This is a particularly important consideration with parasites, where proteins may be synthesized hours to days before they are required, e.g. for cell invasion or tissue penetration, and stored in intracellular organelles or glands. Indeed, as the desired drug targets and vaccine candidates are to be found among the translation products, it is the parasite proteins that must be characterized and linked to the genome data, if we are to reach our goal.

The term "proteome" was coined in 1995 to describe the total protein complement expressed by a genome[11, 12] and has since been extended to include the proteins expressed within a cell type, tissue or developmental stage. The characterization of proteomes has been made possible by recent advances in two core technologies, two-dimensional gel electrophoresis and mass spectrometry.

Tools of the trade

Two-dimensional electrophoresis.

The two-dimensional electrophoretic separation of proteins, on the basis of charge in the first dimension and mass in the second, was first described by O'Farrell [13]. As a technique, it lacked reproducibility, largely due to the problem with cathodic drift of carrier ampholytes that led to degradation of the pH gradient at the basic end. This was solved by the development of immobilised pH gradients[14] but the reagents and

equipment have only recently become widely available. This and other refinements mean that two-dimensional fractionations of protein mixtures are now highly reproducible, and provide incredibly detailed maps of the hundreds to thousands of proteins in an extract (Fig 1). Such mapping requires image analysis of the two-dimensional gel using dedicated software (e.g. Phoretix 2D; PDQuest; Melanie 3). Some researchers advocate a comprehensive high-throughput approach to identify all spots on a two-dimensional map[12, 15]. However, a more focused approach can be taken by comparing maps to select spots for identification, based on a wide range of criteria such as: abundance (Fig 1); induced or developmental changes in expression; modifications e.g. phosphorylation or glycosylation; presence in a particular fraction e.g. membranes or secretions; biosynthetic labelling.

In spite of the capacity of two-dimensional electrophoresis to resolve highly complex mixtures of proteins, it does presently have some limitations[16]. In particular, proteins that are hydrophobic, have extreme pI (<3 or >10) or are relatively insoluble are difficult to separate on immobilised pH gradients. Low abundance proteins may also present problems of detection that can only be circumvented by fractionation or enrichment procedures.

Mass Spectrometry (MS).

Until recently, it has been possible to identify proteins only by their primary amino acid sequence or by their antigenic properties. Edman sequencing[17] has been used for the amino acid sequencing (provided that N-terminal modification do not block the degradation reaction), and specific antibodies for the latter (provided they are available). The methods of choice are now based on MS. They are applicable to all proteins, combining the advantages of greater sensitivity and high throughput without the expensive running costs of the earlier technologies.

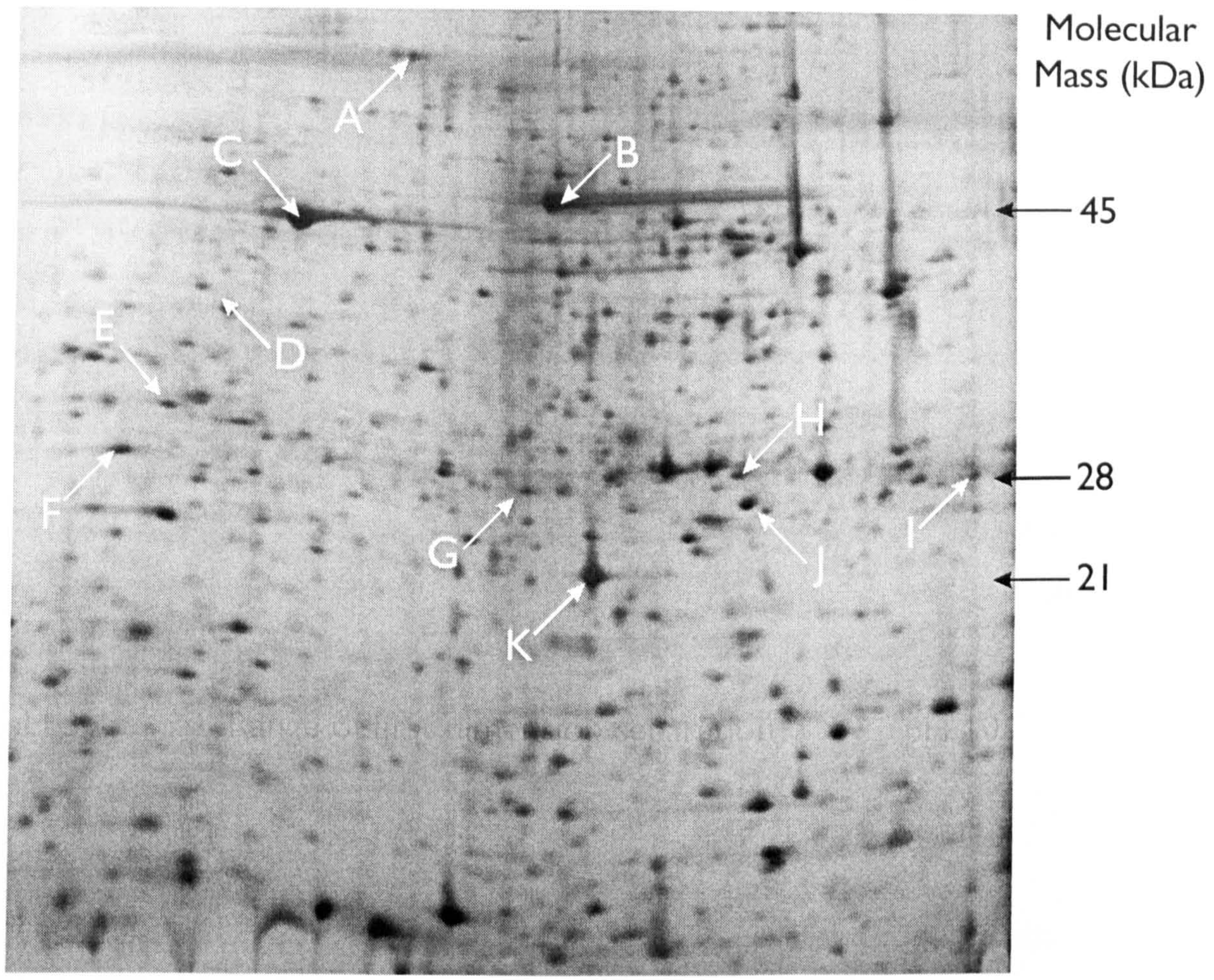


Fig. 1. Soluble *Schistosoma mansoni* cercarial proteins (100 μg) separated by two-dimensional electrophoresis and silver stained. First dimension isoelectric focusing was performed on an IPGPhor (Amersham) with a pH 3-10 linear Immobiline DryStrip (Amersham). Second dimension separation was on a 9-16% gradient SDS-PAGE gel using a Protean II xi gel tank (BioRad). Abundant spots, already identified by MS, are labelled as follows: A – heat shock protein 70; B – Enolase; C – Actin; D – Tropomyosin; E – 14-3-3 ε Homologue 1; F – 14-3-3 Homologue 1; G – GST28; H – TPI; I –Adenylate Kinase; J – GST26; K – Sm21.

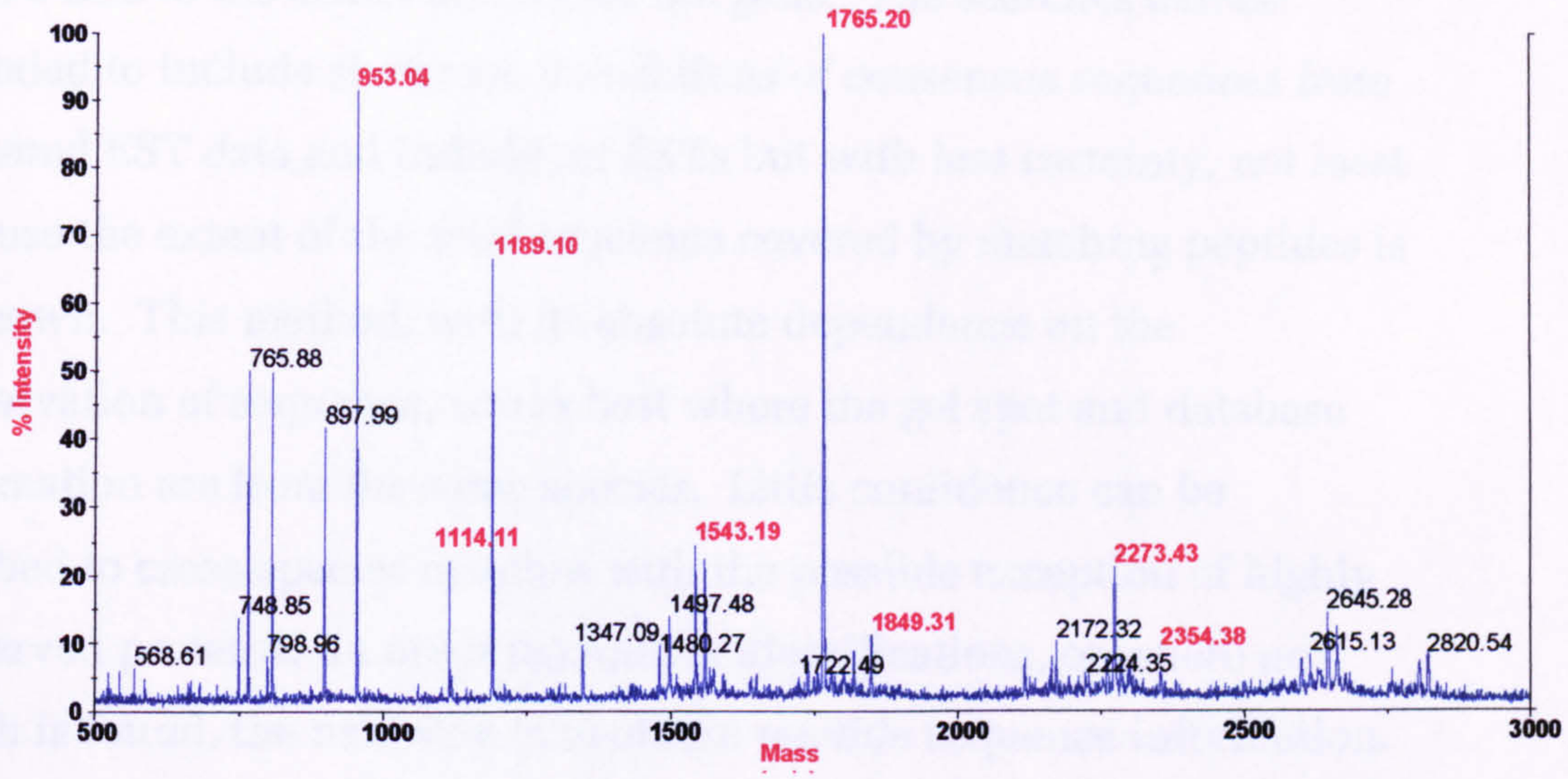
MS is a means of accurately determining the mass of molecules. They are ionised in the gas phase, accelerated by an electric field in a vacuum, and enter a mass analyser, which allows the measurement of their charge to mass ratio (m/z). Traditionally, this involved the deflection of the ion beam by a magnetic field prior to detection, the extent of deflection being directly proportional to charge and inversely to mass. Modern instruments use other parameters such as Time of Flight (ToF) of molecules along a field-free path. Macromolecules in general, and proteins in particular, have traditionally proved difficult to get into the gas phase, but two techniques have been developed to allow MS of a protein, or a mixture of peptides derived from it. These are matrix-assisted laser desorption/ionisation (MALDI)[18, 19] (Box 2) and electrospray ionisation[20, 21] (Box 3).

Application of MS to parasite proteins.

The ability to obtain the precise mass of a given parasite protein is not in itself very useful for identification purposes. In order to extract the maximum information the protein must be converted into a mixture of peptides. Normally this is achieved by excising a spot from a two-dimensional gel (after visualization by a staining procedure that does not permanently modify the protein) and digesting it with trypsin, which cleaves the peptide backbone at the carboxyl side of lysine or arginine residues. The masses of the resulting peptides depend on the primary sequence and thus are diagnostic of a given protein.

The masses of the peptides in the mixture are accurately measured using MALDI-ToF MS (Box 2) to yield a peptide mass fingerprint (PMF; Fig 2). The first step in protein identification is to use that PMF to search full-length protein sequences. These are theoretically digested and ranked according to how many peptide masses they share with the PMF. A perfect match is unlikely (due to poor desorption and ionisation of certain peptides[22]) and unnecessary; as few as three[23] or 4-6 (Ref. 24) matches are claimed to be sufficient to

(a)



(b)

```
  1 MSILTIHARQ IFDSRGNPTV EVDLKTSKGL FRAAVPSGAS TGVHEALELR
 51 DTNSKAYMKK GVLTAVSNVN KIIAPALINK NIPVTNQAAI DKYMIDLDGT
101 ENKEKLGANA ILGVSLAVCK AGAAEAGLPL YRYIARLAGH EDVIMPVPAF
151 NVINGGSHAG NKLAMQEFMI LPTGASSFTE AMQIGTEVYH NLKAVIKREY
201 GLDACNVGDE GGFAPNIQDN MKGLQLLEEA IKIAGYTGKV EIGMDCAASE
251 FHKNGKYDLD FKNPHSAEST WLSPDAMANM YKQMISKFPI VSIEDPFDQD
301 DWETWPKLTS STNIQIVGDD LTVTNPKRIK QAIASKACNC LLLKVNQIGS
351 LTESIEACKL AQDSGWGVMV SHRSGETEDT FIADLVVGLC TGQIKTGAPC
401 RSDRLAKYNQ LLRIEEELGT AAKYAGKNFR HPKV
```

Fig. 2. Peptide mass fingerprint (PMF) of a protein spot excised from a two-dimensional gel, digested with trypsin and analysed on a MALDI-ToF/MS (Applied Biosystems Voyager DE-STR) (a). Searching the SwissProt database with MASCOT using this PMF identified the protein as *S. mansoni* enolase (ENO_SCHMA). Prominent matched peptides are highlighted in red. The full sequence of enolase with the matching peptide highlighted in red; vertical bars show the potential tryptic cleavage sites (b).

make a link to the cDNA and hence the gene. The searches can be extended to include six-frame translations of consensus sequences from clustered EST data and individual ESTs but with less certainty, not least because the extent of the total sequence covered by matching peptides is unknown. This method, with its absolute dependence on the conservation of sequence, works best where the gel spot and database information are from the same species. Little confidence can be ascribed to cross-species matches with the possible exception of highly conserved proteins. In order to confirm identifications, or where no match is found, the next step is to obtain peptide sequence information.

Peptide sequence tags can be obtained from the same peptide mixture by subjecting it to tandem MS (Box 3) and selecting several peptides for detailed analysis. Each is fragmented in the collision cell and the resulting masses analysed to provide one tag (Fig 3). Thus, it is possible to obtain several sequence tags for a given gel spot, but a single tag should be adequate for unequivocal identification (unless it is found within a conserved domain present in several distinct proteins). Unlike PMFs, the sequence tags are just as applicable to database searching with consensus sequences and ESTs as to full-length sequences. They can also be matched to homologous proteins in other species.

Search Tools

The above analyses are made possible by access to parasite sequence databases, and a range of bioinformatics search tools (see Box 1). The PMF search tools all permit a variety of protein sequence databases to be interrogated using a range of options such as the digesting protease, number of missed cleavages, protein modifications, e.g. oxidation of methionine residues, and mass tolerance.

Searching with peptide sequence tags can be done easily, eg., with the standard BLAST tools on the NCBI web site. Unfortunately, ambiguities in the peptide sequence can only be entered for BLAST searching as unknowns but ProteinInfo, for example, will allow

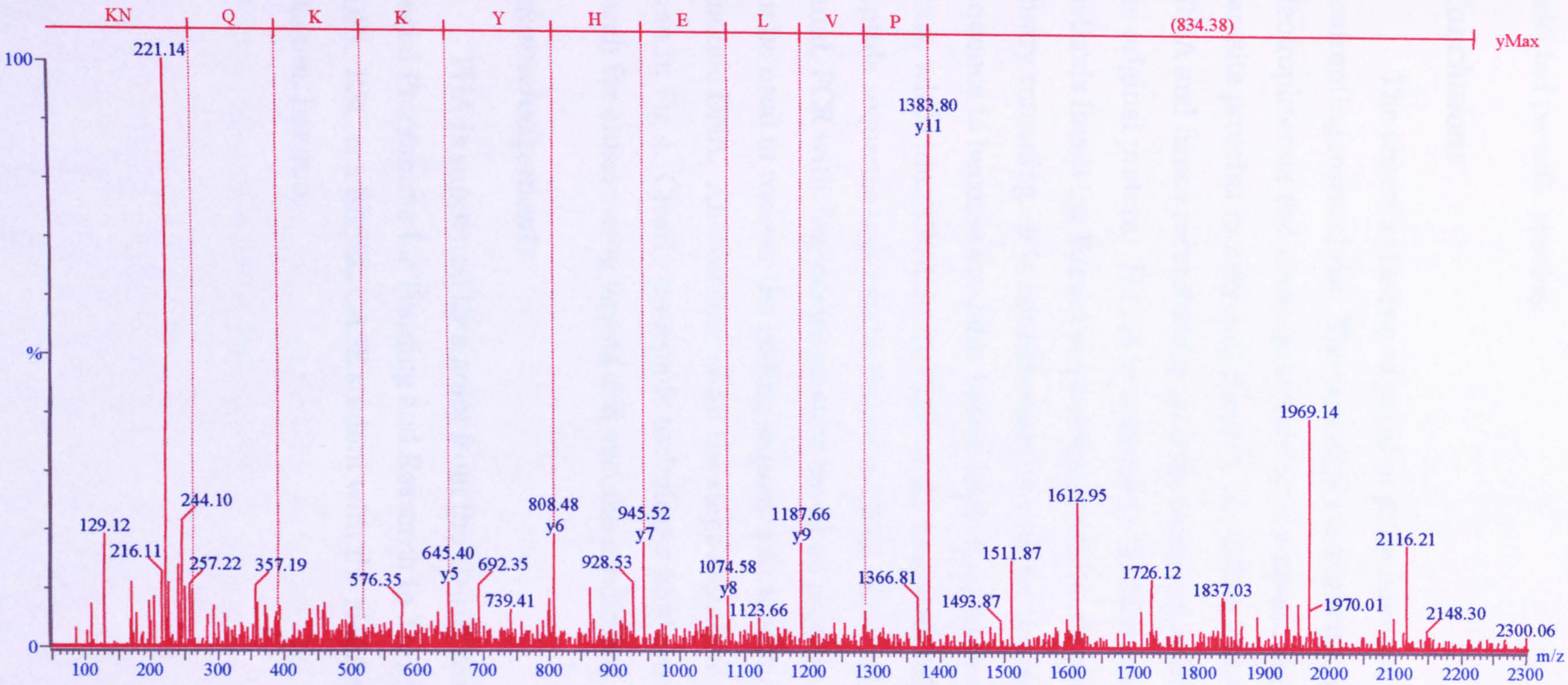


Fig. 3. Tandem mass spectrum of the ions produced by fragmentation of a peptide with a mass of 2220Da. The spectrum was deconvoluted using MaxEnt3 software (Micromass) so that all the charge to mass ratios can be taken to be the true masses of the fragments. Interpretation of the spectrum leads to a partial sequence of PV[IL]EHY[KQ][KQ][KQ]NK. Searching ProteinInfo with this sequence identifies the parent protein as *S. mansoni* adenylate kinase.

alternatives to be specified for a given residue. The parasite genome web site at EBI is of particular value for parasitologists, as it provides the capability to search with ambiguous sequences against databases for selected parasite species.

Conclusions

The characterization of parasite proteomes is a major task confronting researchers. The powerful combination of two-dimensional electrophoresis and mass spectrometry is a means to identify selected parasite proteins rapidly and cheaply, providing a direct route to the cDNA and hence recombinant protein, requiring as little as 500 fmol of the original protein. This is in contrast to the laborious and protracted methods involving Edman sequencing, or antibody production and library screening. It is not necessary to wait for the complete genome sequence to become available before beginning proteome analysis. Thus, while identification by PMF works best with full-length cDNAs, peptide sequence tags can be linked to ESTs. Even when no match is found, PCR with degenerate primers based on peptide sequence tag data can be used to recover the coding sequence from cDNA libraries or genomic DNA. An overview of all the steps described in this article is given in Fig 4. Clearly proteomic techniques hold great promise in the search for elusive drug targets and vaccine candidates.

Acknowledgements

PDA is supported by a grant from the UNDP/World Bank/WHO Special Programme for Training and Research in Tropical Diseases (TDR). RSC is a BBSRC CASE student with the Natural History Museum, London.

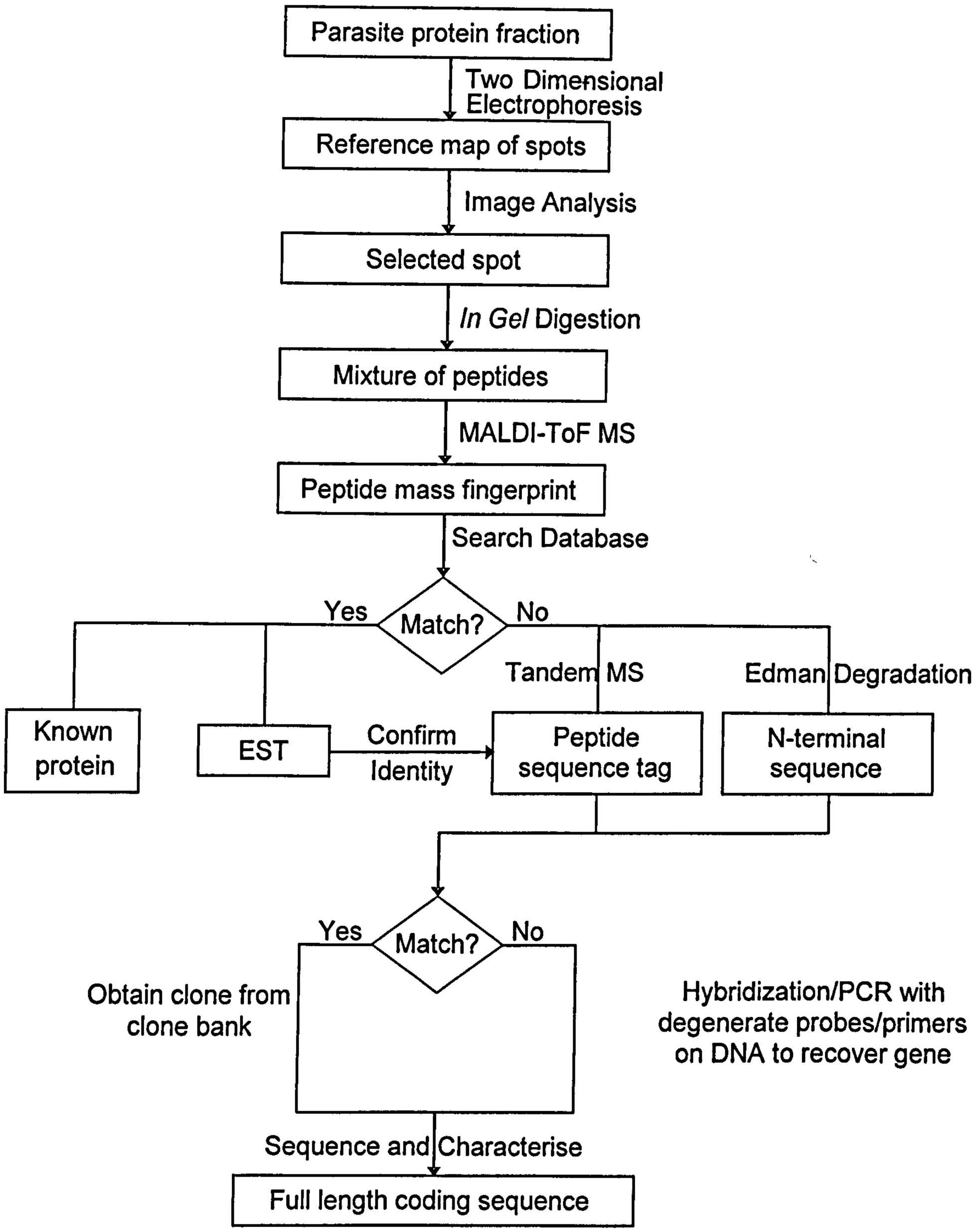


Fig. 4. Flowchart of the steps involved in the identification of parasite proteins using proteomics techniques.

Box 1. Proteomics search tools

For peptide mass fingerprints:

- PeptIdent: http://www.expasy.ch/tools/peptident.html
- ProFound: http://prowl.rockefeller.edu/cgi-bin/ProFound
- MASCOT: http://www.matrixscience.com/
- MS-Fit: http://prospector.ucsf.edu/ucsfhtml3.2/msfit.htm

For peptide sequence tags:

- BLAST: http://www.ncbi.nlm.nih.gov/BLAST/
- ProteinInfo: http://prowl1.rockefeller.edu/prowl/proteininfo.html
- Parasite Protein Motif Search: http://www.ebi.ac.uk/cgi-bin/parasites/motif_search.p

Parasite EST databases:

- http://www.cbil.upenn.edu/ParaDBs/
- http://www.ebi.ac.uk/parasites/paratable.html
- http://genome.wustl.edu/est/toxo_esthmpg.html

Box 2. Matrix-assisted laser desorption and ionization-time of flight mass spectometry for peptide mass fingerprints

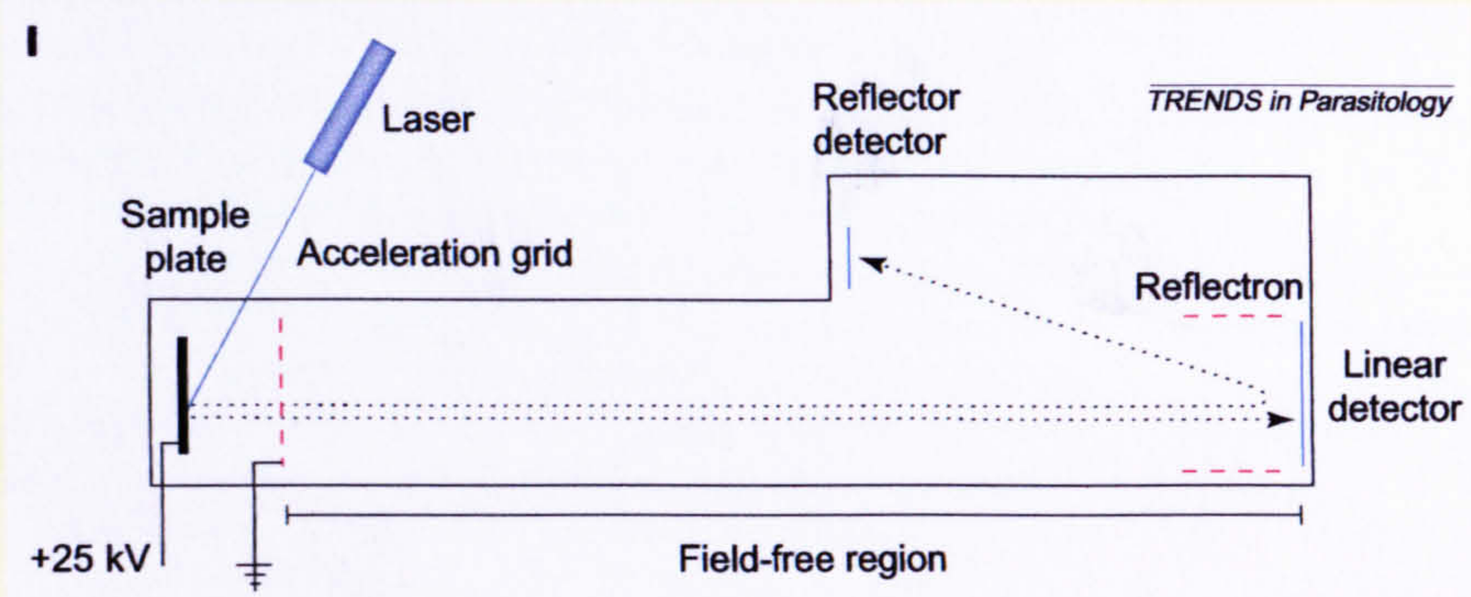
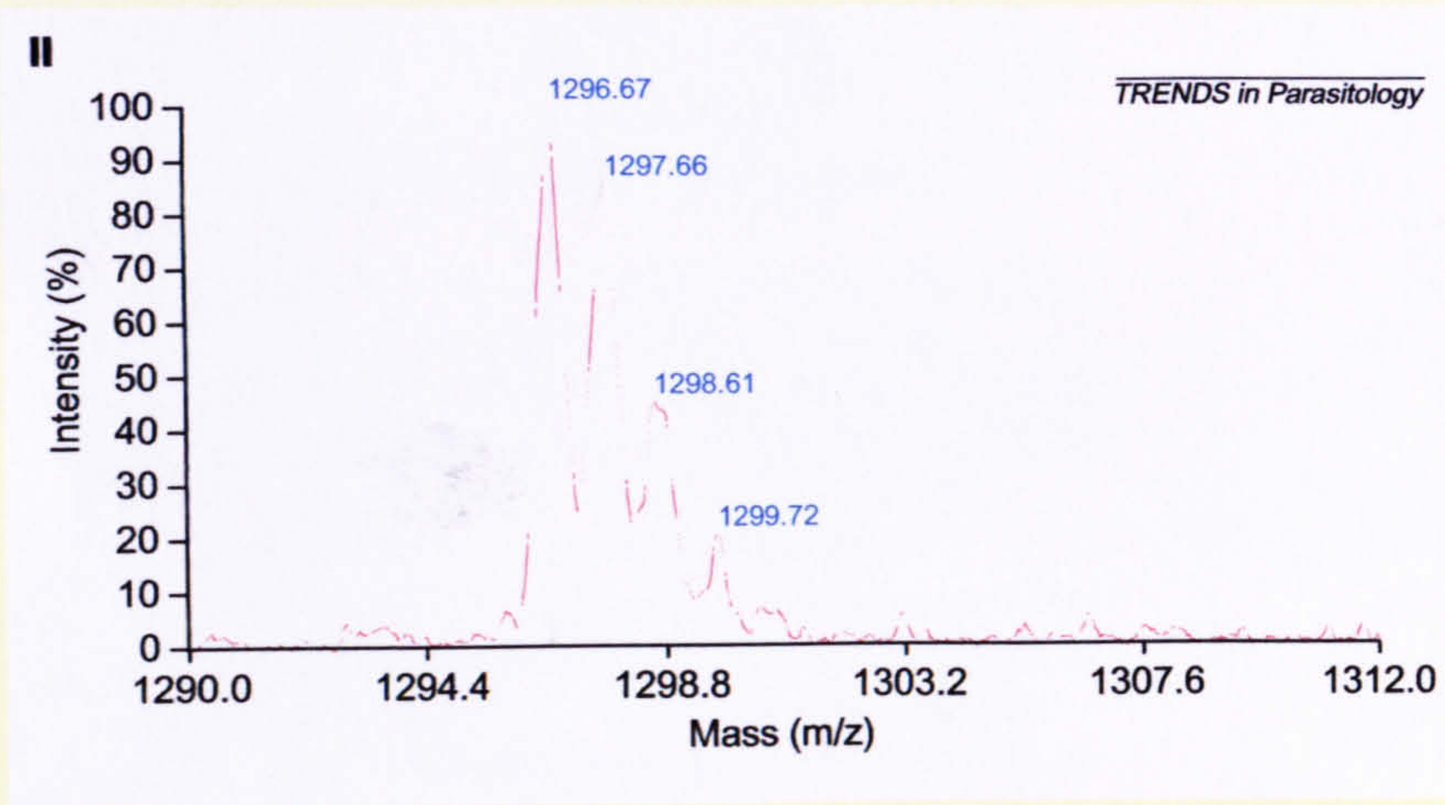


Fig. I. Matrix-assisted laser desorption and ionization-time of flight (ToF) mass spectrometer.



The starting point is a digest of peptides derived from a single gel spot. This is cocrystallized on to a sample plate with a UV-absorbing matrix. The transfer of laser energy from the matrix to the peptides causes their desorption into the vacuum where they are accelerated through an electric field (Fig. I). At this point, all the peptides have the same kinetic energy ($\frac{1}{2}mv^2$, where m is the mass and v is the velocity) so smaller peptides will fly faster than larger ones. By measuring the time of flight (ToF) to the detector at the far end of the field-free region of the flight tube, the mass to charge ratio of each ionic species can be determined with an accuracy of better than ±0.5 Da, and with a resolution such that the ^{13}C content of the peptide is apparent (Fig. II). This accuracy can be increased further, at the expense of sensitivity, by extending the time between ionization and acceleration (delayed extraction), and by positioning an ion mirror (reflectron) to extend the flight path and focus the ions. Fortunately, the laser ionization technique predominantly produces singly protonated ions so that the mass to charge ratio is the true mass +1 Da. A calibrated mass spectrum (see Fig. 2 in the text) and a list of masses can be output, together representing the peptide mass fingerprint, for which an identity can be sought.

Fig. II. (left) The ratio of ^{12}C to ^{13}C is 99 : 1 so that some of the molecules of a given peptide will contain one or more ^{13}C atoms, yielding a series of peaks, which differ in mass by 1 Da.

Box 3. Electrospray tandem mass spectrometry for peptide sequence tags

For this technique, a digest of peptides derived from a single spot is dissolved in a volatile solvent and microlitre volumes are sprayed into the instrument through a fine needle over a period of 20–30 minutes. As the name implies, tandem mass spectrometers consist of two consecutive analysers; in the case of the Q-ToF (Micromass), a quadropole (MS1) followed by a time of flight (ToF) analyser (MS2) (Fig. I). Once in the vacuum of the instrument, the solvent rapidly evaporates to give a constant stream of peptide ions.

An initial spectrum of the peptide ion masses is collected using only the ToF analyser. Inspection of the carbon isotope pattern for each peak reveals its charge state (Fig. II). Multiply charged ions will probably be peptides, and these are selected using the quadropole mass analyser for further analysis. The ion beam enters the quadropole, which is tuned to allow the transit of only those ions with a selected mass to charge ratio. These enter a cell where they are fragmented by collision with an inert gas, usually argon. Cleavage can occur at any bond within the peptide molecule but, under conventional conditions, is predominantly restricted to the peptide backbone (Fig. III); y- and b-ions are the most common products. The fragments are accelerated orthogonally by applying a pulsed electric field to the pusher plate and their masses determined in the ToF analyser, exactly as in the later stages of matrix-assisted laser desorption and ionization (MALDI)-ToF mass spectroscopy, but with an accuracy of better than 0.05 Da.

Unlike the MALDI instrument, the fragments can be multiply protonated and the resulting complex spectrum must be deconvoluted to produce a theoretical mass spectrum of the ions as though they were all singly charged. To obtain the amino acid sequence, the trick is to select the series of y-ions that differ consecutively by the exact mass of one of the 20 naturally occurring amino acid residues (see Fig. 3 in the text). It should be noted that Leu (L) and Ile (I) have identical mass, whereas Gln (Q) and Lys (K) differ by 0.04 Da and cannot be distinguished with complete confidence. The amino acid sequence of the peptide is then read off from the intervals between the peaks of the y-ion series to yield the sequence tag.

I

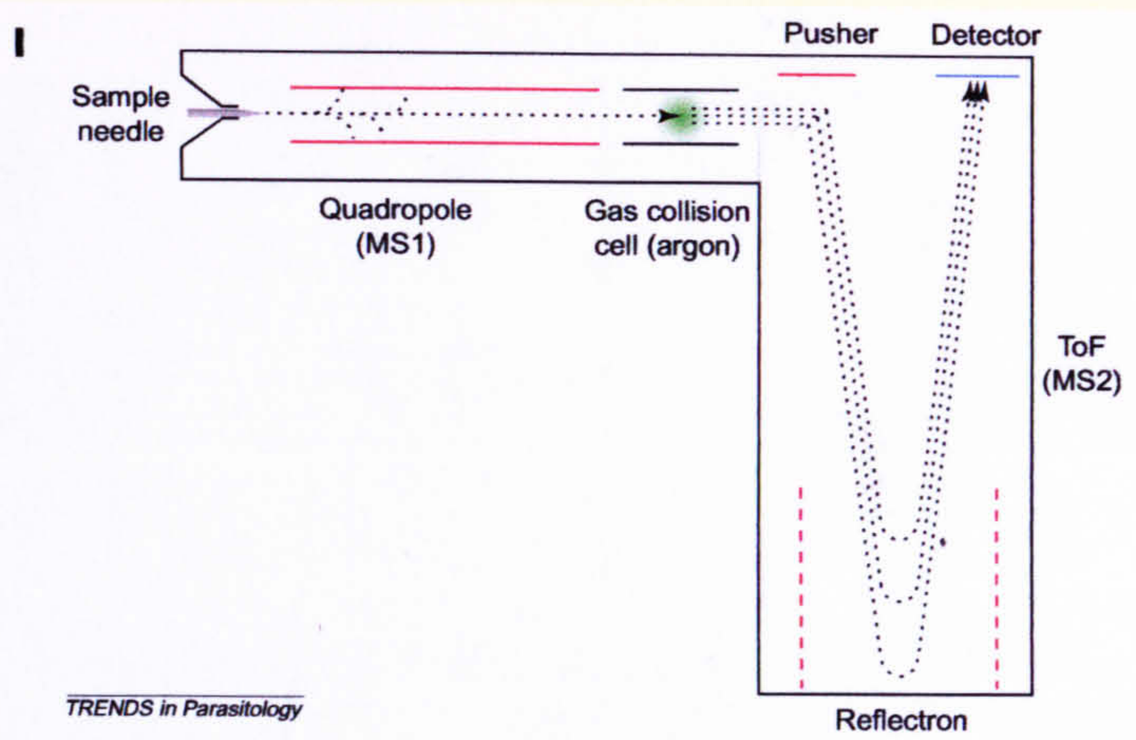


Fig. I. Electrospray tandem mass spectrometer

II

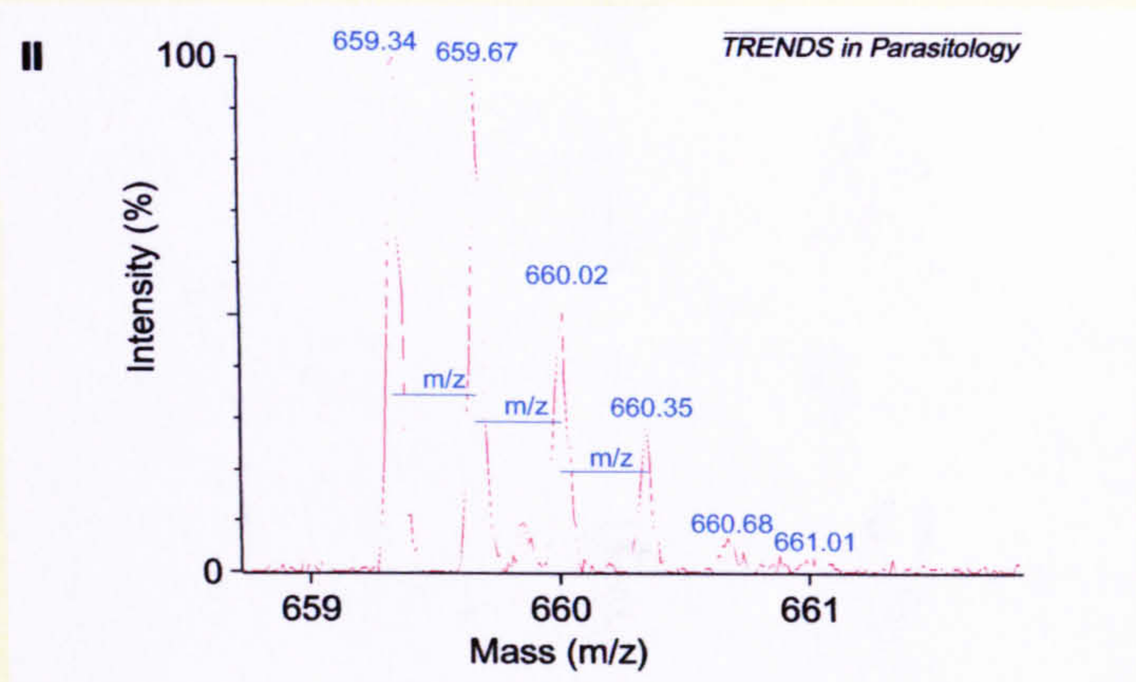


Fig. II. The series of peaks from a given peptide differ in mass to charge ratio (Δm/z) by 0.33 in this example. (Mass difference is 1 Da, charge = 1/0.33 = 3.)

III

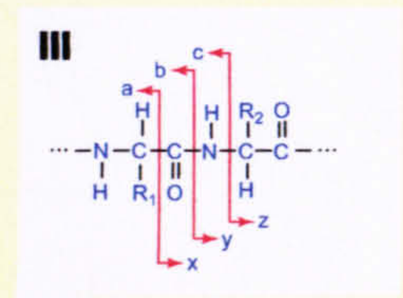


Fig. III. Potential cleavage points in the peptide backbone.

References

1 Galperin, M.Y. and Koonin, E.V. (1999) Searching for drug targets in microbial genomes. *Curr. Opin. Biotechnol.* 10, 571-578

2 Huynen, M. *et al.* (1998) Differential genome analysis applied to the species-specific features of *Helicobacter pylori*. *FEBS Lett.* 426, 1-5

3 Tomb, J.F. *et al.* (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388, 539-547

4 Cole, S.T. *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-544

5 Johnston, D.A. *et al.* (1999) Genomics and the biology of parasites. *BioEssays* 21, 131-147

6 Blackwell, J.M. and Melville, S.E. (1999) Status of protozoan genome analysis: trypanosomatids. *Parasitology* 118, S11-S14

7 Williams, S.A. *et al.* (1999) Helminth genome analysis: the current status of the filarial and schistosome genome projects. *Parasitology* 118, S19-S38

8 Franco, G.R. *et al.* (1995) Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene* 152, 141-147

9 Anderson, L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundancies in human liver. *Electrophoresis* 18, 533-537

10 Gygi, S.P. *et al.* (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720-1730

11 Wasinger, V.C. *et al.* (1995) Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 16, 1090-1094

12 Wilkins, M.R. *et al.* (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* 13, 19-50

13 O'Farrell, P.H. (1975) High-resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250, 4007-4021

14 Bjellqvist, B. *et al.* (1982) Isoelectric focussing in immobilised pH gradients: principle, methodology and some applications. *J. Biochem. Biophys. Methods* 6, 317-339

15 Shevchenko, A. *et al.* (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. U. S. A.* 93, 14440-14445

16 Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. and Aebersold, R. (2000) Evaluation of Two-Dimensional Gel Electrophoresis-Based Proteome Analysis Technology. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9390-9395.

17 Edman, P. and Begg, G. (1967) A Protein Sequenator. *Eur. J. Biochem.* 1, 80-91

18 Karas, M. and Hillenkamp, F. (1988) Laser Desorption Ionization of Proteins With Molecular Masses Exceeding 10000 Daltons. *Anal. Chem.* 60, 2301-2303

19 Hillenkamp, F. *et al.* (1991) Matrix-Assisted Laser Desorption Ionization Mass-Spectrometry of Biopolymers. *Anal. Chem.* 63, A1193-A1202

20 Fenn, J.B. *et al.* (1989) Electrospray Ionization For Mass-Spectrometry of Large Biomolecules. *Science* 246, 64-71

21 Smith, R.D. *et al.* (1990) New Developments in Biochemical Mass-Spectrometry - Electrospray Ionization. *Anal. Chem.* 62, 882-899

22 Krause, E. *et al.* (1999) The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Anal. Chem.* 71, 4160-4165

23 Pappin, D.J.C. *et al.* (1993) Rapid Identification of Proteins by Peptide Mass Fingerprinting. *Curr. Biol.* 3, 327-332

24 Mann, M. *et al.* (1993) Use of Mass Spectrometric Molecular Weight Information to Identify Proteins in Sequence Databases. *Biol. Mass Spec.* 22, 338-345

Chapter 4: Electrophoretic and MALDI-MS Analysis of Egg Secreted Proteins

Introduction

The SDS-PAGE gel of ESP shown in Chapter 2 demonstrated a simple pattern of six bands, but it is possible that there were more proteins present in the mixture. This is because SDS-PAGE separates proteins solely based on molecular weight, and so proteins with the same or similar masses will migrate as a single band. It was therefore decided to use the greater resolving power of 2-D electrophoresis to determine whether this was the case. Since the proteins are separated on the basis of the distinct properties of isoelectric point in the first dimension, and molecular weight in the second dimension, the chance that a single spot would contain more than one protein is significantly reduced. A 2-D gel therefore forms a better starting point to identify the individual proteins by peptide mass fingerprinting.

Although the advances in 2-DE described in Chapter 1 have greatly increased its utility, there are still significant limitations to the technique and the most significant of these relates to solubilisation. In SDS-PAGE, the SDS (sodium dodecyl sulphate or sodium lauryl sulphate) applies very strong denaturing conditions to the proteins in the mixture. The combination of the long chain hydrocarbon and the strongly polar sulphate group ensure that even the most hydrophobic proteins (e.g., membrane proteins) are fully solubilised after boiling in sample buffer. However, at the same time, the SDS imparts a significant negative charge to the molecules it binds. Since the purpose of IEF is to separate the proteins based on their native charge, this rules out the use of significant concentrations of SDS during IEF.

In general, the ionic concentration of the IEF sample buffer must be kept as low as possible to avoid any endoosmotic flow within the gel, where the water in the gel follows small ions (e.g., salt ions) as they are

attracted to the electrodes. In extreme cases, this can lead to the drying of portions of the gel and the interruption of the electric field that is applied. The only exceptions to this rule are zwitterions, which carry no net charge and are therefore not attracted to either electrode. Indeed, zwitterionic carrier ampholytes are used to maintain the ionic strength of the buffer during the focussing. Although there are readily available non-ionic or zwitterionic detergents (e.g., Triton X-100, NonIdet-P40, CHAPS), none of these has the same solubilising ability as SDS, and although some newer sulpho-betaine derivatives have been successfully used to solubilised some membrane proteins [Santoni *et al.*, 2000], these are not yet widely applicable, or commercially available. The separation of hydrophobic and membrane proteins therefore remains a significant problem.

Once the proteins have been separated, the challenge then is to identify the individual components of the mixture. The first step that is usually taken in proteomics studies is to produce peptide mass fingerprints (PMFs) [Cottrell, 1994]. Here the spots are excised from the gel and digested with a suitable protease (e.g., trypsin), and the masses of the resulting peptides measured, using a MALDI-ToF mass spectrometer. The resolving power of this machine is such that families of peaks are observed for each peptide in the mixture, representing the different amounts of ^{13}C present (see Chapter 3). The only mass that is of interest is the all-^{12}C, or monoisotopic, peak, since this is used to search the available sequence databases. To perform this search, each of the protein sequences in the database is subjected to a theoretical digestion according to the specificity of the protease used to digest the sample proteins (e.g., for trypsin, the program would "cleave" each sequence at every lysine or arginine residue which is not followed by a proline). The theoretical monoisotopic mass of each of these peptides is then calculated by summing the monoisotopic masses of its amino acid residues. The list of theoretical masses can then be compared to the list obtained on the mass spectrometer and a large number of matches

between the database entry and the peptide mass fingerprint provides a good indication of the identity of the target protein.

A number of systems now exist to search in this way and most are freely available on the Internet (e.g., MS-Fit, ProFound, PeptIdent and Mascot). These are all more complicated than the description given above, and allow a wide range of options. Most systems allow the digesting protease to be specified, as well as the maximum number of missed cleavages, to allow for incomplete digestion, or situations where cleavage of a particular residue might not be favourable (e.g., when the basic lysine or arginine residues are followed by an acidic one, such as glutamic acid or aspartic acid). Additionally, a number of options are provided to indicate which post-translational or artefactual modification may have occurred on peptides, e.g., phosphorylation of serine, threonine or tyrosine residues, alkylation of cysteines, or the oxidation of methionines. All the Web sites listed above provide a rapid means of screening peptide mass fingerprints against a range of protein sequence databases (e.g., SwissProt, OWL, trEMBL and the NCBI nr database) and potentially identifying the proteins.

In this chapter, the results of separation of the ESP proteins by 2-D electrophoresis are shown. In addition, assays for proteolytic activity and glycoprotein content are undertaken on the gel-separated proteins. Finally, peptide mass fingerprints were obtained for each of the spots on the 2-D gel. Similarities observed between several of the PMFs suggested these proteins might be related in some way, and this was investigated using maximum parsimony methods.

Materials and Methods

Parasite Material and Sample Preparation

Eggs were collected, cultured and the culture medium concentrated, and SEA prepared as described previously (Chapter 2). For each gel 50μg ESP or 500μg of SEA was diluted using rehydration

buffer containing 7M urea, 2M thiourea, 2% 3-([3-Cholamidopropyl]dimethylammonio)-1-propanesulfonate (CHAPS), 10mM dithioerythritol (DTE) and 2% pH 3-10 carrier ampholytes (Resolytes, BDH).

Two-Dimensional Electrophoresis

All the steps for the isoelectric focussing were performed on an IPGphor system (Amersham). Samples were introduced into 18cm pH 3-10 (linear) immobilised pH gradient (IPG) strips (Amersham) by rehydration [Sanchez, Hochstrasser & Rabilloud, 1999] for 18 hours @ 20°C then the following voltages were applied: a gradient from 300 to 3,500V over 3 hours, 3,500V for 3 hours and finally sufficient time at 5,000V to bring the total Volt-hour product to 100,000. Immediately after completion of the first dimension separation, the strips were taken out of the strip holders, rinsed briefly in deionised water, and incubated in the second dimension equilibration solution (50mM Tris, pH 6.8, 6M urea, 30% glycerol, 2% SDS, trace bromophenol blue) containing 1% DTE for 12 minutes. They were then transferred to the same solution containing 4% iodoacetamide instead of the DTE (to block the reduced cysteines residues) for 5 minutes, before being placed on a 15% SDS-PAGE gel (19cm×20cm) containing PDA as a cross-linker instead of Bis-acrylamide [Hochstrasser, Patchornik & Merril, 1988] and run in a Protean II XL tank (Biorad) at 40mA per gel until the dye front left the gel. The gels were then fixed in 40% methanol/10% acetic acid and stained using colloidal Coomassie Blue G-250 concentrate (Sigma). Gels were scanned on an Expression 1600 Pro scanner (Epson) using the transmission unit, and spots detected with 2-D Advance software (Phoretix).

Ammoniacal Silver Staining

The silver staining protocol used is based on the method used by the SWISS-2D PAGE service in Geneva (available at http://www.expasy.ch/ch2d/protocols/protocols.fm4.html). Briefly, gels were fixed overnight in 40% methanol/10% ethanol, then washed for 5

minutes in water. The gels were sensitised by soaking in 1% glutaraldehyde in 0.5M sodium acetate for 30 minutes and washed three times in water for 10 minutes. They were then soaked twice in a 0.05% solution of 2,7-naphthalene disulphonic acid (Fisher) twice for thirty minutes and washed for times for 15 minutes in water. Ammoniacal silver nitrate was prepared by dissolving 6g of silver nitrate in 30ml of water and adding this slowly to mixture of 160ml of water, 10ml of 25% ammonia and 1.5ml of 10M NaOH and adjusting the final volume to 750ml. This was then added to the gel and incubated for 30 minutes, followed by a further four washes of 4 minutes each. The gels were developed in a solution of 0.01% (w/v) citric acid and 0.1% (v/v) formaldehyde until slight background staining appeared and then development was stopped with a solution of 5% (w/v) Tris and 2% (v/v) acetic acid.

Two-Dimensional Zymography

Samples of 50µg of ESP were run on 7cm, pH 3-10 IPG strips as above, with the omission of DTE from the rehydration buffer, and using a single equilibration step in the absence of both DTE and iodoacetamide, in order to preserve any disulphide bonds in the proteins which may be required for their activity. In addition, the second dimension gel contained 0.1% casein to act as a substrate for any proteases in the mixture [Lockwood *et al.*, 1987]. After the gels were run, they were washed twice for 30 minutes in 1% Triton X-100 to remove SDS and renature the proteins, before being incubated overnight @ 37°C in a pH 7.5 50mM Tris-Cl buffer containing 200mM NaCl, 5mM $CaCl_2$ and 5µM $ZnSO_4$. The gels were then stained using colloidal Coomassie, any proteolytic activity being revealed by the presence of clear patches in the uniformly stained background, where the casein substrate had been digested.

Tryptic Digestion of Gel Spots

All the spots from the 2-D gels of ESP were excised and destained by washing three times in 50% acetonitrile/25mM ammonium bicarbonate for 10 minutes. Once the gel pieces appeared shrunken and white, they were dehydrated in a SpeedVac (Savant) for 30 minutes, reswelled in 25mM ammonium bicarbonate containing 12.5µg/ml trypsin (sequencing grade, Roche) and incubated at 37°C overnight. The peptides were recovered by washing the gel pieces three times in 50% acetonitile/5% formic acid. The washes for each gel pieces were combined along with the supernatant from the overnight digestion, dried in the SpeedVac, washed twice with 100µl of water to remove volatile salts and resuspended in 10µl 0.1% formic acid. Since reduction and alkylation of the proteins is performed between the first and second dimension of the 2-D gel, these reactions were not performed again at this stage.

Glycan Staining of ESP

Mini 2-D gels of ESP (7cm, pH3-10 IPG strips, followed by a pre-cast minigel; Novex) were run as above and stained for glycan content using a fluorescent labelling technique [Eckhardt, Hayes & Goldstein, 1976]. Briefly, after overnight fixation in 40% methanol/10% acetic acid, the gels were treated for 2 hours in 0.7% sodium metaperiodate/5% acetic acid to oxidise the glycan residues. They were then washed four times for 15 minutes in deionised water, and the oxidised glycans labelled by incubation for 2 hours at 60°C in equal volumes of acidified dimethylsulphoxide (DMSO) (0.6ml concentrated hydrochloric acid/litre DMSO) and a freshly prepared solution of dansyl hydrazine (2mg/ml) in DMSO (Sigma). The gels were then reduced for 30 minutes with sodium borohydride (0.2mg/ml in DMSO) to convert the resulting hydrazones to stable hydrazine derivatives, rinsed in deionised water and destained in 1% acetic acid. The staining was visualised on a GelDoc UV transilluminator (FlowGen).

MALDI Analysis

For each of the gel spots, the tryptic peptide mixture (0.5µl) was spotted onto the MALDI target plate and mixed with 0.5µl of a saturated solution of α-cyano-4-hydroxycinnamic acid (CHCA; Sigma) in 50% acetonitrile/0.1% TFA. The plate was then dried under a constant stream of warm air, loaded into the Voyager DE-STR mass spectrometer (Perseptive) and spectra acquired using the Perseptive Voyager 5.0 software. Close external calibration was performed for all spots, using the standard Perseptive peptide calibration mix 2.

Peptide Mass Fingerprint Searching

Prior to database searching the baselines of the spectra were corrected, noise removal performed (SD=2.0) and the peak list deisotoped (so that only the mass of the monoisotopic peptide peak was included in the list), using the Perseptive Data Explorer Software. PMFs were used to search against the NCBI nr dataset using the Matrix Science MASCOT system (www.matrixscience.com). The mass tolerance was set to 0.5Da, and the appropriate options selected to indicate that that all cysteines would be alkylated and that methionines might have been oxidised during electrophoresis.

Similarity Analysis

The lists of masses in the PMFs were converted into a suitable matrix in NEXUS format [Maddison, Swofford & Maddison, 1997] using a set of Microsoft Excel macros (See Appendix 2). This reformatting was performed in two steps. Firstly, the lists of peptide masses in columns of an Excel spreadsheet were spread out so that each row contained similar masses, with a tolerance of 0.5Da (with blank cells in the columns for PMFs which did not contain that mass, and with the masses corresponding to autolytic trypsin peptides removed). This was then converted to NEXUS format, by creating a line in the NEXUS file for each PMF, which contained a 0 for each blank cell in the column and a 1 for each cell containing a peptide mass.

Bootstrap analysis was then performed using PAUP* software [Swofford, 1993]. A peptide mass fingerprint of actin (SEA2) obtained from the 2-D gel of *S. mansoni* soluble egg antigen (SEA) was used as an "outgroup", to enable the resulting tree to be rooted.

Results

Worked Examples of Peptide Mass Fingerprinting

A 2-D gel of SEA was run and is shown in Fig. 4.1. This demonstrates a complex pattern of spots, as would be expected since the preparation contains all of the soluble proteins present in the egg. To establish whether peptide mass fingerprinting would be appropriate for the egg proteins, ten spots were excised from the gel, digested with trypsin and peptide mass fingerprints generated. An example fingerprint from the highlighted spot in Fig. 4.1 is shown (Fig 4.2). Searching using the peptide masses yielded an identification of the proteins as SmP40, a previously characterised egg antigen. The search results from the MASCOT search engine are shown in Table 4.1. Of the top four matches, three were statistically significant ($p<0.05$), although one of these was to an *E. coli* protein. The remaining two significant hits and the fourth, non-significant, hit were to different database entries for the same protein, the major egg antigen p40. The last of these fell below the level of significance presumably because it contained one less matching peptide than the other two.

Of the ten spots taken for peptide mass fingerprinting from the gel of SEA, five yielded identifications (annotated on Fig 4.1), and two of these appear to be isoforms of the p40 protein.

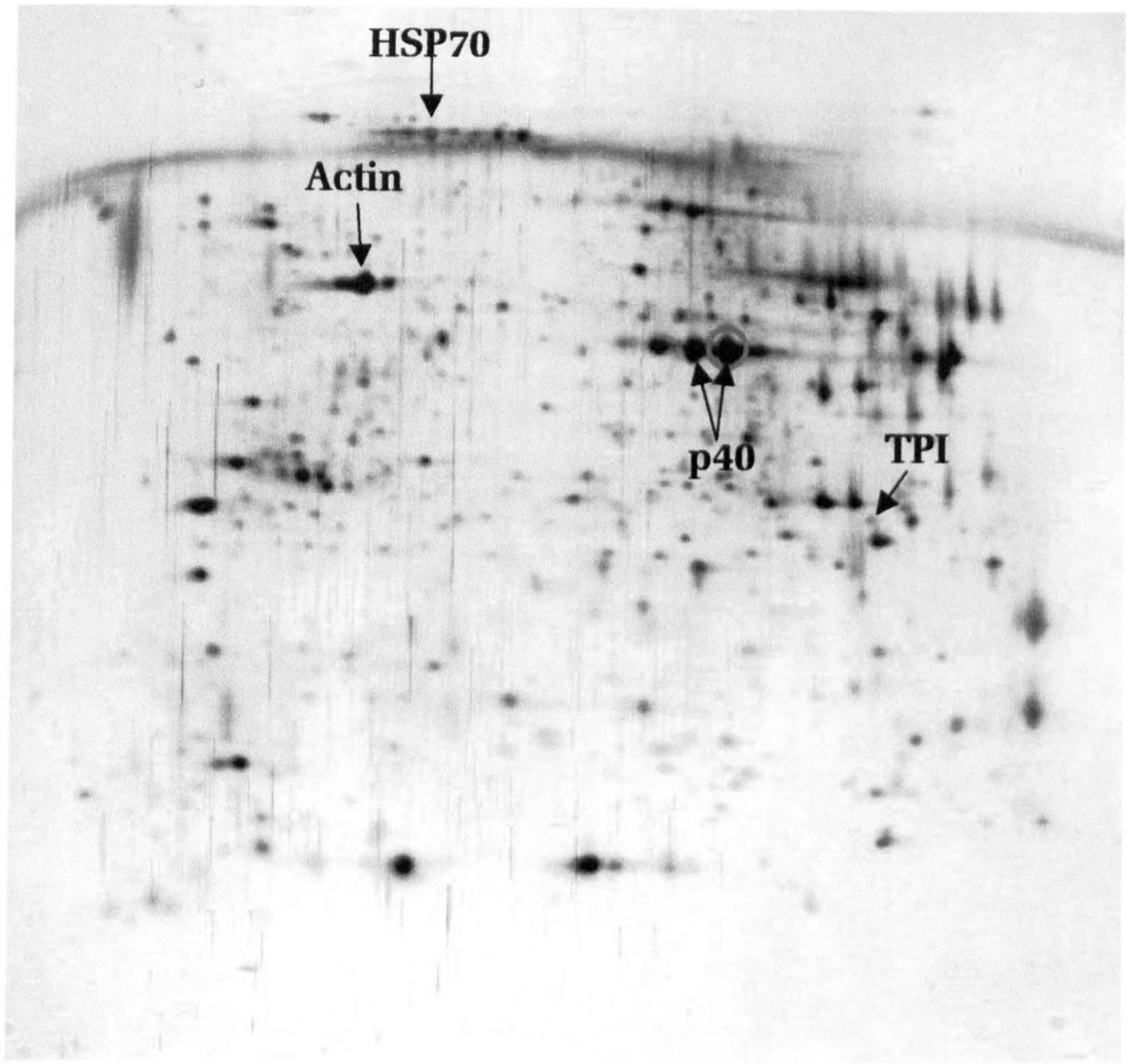


Fig 4.1: A silver-stained 2-D PAGE gels of 100μg of SEA proteins.

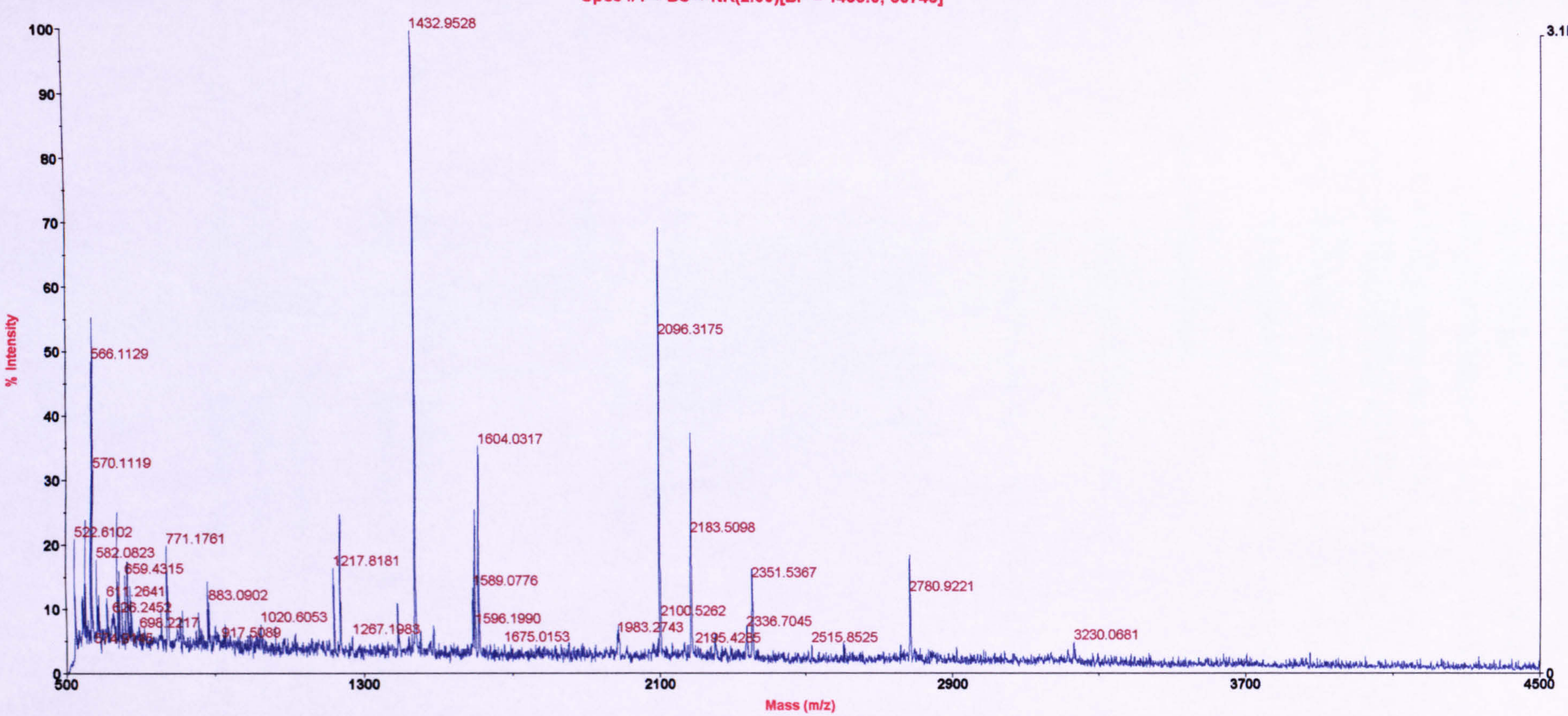


Fig 4.2 Peptide mass fingerprint of a spot from the 2-D gel of SEA (circled in Fig 4.1), identified as the "major egg antigen" SmP40

BLANK IN ORIGINAL

Table 4.1 (Opposite) Output of the MASCOT program when run with the peptide masses from the spectrum in Fig 4.2. The top section of the output shows the distributions of the "probability-based MOWSE scores" for the proteins that were retrieved. The section headed "Index" shows a list of the highest scoring proteins, which contain the most matching peptides. The section headed "Results List" lists, for each protein, the peptides which have matched. Of the top four entries, three refer to different database entries for the same protein SmP40.

{MATRIX SCIENCE} Mascot Search Results

User : Peter Ashton

Email : pda2@YORK.AC.UK

Search title : SEA3

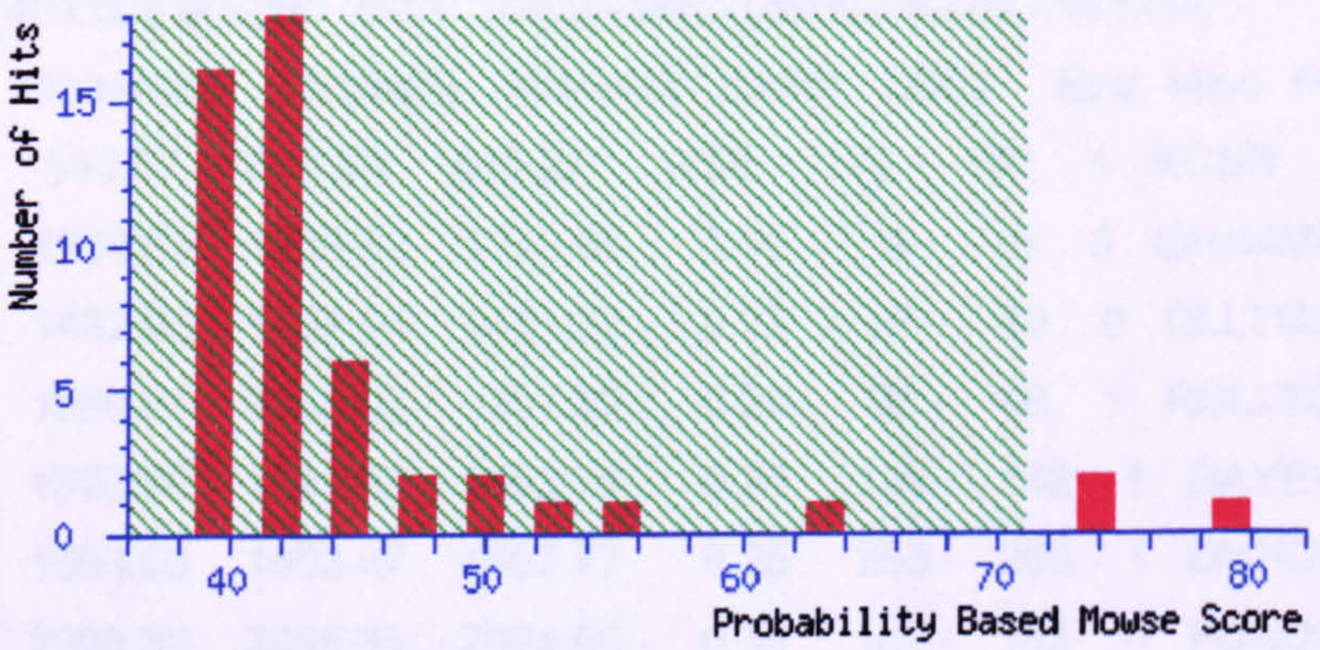
Database : NCBInr 20010915 (755535 sequences; 240612516 residues)

Top Score : 79 for gi|12519290, (AE005659) Z5888 gene product [Escherichia coli O157:H7 EDL933]

Probability Based Mowse Score

Score is -10*Log(P), where P is the probability that the observed match is a random event.

Protein scores greater than 71 are significant (p<0.05).



Protein Summary Report

Index

	Accession	Mass	Score	Description
1.	gi|12519290	17229	79	(AE005659) Z5888 gene product [Escherichia coli O157:H7 EDL933]
2.	gi|627088	39557	73	40k egg antigen (clone 10F5) - fluke (Schistosoma mansoni)
3.	gi|129359	39575	73	MAJOR EGG ANTIGEN (P40)
4.	gi|161038	38250	63	(M15509) major egg antigen [Schistosoma mansoni]

Results List

1. gi|12519290 **Mass:** 17229 **Score:** 79

(AE005659) Z5888 gene product [Escherichia coli O157:H7 EDL933]

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Peptide	
644.08	643.08	643.40	-0.33	13 -	17	1	VKELR	
659.43	658.42	658.27	0.15	36 -	40	0	SYMSR	1 Oxidation (M)
666.67	665.66	665.31	0.36	126 -	131	0	SMETAK	
881.12	880.11	880.39	-0.28	100 -	108	0	DGSFGVGDK	
1234.83	1233.82	1233.54	0.28	31 -	40	1	CGFARSYMSR	
1432.95	1431.94	1431.70	0.24	114 -	125	0	FGTFSEALEYLR	
1589.08	1588.07	1587.80	0.27	113 -	125	1	RFGTFSEALEYLR	
2096.32	2095.31	2094.99	0.32	114 -	131	1	FGTFSEALEYLRSMETAK	1 Oxidation (M)

No match to: 522.61, 542.18, 546.13, 550.67, 566.11, 568.13, 570.11, 582.08, 638.63, 674.03, 771.18, 773.14, 1217.82, 1388.01, 1593.99, 1604.03, 2183.51, 2249.47, 2351.54, 2780.92

2. gi|627088 **Mass:** 39557 **Score:** 74

40k egg antigen (clone 10F5) - fluke (Schistosoma mansoni)

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Peptide
550.67	549.66	549.27	0.39	178 -	181	1	KCSR
1234.83	1233.82	1233.66	0.16	6 -	16	0	QHNAVSIPVNR
1432.95	1431.94	1431.72	0.22	27 -	40	0	DLLTGLEHGGGAHR
1589.08	1588.07	1587.82	0.25	26 -	40	1	RDLLTGLEHGGGAHR
1593.99	1592.98	1592.75	0.24	129 -	142	1	DAYEVGEDGKVHFK
1604.03	1603.02	1602.77	0.25	253 -	268	1	GVHGLSYVDDGSGGKR
2096.32	2095.31	2094.96	0.35	145 -	163	0	FDAQGFAPQDINVTSSENR
2183.51	2182.50	2182.13	0.37	216 -	235	0	VDQNQSLTLNESGQVAVRPK
2351.54	2350.53	2350.13	0.40	143 -	163	1	VRFDAQGFAPQDINVTSSENR

No match to: 522.61, 542.18, 546.13, 566.11, 568.13, 570.11, 582.08, 638.63, 644.08, 659.43, 666.67, 674.03, 771.18, 773.14, 881.12, 1217.82, 1388.01, 2249.47, 2780.92

3. gi|129359 **Mass:** 39575 **Score:** 74

MAJOR EGG ANTIGEN (P40)

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Peptide
550.67	549.66	549.27	0.39	178 -	181	1	KCSR
1234.83	1233.82	1233.66	0.16	6 -	16	0	QHNAVSIPVNR
1432.95	1431.94	1431.72	0.22	27 -	40	0	DLLTGLEHGGGAHR
1589.08	1588.07	1587.82	0.25	26 -	40	1	RDLLTGLEHGGGAHR
1593.99	1592.98	1592.75	0.24	129 -	142	1	DAYEVGEDGKVHFK
1604.03	1603.02	1602.77	0.25	253 -	268	1	GVHGLSYVDDGSGGKR

2096.32	2095.31	2094.96	0.35	145 -	163	0	FDAQGFAPQDINVTSSENR
2183.51	2182.50	2182.13	0.37	216 -	235	0	VDQNQSLTLNESGQVAVRPK
2351.54	2350.53	2350.13	0.40	143 -	163	1	VRFDAQGFAPQDINVTSSENR

No match to: 522.61, 542.18, 546.13, 566.11, 568.13, 570.11, 582.08, 638.63, 644.08, 659.43, 666.67, 674.03, 771.18, 773.14, 881.12, 1217.82, 1388.01, 2249.47, 2780.92

4. gi|161038 **Mass:** 38250 **Score:** 64

(M15509) major egg antigen [Schistosoma mansoni]

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Peptide
550.67	549.66	549.27	0.39	165 -	168	1	KCSR
1432.95	1431.94	1431.72	0.22	14 -	27	0	DLLTGLEHGGGAHR
1589.08	1588.07	1587.82	0.25	13 -	27	1	RDLLTGLEHGGGAHR
1593.99	1592.98	1592.75	0.24	116 -	129	1	DAYEVGEDGKVHFK
1604.03	1603.02	1602.77	0.25	240 -	255	1	GVHGLSYVDDGSGGKR
2096.32	2095.31	2094.96	0.35	132 -	150	0	FDAQGFAPQDINVTSSENR
2183.51	2182.50	2182.13	0.37	203 -	222	0	VDQNQSLTLNESGQVAVRPK
2351.54	2350.53	2350.13	0.40	130 -	150	1	VRFDAQGFAPQDINVTSSENR

No match to: 522.61, 542.18, 546.13, 566.11, 568.13, 570.11, 582.08, 638.63, 644.08, 659.43, 666.67, 674.03, 771.18, 773.14, 881.12, 1217.82, 1234.83, 1388.01, 2249.47, 2780.92

Two-Dimensional Electrophoresis of ESP

To determine whether ESP was more complex than the six bands present on a 1-D gel would suggest, a 2-D gel of 50μg of ESP was run and is shown in Fig 4.3; it demonstrates that ESP can be further resolved into a total of 21 spots. These spot appears to be clustered into three distinct groups. Firstly, there is a group of very basic spots (p*I* > 9) (1-6) which are arrayed in a regular pattern on the right hand side of the gel. Second, a group of low molecular weight, less basic spots (8-12) and finally a group of slightly acidic spots (13-22). Of these, spots 1-6 are the largest, and analysis of the spot volumes using the 2-D Advanced software suggests that they comprise >50% of the total protein on the gel. As with the 1-D analysis of SEA and ESP (Chapter 2), the ESP proteins cannot be seen as major components of SEA (Fig 4.1).

Two-Dimensional Zymography of ESP

Since proteolytic activity was detected using 1-D zymograms of ESP (Chapter 2), it was decided to attempt to extend this technique to 2-D electrophoresis. A gel was run using 50μg of ESP on a 7cm IPG strip and running the second dimension using a conventional mini-gel, containing 0.1% casein as a protease substrate. Whilst the proteolytic activity was not as strong as that detected on the 1-D gel, two distinct areas of clearing can be observed (Fig. 4.4). These presumably correspond to the two bands originally seen. Although it is difficult to determine exactly which spots are giving the proteolytic activity, it is clear that it resides in one of the slightly acidic spots (in the group 13-22) and can therefore be attributed to one of the minor ESP components.

Glycan Staining of ESP

In order to assess whether the ESP protein were glycosylated a gel was stained using a fluorescent label that covalently bonds to oxidised glycans (Fig 4.5). Staining was observed in 3-4 of the basic proteins on the right hand side of the gel.

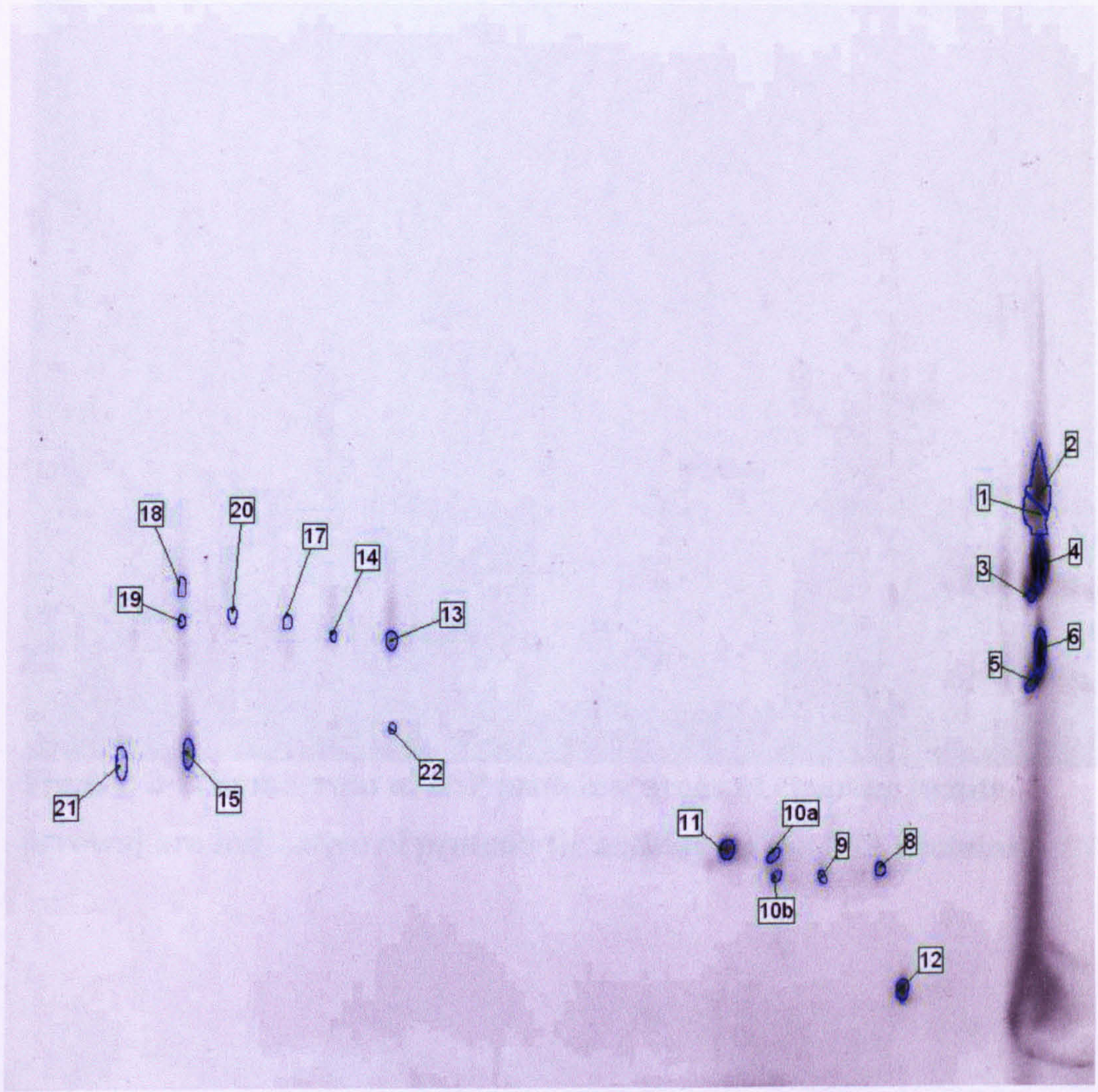


Fig 4.3: A 2-D PAGE gels of 50μg of ESP proteins stained with Coomassie Blue G-250. Spots were highlighted after detection using the Phoretix 2-D Advanced software.

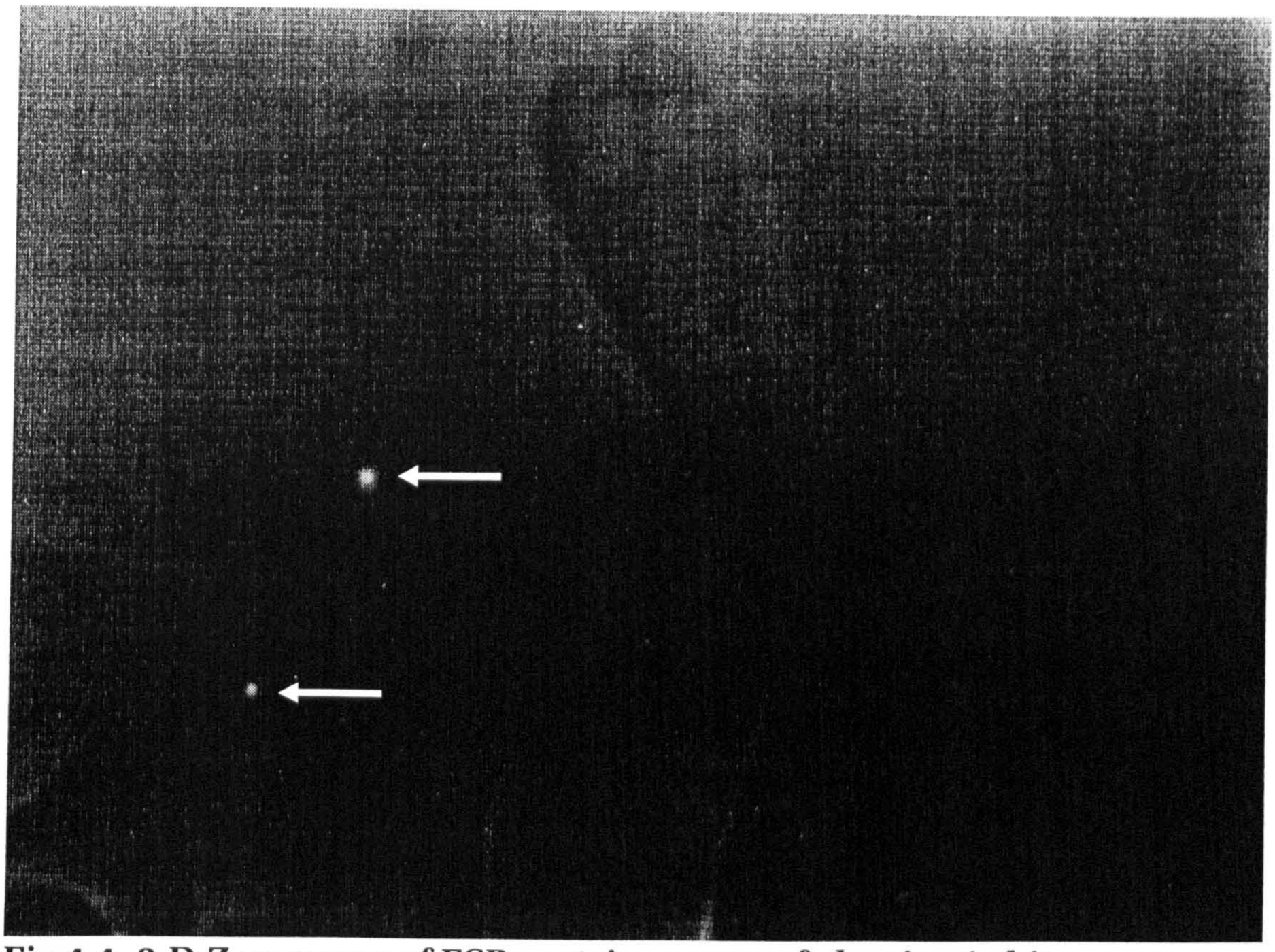


Fig 4.4: 2-D Zymogram of ESP proteins, areas of clearing (white arrows) are indicative of proteolytic activity by the ESP proteins.

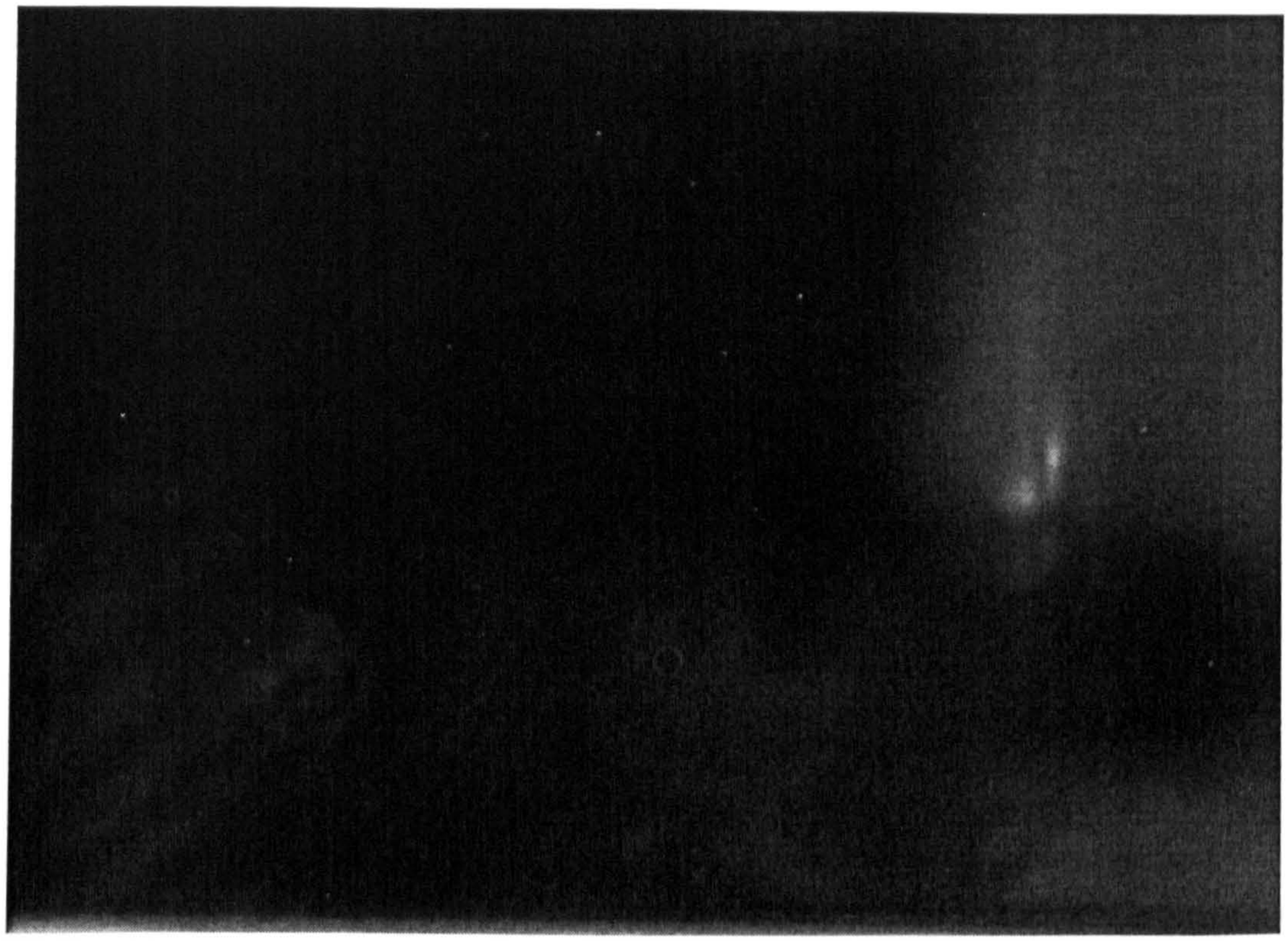


Fig 4.5: Fluorescent staining of a minigel of ESP using dansyl hydrazine. Light areas show the incorporation of the fluorescent label.

Peptide Mass Fingerprints of ESP Proteins

Examples of peptide mass fingerprints of the ESP proteins are shown in figure 4.4A-C. (A complete set of PMFs for the ESP proteins is given in Appendix 1.) As can be seen, clean strong spectra were obtained for all of the spots; however, searching of the NCBI nr databases with the PMFs using the MASCOT system yielded no significant matches, and none for *S. mansoni* proteins.

Similarity Analysis of ESP PMFs

Although no identifications were obtained using the PMFs, it can be seen that several of the spectra share a number of common peaks (e.g., peaks with masses of 1634, 1652 & 2143, amongst others in Figs 4.4B & 4.4C). In order to assess the extent of the similarity between the PMFs, the lists of masses were transformed into a NEXUS format data matrix, which could be used by the PAUP* program (Appendix 2). Due to the large number of number of different masses that were observed, and the relatively large number of protein spots, generation of a single unambiguous tree would have required an immense amount of computer time. In order to avoid this, rather than examine every possible tree that can be constructed for the 21 spots, the software uses algorithms that reduce the number of trees that are considered, providing an approximation to the true tree.

It was found that varying the order of addition of the PMFs to the analysis resulted in different trees being constructed by the software. This suggested either that there was no unique most parsimonious tree or that discovering such a tree using these approximate methods would not give complete confidence in its uniqueness. When working with phylogenetic data, it is assumed that such a unique tree will actually exist and that it will reflect the true evolutionary relationship between the different species. However, this is not the case with the PMFs of the ESP proteins, unless all of the proteins, for example, were derived from a single parent protein. Rather, it would be expected that there are several distinct genes each of which gives rise to one or more spots on

the 2-D gel. In this case, there would be no way to resolve the interrelations between these subgroups, since PMFs in different groups might have no peptide masses in common. An alternative approach is to perform a bootstrap analysis and to construct a consensus tree. Here, a large number of trees are generated by performing a large number of analyses and varying the order of addition of the PMFs. The consensus tree is constructed by only allowing groupings of PMFs that occur either in all the trees generated (for a strict consensus tree) or in a strict majority of tree (majority-rule consensus tree).

Fig 4.5A shows a strict consensus (i.e., groupings are only shown when they occur in all the trees generated in the analysis), whereas in Fig 4.5B, groupings are allowed when they occur in >50% of the trees. The only difference is whether spots 1 and 2 are included with the group containing spots 3-6. In both trees, two major groups were observed. Most obvious, spots 1-6 all cluster into a single group, as might be inferred from the regularity of their distribution on the 2-D gel. The second major group comprises spots 8-12. The remaining spots cluster into four smaller groups.

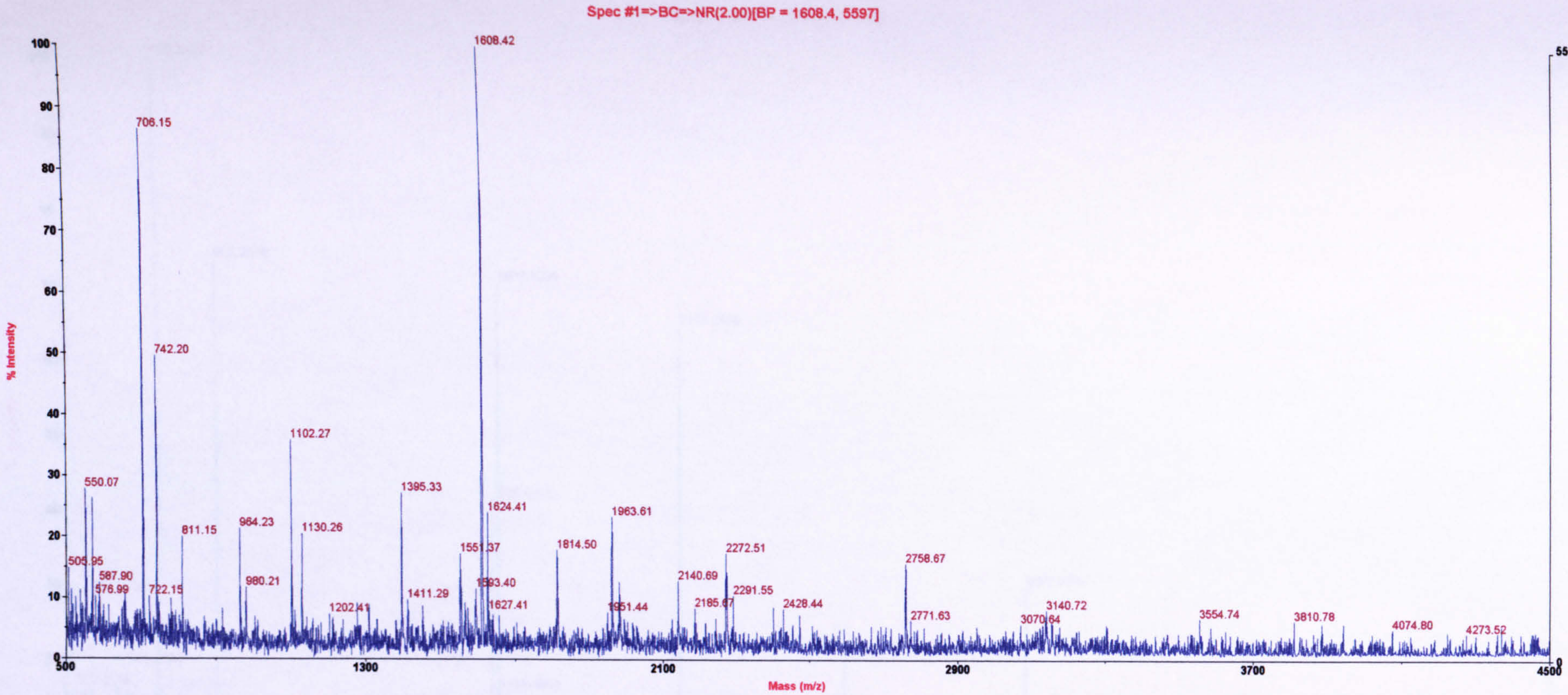


Fig 4.4A Peptide mass fingerprint of ESP spot number 1.

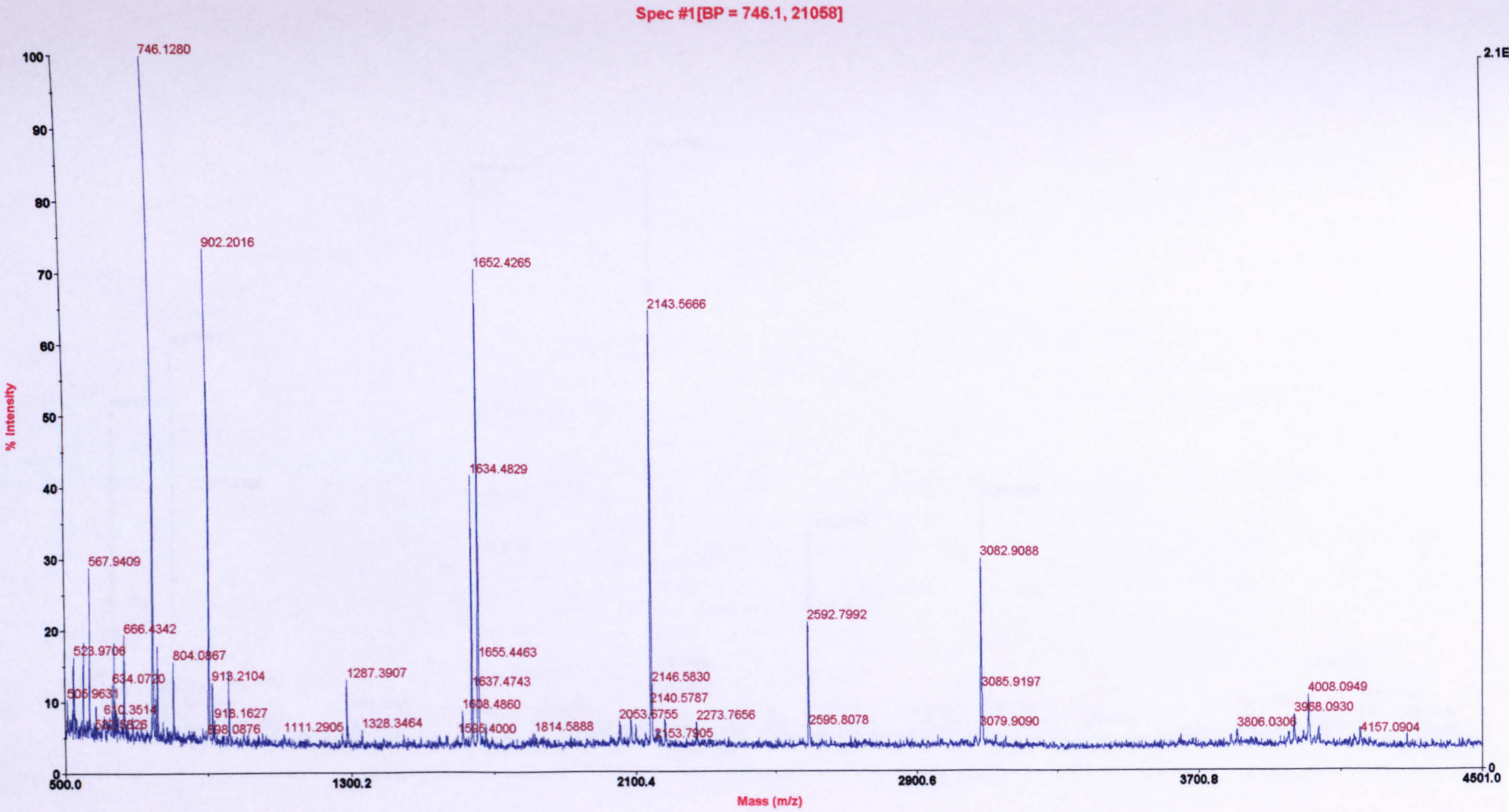


Fig 4.4B Peptide mass fingerprint of ESP spot number 3

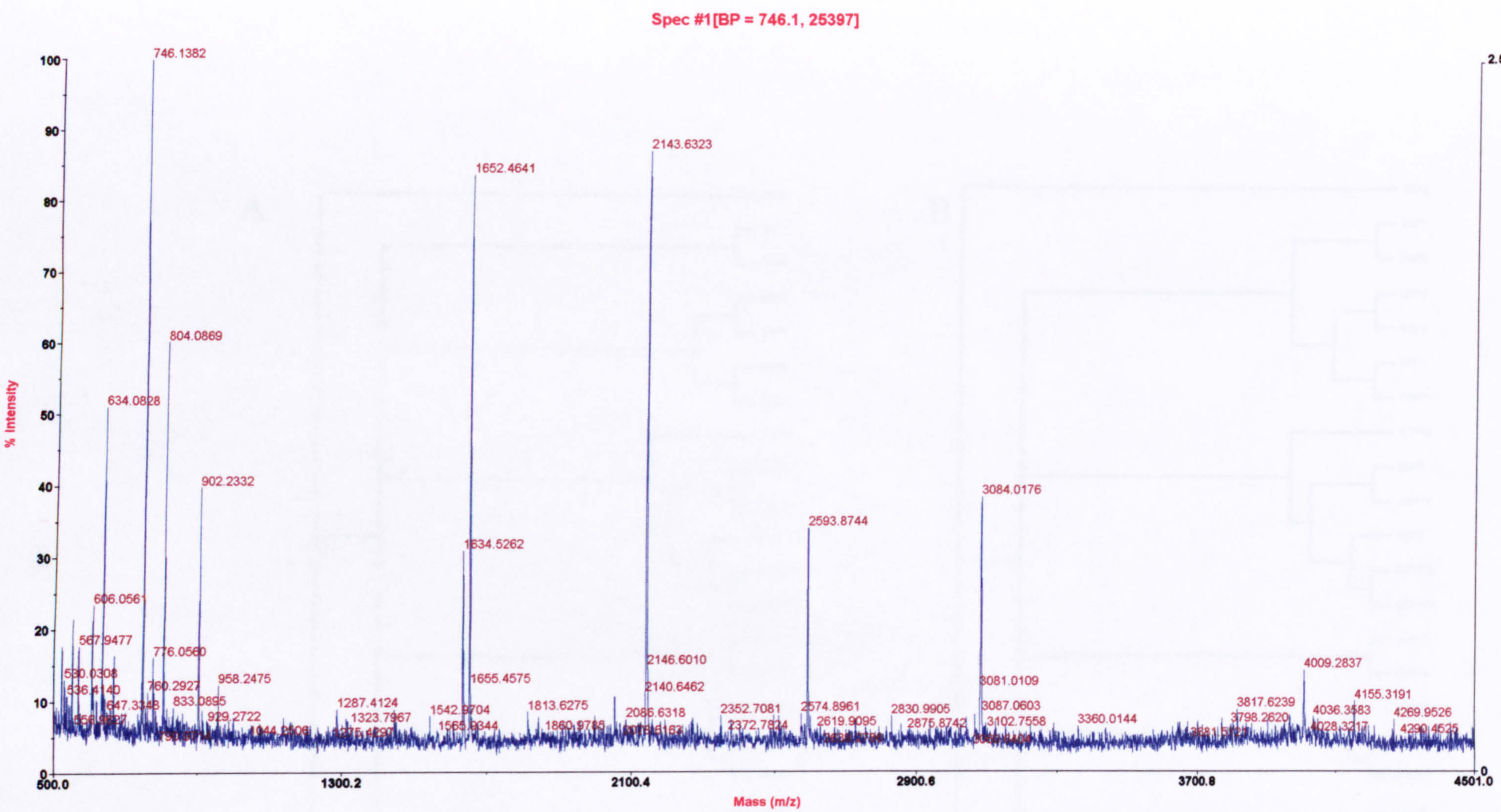


Fig 4.4C Peptide mass fingerprint of ESP spot number 4

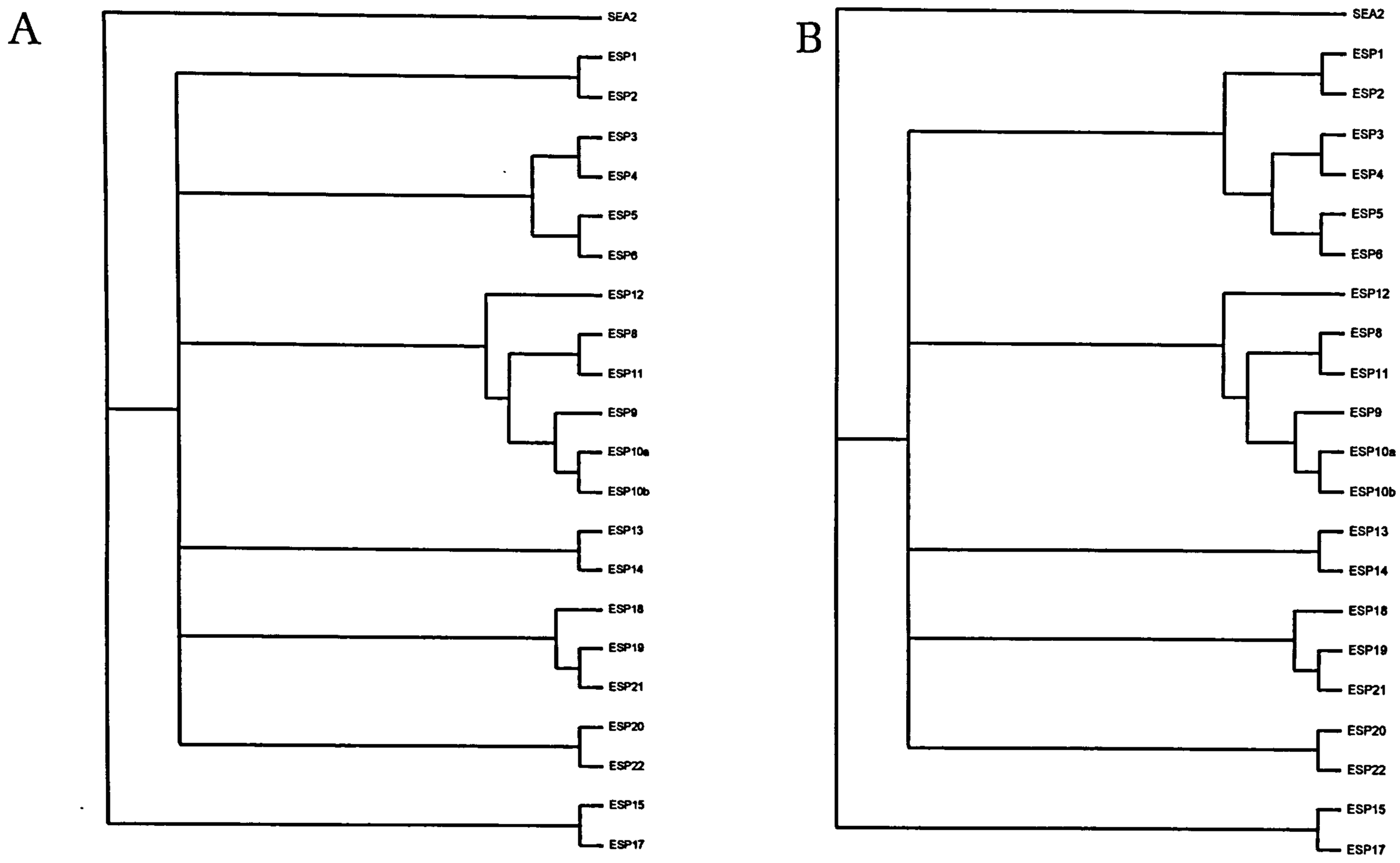


Fig 4.5: Strict consensus (A) and majority-rule consensus (B) trees of ESP proteins based on shared peptide masses in PMFs.

Discussion

Two-dimensional electrophoresis of ESP reveals that whilst there are more proteins present than suggested by 1-D SDS-PAGE analysis, the mixture is still relatively simple. Even more sensitive silver staining fails to show any greater complexity. In addition, as noted with the 1-D gels in Chapter 2, the ESP proteins to not appear to be major components of SEA, and this is probably due to the enrichment of the ESP components during the culture period. Another possible explanation is that the ESP proteins are not present in the same form in ESP and SEA, and are, for example, proteolytically processed after secretion. They would therefore not necessarily appear in the same positions on the 2-D gels.

One of the most noticeable features of the 2-D gel of ESP is the clustering of spots 1-6 at the high-pH side of the gel. Possession of a high pI indicates that these proteins would exhibit a considerable positive charge at physiological pH (7.2). Proteins which have pIs in this range are known to interact with negatively charged partners (e.g., histones, which bind DNA). It is therefore tempting to suggest that the proteins in spots 1-6 also interact with ligands or substrates that are negatively charged. Since one of the possible roles of ESP is the disruption of host tissue, so that the eggs can migrate from the blood vessels of the mesenteric system into the lumen of the gut, the most obvious candidate for this partner would be extracellular matrix components that contain sulphated glycans. However, the zymogram shows that none of the high pI proteins possess any proteolytic activity. It is therefore possible that the interaction of the proteases and their presumed tissue substrates will be indirect, with the high pI proteins acting to bring the tissue and the proteases together or to keep the proteases localised to the immediate vicinity of the egg.

The presence of glycosylation on at least some of the ESP proteins lends additional weight to the supposition that they are true secretions, rather than simply resulting from the leakage of material from eggs that

die during the culture period. In order for the proteins to acquire these glycans, they must have passed through the endoplasmic reticulum, where the initial glycans are attached, and the Golgi, where the glycans are modified and elaborated before the proteins are exported from the cell. The extent of glycosylation presented here is likely to be an underestimate, due to the sensitivity of the staining method, and it is possible that loading more protein onto the gel, or using a more sensitive staining technique would reveal glycosylation on a greater proportion of the spots.

The lack of firm identification of any of the gel spots by peptide mass fingerprinting is disappointing, but perhaps not surprising since the coverage of the schistosome genome by the full-length sequences in both the NCBI nr and SwissProt databases is limited, and the ~400 available full-length *S. mansoni* sequences represent only ~2-3% of the expected 16-20,000 genes. Although it is tempting to say that all of the ESP components must therefore be novel, there is a problem, since thioredoxin peroxidase has been demonstrated to be secreted from the live egg during *in vitro* culture [Williams *et al.* 2001] and should therefore have been identified by this method. There are several possible explanations why this did not happen. Firstly, it is known that the thioredoxin peroxidase is a minor component of the mixture [D.L. Williams, *pers. comm.*] and may therefore not be visible on the Coomassie-stained gel of ESP that was used as the source of the gel spots. However, since silver staining of gels containing similar amounts of ESP shows no new spots, this is unlikely. Secondly, presence of proteases in the mixture could be problematic, since there is no reason why they could not digest the other proteins in the mixture. In fact, to prepare ESP, the eggs are incubated for 3 days at 37°C, and the requirement to keep the eggs alive during this period precludes the use of protease inhibitors in the culture medium. Any proteolysis could therefore be expected to be complete by the time the secretions are collected and this could explain the apparent absence of thioredoxin

peroxidase among secreted proteins, since it could have been digested into fragments that are too small to show up on a 15% PAGE gel. Even if these fragments can be visualised as several distinct spots, the each of the PMFs obtained would contain only a subset of the peptides. There may not be a sufficient number of peptides in each spot to allow the search algorithms to treat thioredoxin peroxidase as a good match.

Although none of the ESP proteins were identified by peptide mass fingerprinting, this is not simply a failure of the technique, since spots were successfully identified from the gel of SEA. However, this exercise has ruled out all of the known, well characterised *S. mansoni* proteins as components of ESP. In particular, p40 and the known schistosome proteases (e.g., elastase [Newport *et al.*, 1987] and cathepsins B [Asch and Dresden, 1979] and L [Michel *et al.*, 1995]) are not present in ESP. In addition, by using techniques that are more commonly applied to phylogenetic analysis, it has been possible to make some inferences about the possible relationships between the ESP proteins. These methods attempt to construct phylogenetic trees of several species on the basis on the number of "characters" that they share. These were originally phenotypic or gross morphological features, but nowadays are more commonly derived from DNA data, such as the presence of a particular base at a given position in a sequence alignment. To analyse to ESP PMFs, the characters used were the presence (or absence) of a particular peptide mass in a particular peptide mass fingerprint. It is striking that the grouping of the PMFs arrived at by this analysis relates directly to the apparent grouping of the proteins on the gel, with spots 1-6 forming one cluster, spots 8-12 forming another and the remaining spots (13-22) split into four distinct groups.

There are several possible reasons that this clustering of spots can occur. Firstly, members of a particular group may be members of a single gene family which has arisen by gene duplication and subsequent divergence, or they may represent different transcripts arising by

alternative splicing of the mRNA from a single gene. Another possibility is that a small number of proteins are subjected to a range of post-translational modifications such as glycosylation (which would alter the molecular weight) or phosphorylation (which would alter the charge and therefore the pI). Alternatively, it may be that the different spots arise from a single parent protein by proteolysis. Given that some of the ESP proteins are indeed glycosylated and that there are proteases present in the mixture, it is likely that the final explanation will be a combination of these factors. Whilst the final explanation of this relatedness must wait for the full sequences of all the ESP components to become available, taken together these data suggest that there may be as few as 6-7 distinct genes which account for the 21 spots on the 2-D gel of ESP, and that most, if not all, of these spots are encoded by previously uncharacterised genes.

Chapter 5: Tandem MS Analysis of ESP

Introduction

MALDI-ToF/MS provides a rapid and accurate means of measuring the masses of a mixture of peptides resulting from the proteolytic digestion of a protein, and thereby identifying it. However, this method only works if the full sequence of the protein in question is available. In the case of *Schistosoma mansoni*, only a small number of such sequences are present in the various online databases and so it is necessary to attempt to link the proteins with ESTs, which are short single pass nucleic acid sequences obtained from cloned mRNAs [Franco *et al.*, 1995]. To achieve this link, *de novo* peptide sequence information must be obtained, which can then be compared to six-frame translations of the ESTs. One method that can be applied is Edman degradation of a purified protein or from a band from an SDS-PAGE gel blotted onto PVDF [Edman & Begg, 1967], but this has two major drawbacks. Firstly, the N-terminus of the protein may be blocked (e.g., by formylation or any of a number of other post-translational modifications) and the sequencing reaction prevented from proceeding. Secondly, any sequence that is obtained will be from the N-terminal end of the protein and will therefore be encoded by the far 5' end of the mRNA. ESTs are typically generated by reverse transcription using an oligo-dT primer that anneals to the poly-A tail at the 3' end of the mRNA. However, reverse transcription is an imperfect process and often terminates prematurely, before the 3'-end of the cDNA (corresponding to the 5' end of the mRNA) has been synthesised. It is, therefore, precisely the N-terminal sequence information that will be missing from the EST, and this makes the matching of Edman-obtained sequences with ESTs unlikely.

Although it is possible to generate internal peptide sequence data using the Edman technique, e.g., by performing a partial proteolytic or cyanogen bromide digest of a protein and separating the peptides by

HPLC or using a Tris-Tricine gel [Schagger & von Jagow, 1987], it is now becoming more common to use one of several mass spectrometric techniques [Mann & Wilm, 1994]. These have the advantage of increased sensitivity and do not rely on having to purify the individual peptides resulting from the digest of a protein. Methods are available for peptide sequencing using MALDI-ToF/MS systems, but they all have significant drawbacks and the approach usually taken is to employ some form of tandem mass spectrometer.

MALDI Post-Source Decay

By increasing the laser energy during a MALDI experiment, the peptide ions produced can be caused to fragment after they have left the ion source, and this is referred to as metastable decay or post-source decay (PSD). Since there is no accelerating force on the peptides in the flight tube, all of the fragments of each ion will travel together at the speed of the original ion. A timed ion selector placed immediately before the reflector can then be used to select all the daughter ions of a particular parent ion. As these enter the reflector, smaller fragments will penetrate less than large ones, and so they will be separated, and their masses can then be measured as they arrive at the reflector detector. In the case of MALDI systems containing a linear reflector (the vast majority of current systems), it is not possible to set a single reflector voltage that will accurately focus all of the fragment ions. The product ion spectrum is, therefore, usually acquired in a series of steps, altering the reflector voltage for each step in order to focus a subset of the ions correctly at the detector. The spectra for each of these steps are then stitched together to produce a single product ion spectrum. There is one system available now which contains a reflectron with a quadratic field, and it has been reported that it is possible to acquire full PSD fragment spectra in a single step [Colburn *et al.*, 2000].

By interpreting the masses of the fragment ions, it should be possible to deduce the amino acid sequence of the parent peptide [Kaufmann, Spengler & Lutzenkirchen, 1993]. However, the

fragmentation of the peptide is not controlled in any way and therefore, PSD spectra tend to be complex, with not just C- or N-terminal fragments, but also many internal fragments that result from multiple cleavages of a single peptide. If the sequence of the peptide is known (or at least suggested, e.g., by peptide mass fingerprint searching), it is possible to back-label the spectrum with the identities of fragments that are predicted from the sequence. Although this provides useful supporting evidence if the identity of the protein has already been proposed (e.g., by peptide mass fingerprinting), it again relies on the full-length sequence of the protein being available in the database.

One alternative is to attempt to modify the peptides in a way that will simplify the PSD spectrum, e.g., by covalently coupling a group with a negative charge to the N-terminus of the protein [Keough *et al.*, 1999]. Peptides are normally ionised by the addition of a single proton to a basic group on the peptide (e.g., the C-terminal arginine or lysine for tryptic digests), but in the presence of a fixed negative charge at the N-terminus, this will lead to the formation of a neutral ion. A second ionising proton must also be acquired in order for the peptide to be accelerated and this extra proton has two effects. Firstly, it will usually attach to the peptide at one of the nitrogen atoms in the peptide backbone (which acts as a Lewis base). The peptide bond at the site of attachment will then become more labile, and therefore be preferentially cleaved during PSD (referred to as charge-site directed fragmentation). Secondly, after cleavage, the N-terminal fragment will contain the negatively charged group plus the second ionising proton, and since it possesses no overall charge it will not be stopped by the reflection. The PSD spectrum will be considerably simplified, theoretically containing only the C-terminal fragments of the peptide (the y-ions). Whilst this technique appears promising, there are some practical problems associated with it, in particular, the requirement that the derivatisation of the peptide must be performed in anhydrous conditions for the yield to be sufficient to produce an interpretable spectrum.

Carboxypeptidase Digestion

Another approach that has been taken is to purify a single peptide from the mixture and digest it with one or more carboxypeptidases (usually Y and P) [Bonetto *et al.*, 1997]. These cleave successive amino acid residues from the C-terminal end of the peptide, leaving a range of fragments which all contain the N-terminus. The mass differences between the fragments will correspond to the masses of the residues and hence the amino acid sequence can be deduced. However, this approach cannot be applied to mixtures of peptides, and it relies on having sufficient material available to purify an individual peptide from a tryptic digest of a gel spot. A better approach is to use a true tandem MS method.

"True" Tandem MS

Tandem mass spectrometry refers to the technique of coupling two separate mass analysis steps. The first step allows a single peptide to be selected from a mixture. This peptide is then generally fragmented using collision-induced dissociation (CID) and the fragments analysed by the second step. Several systems are available to do this, including a new MALDI-ToF-ToF system [Medzihradszky *et al.*, 2000]. However, the most popular systems used to date have been either triple quadrupoles or quadrupole-ToF systems. These both use a different method of ionisation in which the sample is introduced into the machine in an acidified volatile solvent, which evaporates rapidly once in the vacuum to produce a beam of ionised peptides. Generally, the needle used to inject the sample is held at a high voltage, and the ionisation technique is known as electrospray, or nanoelectrospray, ionisation (ESI).

The first analyser in both systems is a quadrupole: a simple arrangement of four metal rods, to which oscillating electric fields are applied. By altering the frequency and phase variation of these fields, it is possible to tune the quadrupole to permit through ions within a specific range of charge to mass (m/z) ratios. In this way, it is possible to

either select ions in a specific m/z range, or to scan across the whole range of possible m/z values to produce a spectrum similar to a peptide mass fingerprint. However, unlike MALDI, ESI tends to produce multiply charged ions, and the spectrum may well contain several different charge states of the same peptide with different m/z values. In order to determine the charge state of a given peak in an ESI spectrum, it is necessary to examine the spacing of the individual isotope peaks for a given peptide. Since the isotope peaks will differ by exactly one Dalton, the charge state will simply be the reciprocal of the m/z difference between successive peaks. In such spectra, peptides in particular will tend to be multiply charged since they contain multiple potential protonation sites, and this helps to distinguish them from other, contaminating, molecules. Software is available that will perform "deconvolution" of the spectrum (e.g., MaxEnt; Micromass) to display the spectrum as if all the ions were singly charged [Ferrige, Seddon & Jarvis, 1991]. It is necessary, however, to examine the raw spectrum, since, when performing the tandem MS, the first analyser must be set to the *m/z* and not to the *mass* of the peptide of interest.

Once through the first analyser, the beam of peptide then enters a collision cell containing a small amount of an inert gas (e.g., argon), inducing the peptides to fragment. Although this fragmentation can occur at any bond on the molecule, under conventional conditions, it is restricted to the peptide backbone, and in particular, the predominant cleavage occurs at the C—N bond between amino acid residues, yielding a series of y- and b-ions (Chapter 3, Box 3, Fig. III). The fragments then enter the second analyser, which is either a second quadrupole, which is scanned across the mass range to give the product ion spectrum, or can also be a time-of-flight (ToF) analyser. In the case of a ToF, the constant stream of ions must be converted to pulses, so that the time between the start of the pulse and the ions arriving at the detector can be measured. This is achieved by positioning the ToF flight tube orthogonally to the ion stream emerging from the collision cell and applying a pulsed

voltage to a pusher plate. The arrival time at the detector allows the m/z of the ions to be calculated exactly as for the MALDI-ToF systems.

Once the product ion spectrum has been deconvoluted, it can be interpreted by finding a series of peaks, which differ by exactly the monoisotopic masses of given amino acid residues. The first step is generally to identify the C-terminal residue (the y_1-ion) that, for tryptic peptides, will usually be a lysine (147Da) or arginine (175Da). From this starting point, and with a sufficiently good spectrum, is it possible to proceed step by step to the mass of the parent ion. Alternatively, it is possible to work from the parent ion mass to the C-terminal end of the peptide, or to look for the b-ion series rather than the y-ions. If neither the y_1-ion nor the parent ion is obvious in the spectrum, it may also be possible to work from a peak that exhibits one of the neutral losses (e.g., -17Da for NH_3 or –18Da for H_2O) that are characteristic of y-ions. Once again, though, software is generally used to simplify this process. This also has the advantage that, if a particular sequence assignment is made, the software can automatically look for the y- or b-ions in the spectrum, or indeed the a- and z-ions which can also be present. Although this process sounds simple, it relies on the quality of the spectrum being sufficiently high. Since this is not always possible, often no sequence or only a partial sequence (peptide sequence tag; PST) is produced. Whilst this technique is very powerful, particularly for unknown proteins, there are some limitations. In particular, leucine and isoleucine have identical masses, and so cannot be distinguished. In addition, since the mass accuracy of the Q-ToF machine is of the order of 0.05Da, it is not possible to reliably discriminate between lysine and glutamine, which differ by only 0.04Da.

In this chapter, peptides from each of the 21 spots on the 2-D gel of ESP are subjected to tandem mass spectrometry. Forty-two peptide sequences were generated, and used to search the GenBank database. This enabled two ESTs, which encode the ESP proteins, to be identified.

Materials and Methods

2-D Gels and Tryptic Digests

2-D gels of ESP and tryptic digest of the gel spots were produced exactly as in chapter 4; in fact, the same digests were used for both the MALDI and the Q-ToF analysis.

Q-ToF Analysis

Tandem MS analysis of the peptide mixtures was performed on a Q-ToF instrument (Micromass, Manchester) at the Department of Biochemistry and Molecular Biology, University of Leeds and was operated by Dr. Alison Ashcroft. The dry peptide samples were dissolved in 0.1% formic acid and an equal volume of methanol added. 2-3μl of each sample was loaded into a nanospray needle (Protana, Odense, Denmark) and mounted in the ion source. The voltage was applied (400kV) and the needle observed to ensure that spraying had begun. If the needle did not spray, sheath gas was turned on, or the needle re-loaded, until spraying started. For each peptide mixture, a ToF only spectrum was acquired initially, and the charge to mass ratios and charge states of all visible multiply charged ions was noted. For each multiply charge ion, the quadrupole was set to pass through the m/z of interest (plus any isotopic peaks of the same peptide), the collision energy turned on and the fragment ion spectrum acquired. The collision energy was manually adjusted throughout the acquisition to ensure that the fragment ions of different sizes were represented roughly equally in the spectrum.

Data Analysis

Spectra were analysed using Micromass MassLynx software. The fragment ion spectra were first deconvoluted using MaxEnt 3, to present the spectrum as if all fragments were singly charged. Initially, automatic peptide sequencing was attempted for all the spectra using the PepSeq software (Micromass). If this failed, or did not give convincing results, manual interpretation of the spectra was performed.

Searching using Peptide Sequence Tags

The NCBI nr and dbEST database were interrogated using the standard BLASTP software at http://www.ncbi.nlm.nih.gov/BLAST/. For most searching, the query was limited to matches from *Schistosoma* spp. Where no match was found, the query was re-run against the entire database. Since the searching relied on matching short peptide sequence tags to long protein sequence (or translated DNA sequences), the standard parameters were adjusted to use an expect value of 10000 and a word size of two. This enabled weaker matches between the short peptide sequences and longer ESTs to be retrieved easily.

Resequencing of Identified Clones

Plasmids containing the inserts identified by BLAST searching were obtained from Dr. Gui Oliveira (FIOCRUZ, Belo Horizonte, Brazil). who originally generated the ESTs, and were transformed into competent J101 strain of *E. coli*. The bacteria were plated out on carbenicillin plates and single colonies were selected and grown in 5ml liquid cultures overnight. The cells were then harvested by centrifugation and the plasmid purified using a Qiagen mini-prep kit according to the manufacturer's instructions. Sequencing was performed either by PCR-based dye-terminator chemistry using a ABI kit and sequenced in the Cancer Research Unit at York, or the plasmids were sent to MWG-Biotech (Ebersberg, Germany), for sequencing.

Results

Initial ToF Scans

Whilst all spots were run on the MALDI and gave satisfactory spectra, when analysed using the Q-ToF, every sample was found to contain significant amounts of contaminating polyethylene glycol which can be recognised as a series of peaks which differ by 44Da (a representative spectrum for spot 4 is shown in Fig 5.1). It was assumed that this arose from the Eppendorf tubes in which the samples had been prepared and stored. However, in most cases, it was possible to find

multiply charged ions outside the m/z range of the contamination, or between the peaks of the PEG molecules.

Fragment Ion Spectra

A typical fragment ion spectrum is shown in Fig. 5.2. The interpretation of this spectrum was obtained using the MassSeq software, and is given a high confidence, since not only is the complete y- -ion series present, but also some of the b- and a-ions. Similar spectra were obtained for all of the multiply charged ions. Table 5.1 lists the spot number, the parent ion mass and the interpreted sequence for all of the peptides which fragmented sufficiently well to yield data. In the sequences, [IL] indicates either leucine or isoleucine, and [KQ] indicates either lysine or glutamine. Where the software called either of these, it has been replaced by both possibilities, except in the case of C-terminal lysines. These can be assume to be correct, since all the peptides arising from tryptic digestion would contain either a lysine or arginine at the C-terminus, except for the C-terminal peptide of the whole protein. Groups of two or more residues in round brackets, e.g., (YK) indicate that the order of the residues is not known, usually because the y-ion between the two residues is not present in the spectrum, but that the mass difference between the y_n- and y_{n+2}-ions is equal to the sum of the masses of the residues shown.

It was possible to obtain at least partial sequence from a total of 42 peptides from spots 3, 4, 5, 6, 11, 12, 13, 15 & 17. Spot 15 was most successful, with 11 different peptides fragmenting sufficiently well to yield sequence. It should be noted that several of the peptide sequences are similar (e.g., (ER)(PY)WY[IL]FN... and ERPYWY[IL]FD... from spots 5 and 6, respectively), although not identical, and these are highlighted in Table 5.1.

Searching Databases using Peptide Sequence Tags

The BLAST tool used for the searching of the peptide sequences against the dbEST database is designed to perform local alignments and

similarity searching between homologous proteins. Here it was used to search for short subsequences, which should be present as an exact match in one of the six-frame translations of the EST sequences. As a result, the parameters supplied to the BLAST program must be adjusted in order to allow the program to return appropriate matches. To achieve this, the expect value was set to 10000 and the word size to 2, in line with the recommendations on the NCBI web site. Searching using all of the peptide sequences matched three entries on the dbEST division of GenBank. Two of these ESTs were virtually identical, and matched the same single peptide sequence from spot 15 (Accession numbers AW061382 & AI764941; Fig 5.3). The third EST (AI820476) matched a total of five peptides from spots 3, 4, 5 and 6 (Fig 5.4), and one of the peptide sequences was split over two different reading frames, suggesting that a frame shift had occurred during the sequencing of the EST.

The ESTs retrieved were then used to search the GenBank database, to see if they shared any homology with proteins of known function. No significant homology was found.

Resequencing of EST AI820476

To see if the peptide sequence (ER)(PY)WY[IL]FNGGVNYTGR could be brought into a single reading frame, the EST AI820476 was resequenced. Three sequences runs (two in York and one by MWG) were then aligned with the original EST sequence and a consensus sequence constructed. The relevant part of the translation sequence is shown in Fig 5.5, and demonstrates that the EST now contains both the parts of the peptide sequence obtained by tandem MS, although the middle part of the sequence varies, presumably due either to further DNA sequence errors, or to errors in the peptide sequence.

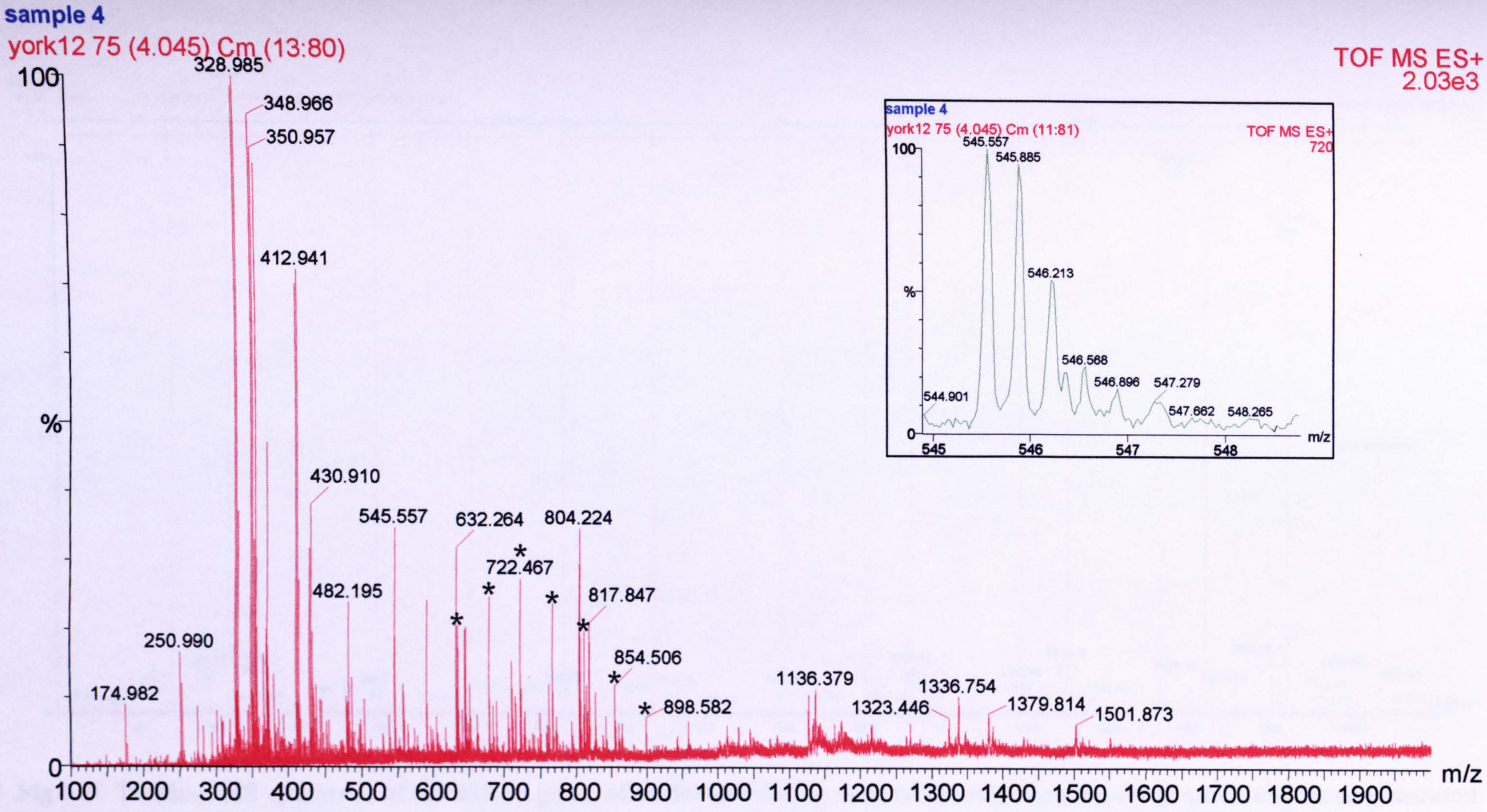


Fig 5.1: TOF MS spectrum from spot 4. The smaller spectrum (inset) shows a close-up of the 545.56Da peak, demonstrating that it contains a triply charged ion. Peaks marked with stars represent the series of contaminating PEG molecules.

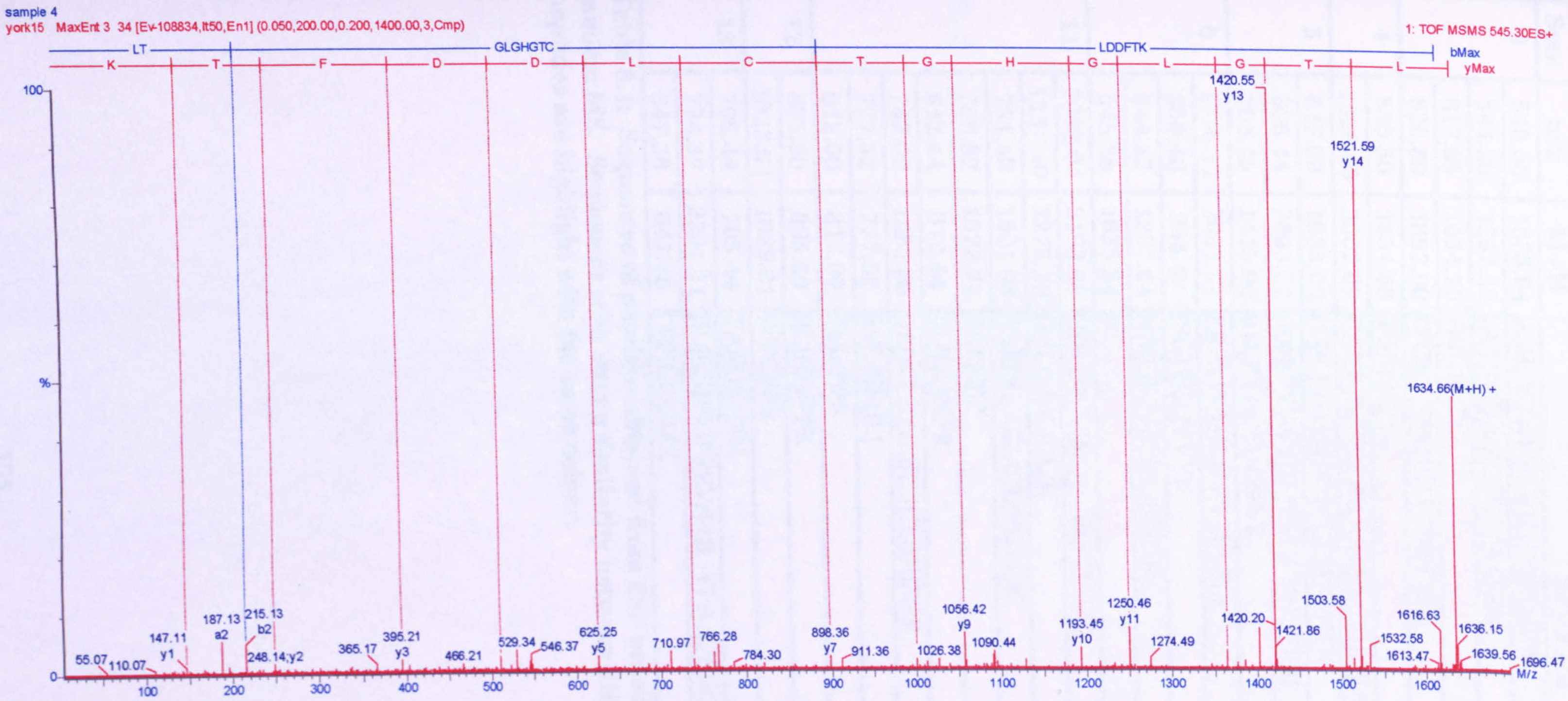


Fig 5.2: Tandem MS spectrum of the 545.56 peak, after deconvolution with MaxEnt3, showing the peptide sequence generated by the PepSeq software (Micromass). This shows that the peptide has the sequence [IL]TG[IL]GHGTC[IL]DDFTK.

Spot	m/z	M+H	Sequence
3	545.88	1635.64	A[IL]YYWD[IL][IL](Y[KQ])ER
	644.26	1287.52	(AT)[KQ]VEC[IL]N[KQ]PK
	817.85	1634.70	DVGPP[IL]GTC[IL]DDFTK
	826.80	1652.60	(YC)[IL][KQ][IL]YDETYER
4	545.56	1634.68	TG[IL]GHGTC[IL]DDMTK
	1028.03	3082.09	CATPEMFR(P[KQ])S([IL][IL]C)(RPDP)VREACFK
5	817.00	1633.00	(T[IL])HHGHCGT[IL]EPDFTK
	698.44	2093.32	(ER)(PY)WY[IL]FNGGVNYTGR
	708.45	1415.90	RD(PY)WY[IL]FVR
6	1046.41	2091.82	ERPYWY[IL]FDTEPYWK
	958.40	958.40	(SG)Y[IL]EVYK
	644.32	1287.64	N[KQ]GVEC[IL]NAGPK
	545.88	1635.64	HGTC[IL]DD(FT)K
11	1082.04	2163.08	H[IL]NVV[KQ]…PLHSASK
	1137.10	2273.20	D[IL]M[IL][IL]K
	751.49	1501.98	SAEDG…GM[IL]FGVGK
	786.87	1572.74	SAENEF
	612.44	612.44	[IL][IL]NPR
	743.00	1485.00	[IL]PSCD[IL][IL][IL]DENPR
	777.52	777.52	RWESS[IL]
	612.00	612.00	[IL]EPPR
12	895.50	895.50	[IL]FFPPFK
	1023.61	1023.61	WVCFPS
13	705.44	705.44	TA[KQ]ETR
	734.37	2201.11	[IL]GSPEGPPRWP[IL]YPAAPPGGGK
	647.38	647.38	(TP)GGGMK

Table 5.1: Sequences of peptides obtained from ESP proteins by tandem MS. Sequences with strong similarity between different peptides are highlight with the same colour.

Spot	m/z	M+H	Sequence
15	1008.62	2016.24	CP[IL]PSTPD[IL][IL][IL]GNH
	1015.62	2030.24	N(TA)P[IL]PST(WC)
	1025.55	2050.10	[IL]YVF[KQ][KQ]EY[IL][IL][IL]F[IL]YTP
	1032.57	2064.14	N[IL]Y[KQ]F[IL]YVF[IL][IL][IL]Y[IL]GGK
	955.55	2864.65	(P[IL]G)(A[KQ])[IL]G[IL][IL]GWDG[IL]PST[IL][KQ][IL]([KQ][IL])[IL]([IL]Y)GK
	950.89	2850.67	[IL]SG[IL][IL]ATV[IL][IL][IL][IL]
	912.17	2734.51	YP[IL]PST(V[IL])[IL][IL][IL]VV[IL]([IL]PP)
	685.39	2054.17	[IL][IL]([KQ][KQ]P)G[IL][IL][IL]DPG([IL]Y)C[IL][IL]K
	696.73	2088.19	M[KQ]DTY[IL]EAMVR[IL]Y[IL]WR
	690.05	2068.15	P[IL][KQ][KQ][IL]G[IL][IL][IL]DGYR[IL]NAVGK
	771.14	2311.42	[IL]SVHGG[KQ][IL]YR[IL][IL][IL]ED[IL]M[IL][IL]K
17	678.58	678.58	E([IL]T)CR
	804.39	2411.17	M(AY)SRCSWWESWW[IL]ATK
	734.43	2201.29	[IL]GGTHARWY[IL](PT[IL]FVT[KQ]K
	854.68	854.68	V[KQ][IL][IL][IL][IL]R

Table 5.1 continued.

A

```
ATCGGGAGCA AACTGTNNGG TAGTCTTCAG CCTACTACAA CTTCTTGTGG CATTTTCACA
CTGTGATATT AATGACATAA CATGCAACAA GACAGTTTGT TGCGCATCAG AAGACGGTAA
AAAAGGTTCC CTATGTTGTG AGAAAGATGG TTGTCCAATT CCAAGCACTC CAGATCTTTT
GCTTGGAAAT TACCAGCGCC ATCAACGAAT GAAAATTATT TAGAGGAGTG TGCGAAAATT
TCATATACAC GCCATAAAAT ACCGGAAAAT TGGTTGTTCA AGATTAATTA CGTTTAAAAA
AATG
```

B

```
SGANCXVVFS LLQLLVAFSH CDINDITCNK TVCCASEDGK KGSLCCEKDG CPIPSTPDLL
LGNYQRHQRM KII*RSVRKF HIHAIKYRKI GCSRLITFKK M
```

Fig 5.3: (A) Sequence of the EST AW061382. (B) Frame 2 translation of the EST, with the matching peptide sequence highlighted in red.

A

```
CATCAGTTCA TATGCGTATA TTTAGTTAAG AGCACACAAG AATTATAAGG CGCATTACAT
TTCCTTGGAT TAGAACGGAC AGTATGTCCT TCTCAGTTTC GGGTTGATTG ATACATTCAA
CTTTTCCATC CTTTGTTTGA ATACACCGTT TAATAGAGGA ATGCCTTTGA ATCGGGATTT
CGTGAAGTCA TCAATGCAGG TACCATGTCC GAGACCAGTA ATCCGACCTG TGTAATTGCC
GAAGTCAATA ATACCAGTAT GGACGCTCTC TCTTCCGTGT CATTTACGAA GGGTACACAA
CAGTTCCAGG TTGTCCATGG TGGTGAGAGT GAGCAACCGC TTTTGTAGAC TTCATTATAT
GACCCCTCTC ATACGTTTCA TCGTACAATT GTAGACATTT TTTGCTGGAT CCGGTGTGGT
TATCCCATTG GCTTATCATG TGTTGACATA CGGCATAAGA ACCTCGATTT CGTGGCTTTG
TTTATCTGGG GGGGAACTTC TCTGTAACCC TTCCTCTCTG ATTTTCCAT
```

B

```
MENQRGRVTE KFPPR*TKPR NRGSYAVCQH MISQWDNHTG SSKKCLQLYD ETYERGHIMK
STKAVAHSHH HGQPGTVVYP S*MTRKRERP YWYY*LRQLH RSDYWSRTWY LH**LHEIPI
QRHSSIKRCI QTKDGKVECI NQPETEKDIL SVLIQGNVMR LIILVCS*LN IRI*TD

WKIREEGLQR SSPPDKQSHE IEVLMPYVNT **ANGITTPD PAKNVYNCTM KRMRGVI**S
LQKRLLTLTT MDNLELLCTL RK*HGRESVH TGIIDFGNYT GRITGLGHGT CIDDFTKSRF
KGIPLLNGVF KQRMEKLNVS INPKLRRTYC PF*SKEM*CA L*FLCALN*I YAYELM
```

Fig 5.4: (A) Sequence of the EST AI820476. (B) Translation frames -1 and -2 of the EST with the matching sequences highlighted in red. The peptide sequence highlighted in blue is split across the two reading frames.

```
DXAIXLPLYX ESMRGFIY*S LQKGXSXSPP WTPGSVCVPF VNDTKRERPY WYLFDNVNYT
GRITGLGHGT CIDDFTKSGFK GISS
```

Fig 5.5 Translation of the resequenced EST AI820476, showing the two parts of the peptides sequence (colour blue in figure 5.5B) in the same reading frame

Discussion

In the absence of identification of any of the protein spots from a 2-D gel of ESP by peptide mass fingerprinting, it became imperative to obtain *de novo* sequence data in order to attempt to link the proteins to ESTs from the GenBank database. Using tandem MS on a Q-ToF machine has enabled a considerable number of peptide sequences to be generated from the tryptic digests of the EST spots. Unfortunately, not all of the samples were amenable to sequencing for a variety of reasons. Firstly, some of the samples contained too much PEG contamination to identify any multiply charged peaks in the initial TOF MS scans. This probably arose from the Eppendorf tubes in which the digests were performed. Although attempts were made to clean up the samples using Millipore ZipTips, no suitable balance was found between having a high recovery of peptide molecules and retaining the contamination in the tips. So, elution with 50% acetonitrile gave poor yields of relatively clean peptides, whereas 80% acetonitrile also eluted the contaminating molecules. Secondly, some of the peptides were not sufficiently concentrated to yield clear fragmentation spectra, or simply would not fragment sufficiently well to give a spectrum with good representation of all of the y-ions. This latter case particularly applied to singly charged peptides, and this should perhaps be expected, since a single ionising proton will generally reside on the most basic group in the peptide. In the case of tryptic peptides, this is usually the C-terminal lysine or arginine residue. As a result, singly charged peptides do not have a proton associated with the peptide backbone, which would normally make one of the peptide bonds more labile and subject to preferential cleavage during fragmentation in the collision cell.

As mentioned above, several of the peptide sequences obtained are similar and this similarity occurs in two ways. Firstly, some peptides appear to be shared (or at least very similar) between distinct spots. The differences that do occur between these sequences might well arise due to errors in the peptide sequencing. Alternatively, it

could be, as suggested in Chapter 4, that the genes involved form part of a gene family that have diverged since a duplication event. Only comprehensive molecular analysis of the genes involved will elucidate which of these possibilities is correct, but the existence of these shared sequences provides additional supporting evidence for the grouping of the gels spots based on the peptide mass fingerprints presented in Chapter 4.

Secondly, repetition, or near repetition of peptide sequences from peptides with different masses within one spot suggests that the protein may have undergone some post-translational or artefactual modification. The most likely candidates for this are the oxidation of methionine residues during electrophoresis (giving methionine sulphoxide) and the alkylation of the cysteines residues in the peptides by treatment with iodoacetamide. In particular, if this alkylation was not complete peptides which contain multiple cysteine resides could give rise to a number of different peaks with different numbers of carbamidomethyl groups (recognisable as a mass shift of 57.02Da per group). Although the alkylation should be complete under the conditions employed, a recent report has suggested that this reaction may be inhibited by the thiourea present in the rehydration buffer [Galvani *et al.* 2001].

The *S. mansoni* ESTs present in the GenBank database were generated as part of the WHO Schistosome Genome Initiative from a variety of life-cycle stages. Unfortunately of the 14,039 entries for *S. mansoni* on dbEST, only 1,858 come from egg-derived cDNA libraries. To match more of the ESTs with the peptide sequences would require a major increase in the number of sequences generated from egg-derived cDNA libraries. However, one inference that can be drawn is that since the peptide sequences did not match any of the ESTs from other life-cycle stages, it is likely that the ESP proteins are egg-specific. Whilst two distinct ESTs were identified as containing peptide sequences from the ESP proteins, it is still disappointing that more of the peptide sequence were not found in the database, especially since, as

demonstrated by the biosynthetic labelling of ESP in Chapter 2, the secretions appear to form the major synthetic activity of the mature egg. Their mRNAs should therefore be abundant and well represented in the cDNA library. One possible explanation for this is that the mRNAs of at least some of the ESP proteins are, for some reason, difficult to clone. Suspicion that this is the case is aroused by the fact that the peptide sequences were found on the negative strand of a cDNA from a directional library. Since the library was constructed using oligo-dT priming, and then inserted into the vector after digestion with two different restriction enzymes (EcoRI and XhoI), all of the coding strands should be on the positive strand when sequenced with the M13R primer. The presence of the peptide sequences on the negative strand suggests that the mRNA contains an internal A-rich region, causing the primer to anneal internally and insert in the wrong orientation into the vector.

Whatever the case, to obtain further sequence information from the ESP proteins will require a number of different approaches. Firstly, the full length sequences of the identified ESTs could be obtained using e.g., 5'-RACE. Secondly, the peptide sequences which did not match to any ESTs could be used to design degenerate probes by hybridisation studies, or primers for degenerate PCR to identify or amplify clones from cDNA libraries which encode them. Finally, further attempts would be made to generate further peptide sequence information from the ESP proteins. However, the ESTs identified provide a good starting point for the detailed molecular characterisation of the ESP proteins.

Chapter 6: Concluding Discussion

This chapter summarises the work undertaken in this project, discusses alternative approaches that could have been taken and provides some suggestions about future work that might increase the knowledge of the *S. mansoni* egg secretions.

The initial rationale for this project was that by studying the proteins secreted by the egg, understanding could be gained about two major aspects of schistosomiasis: the means by which the eggs, laid in the blood vessels of the mesenteric system, migrate through the tissues of the gut wall to be voided in the faeces, and the coincidental formation of granulomatous lesions around those eggs which fail to exit from the host via this route. It was generally believed that these secretions were produced by the miracidium within the egg, and as a result, much effort has previously been dedicated to the identification of egg proteins, usually purified from a homogenate of whole eggs, called SEA. However, an ultrastructural study had identified that, in common with other trematodes, the miracidium of *S. mansoni* is completely surrounded within the egg by a syncytial envelope. This envelope develops with the miracidium and, at maturity, would present a barrier to passive diffusion between the miracidium and the external milieu. If the secretions were truly produced by the miracidium, they would therefore have to be taken up by, and transported across, this envelope in order to be released from the egg.

In order to clarify this situation, the ultrastructure of the egg was revisited (Chapter 2). Rather than using the approach taken in the previous study of using slam freezing and freeze substitution to allow the embedding resin to penetrate the egg shell, each egg was individually punctured using an entomology pin, and this resulted in better preservation of the egg contents than observed in the earlier study. Examination of eggs at different stages of development suggested that the envelope originated in the same way as in other trematodes.

That is, cells which were morphologically distinct from the vitelline cells migrated to the inner surface of the shell where they flattened, spread out and fused to form the envelope. As the miracidium developed, this structure became thicker and a layer of acellular material appeared between it and the shell. In the fully mature egg, the most striking feature of the envelope was the large amount of rough endoplasmic reticulum that it contained, which was immediately suggestive of protein synthesis. Measurements of the amount of protein appearing in the culture medium used to maintain immature and mature eggs *in vitro* showed that no secretion occurred until the envelope was fully developed.

Analysis of the proteins present in the culture medium using SDS-PAGE showed a simple pattern of six bands, which were largely absent from, or only minor components of, SEA. However, when gels of biosynthetically SEA and ESP were visualised using the Phosphorimager, the ESP components were clearly visible in SEA. So, although they form minor components in terms of the amount of protein present, the high incorporation of label into these proteins during the culture period suggests that production of ESP is a major biosynthetic activity of the mature egg. Raising antisera enabled the secretions to be localised with the egg by immunocytochemistry and this confirmed that it is the envelope, and not the miracidium, that is the site of synthesis of ESP. Additionally, it was possible to use this antiserum in a circumoval precipitation reaction, which clearly demonstrated the secretion of ESP from viable eggs.

Additional electrophoretic analyses reveal that there are two proteases in ESP. This correlates well with the hypothesised functions of ESP which are to disrupt the host tissue in order to facilitate the egress of the eggs. However, this process is clearly not due to the action of secreted hydrolytic enzymes alone, since it has been demonstrated that the host immune response is also required for significant numbers of eggs to be excreted in the faeces [Doenhoff *et al.*, 1981; Karanja *at al.*

1997]. In order to understand these processes completely, it will be necessary to fully examine the immune responses that are provoked by the egg secretions. In this project, it has been possible to demonstrate that, after intraperitoneal injection with viable eggs, IgG antibody responses are elicited by ESP and that these appear to be distinct from the responses to SEA, rising earlier presumably because the egg secretions will be seen as soon as mature eggs are present in the host, whilst the anti-SEA response will not become evident until at least some of the eggs die. Further, differences in the composition of the granulomas around Sepharose beads coupled with SEA and ESP suggest that the recruitments of granulocytes to the lesions occurs more with ESP (i.e., when the egg is still alive) and that fibrosis occurs principally after the egg has died and released its contents.

The above evidence all suggests that it is the egg secretions that should be characterised in order to understand the processes of egg escape and initiation of granuloma formation in schistosomiasis. Traditionally, the first approach that might have been taken would use antisera raised against ESP to screen cDNA expression libraries. This was attempted by another member of the laboratory (Dr. Bindiya Shah), but unfortunately, was unsuccessful, and this is probably due to the antibody response being directly principally against the glycans on ESP rather than the protein backbone. Indeed, it has recently been demonstrated that during an experimental infection in chimpanzees, the antibody recognition of ESP was entirely due to the glycans, and could be eliminated by periodate oxidation of the blotted proteins (Eberl *et al.*, 2001). It was, therefore, decided that the best approach to obtain information about the ESP would be to use the techniques of proteomics, which were just becoming available during the period of this project. It was therefore necessary to establish these techniques in the laboratory (Chapter 3), so that they could then be applied to characterise as many of the ESP proteins as possible and, hopefully lead

to linking the spots from the 2-D gel of ESP to the sequences of their encoding genes.

As with the 1-D gel of ESP, two-dimensional electrophoresis revealed a relatively simple pattern, with only 21 spots visible on the gel, compared with many hundreds of spots for SEA (Chapter 4). Repeating the zymogram in two dimensions showed that the proteases seen in the 1-D zymogram could be detected, and that they were both present among the acidic components of ESP. Additionally, it was possible to demonstrate that at least some of the ESP proteins are glycosylated, and this confirms that they are truly secreted proteins, since they must have passed through both the endoplasmic reticulum and the Golgi apparatus in order to acquire their carbohydrates.

Each of the spots from the 2-D gel was then excised, digested with trypsin and subjected to two different mass spectrometric techniques: peptide mass fingerprinting and peptide sequencing. Whilst peptide mass fingerprinting provides a convenient method of rapidly screening a large number of proteins, it does have some limitations and the main one is its inability to yield identification of homologous proteins from different species. Each point mutation in a gene sequence which leads to a change in a single amino acid will affect the mass of one of the tryptic peptides of the encoded protein. Since peptide mass fingerprinting identifies proteins using as few as 4-6 peptide masses, any mutations which change the mass of even a single peptide might lead to the homologous protein not being regarded as a significant match. The identification of the egg secreted proteins by peptide mass fingerprinting, therefore, relies on their full length sequences being present in one of the online databases (e.g., the GenBank non-redundant database, nr). Thus, with only ~400 *S. mansoni* genes represented in this database, it is not surprising that no identifications were obtained. This is not simply a failure of the technique, since proteins from the 2-D gel of SEA were identified with high confidence. However, the attempt has demonstrated that the best studied egg antigens (e.g., p40) are not in

ESP and additionally that the proteases present in the ESP preparation are unlikely to be any of the previously characterised schistosome proteases (e.g., elastase or cathepsin B or L).

Another, unexpected result from the PMF analysis was the degree of similarity noted between the fingerprints of different ESP spots. This enabled a tree of relatedness to be constructed on the basis of these shared peptides using maximum parsimony methods. The most surprising finding from this analysis was that the clusters of proteins generated correlate with the grouping of the spots on the 2-D gel, with spots 1-6 (the most basic and largest components) all forming a single branch of the tree, and with spots 8-12 (moderately basic, smaller and more minor components) forming another, with the remaining acidic spots falling into four further groups.

Although peptide mass fingerprinting did not identify any of the ESP proteins, it was possible to link at least some of the spots to partial DNA sequences by generating peptide sequences using tandem MS (Chapter 5). The failure to obtain links to DNA sequences for all the peptide sequences generated is, again, probably due to the lack of sequence information. This is perhaps surprising, since the number of sequences in the EST database for *S. mansoni* is reasonably large, although sequences from egg-derived libraries do not form a major proportion of this total (1,858 sequences). Whilst there are some 14,000 ESTs in the database, cluster analysis performed by The Institute for Genome Research (TIGR), has shown that these ESTs represent ~7,000 distinct coding sequences, which is only 30-40% of the estimate 15-20,000 genes. Even so, more matches might have been expected on the basis of probability, and this is suggestive that the egg secretions are truly egg specific since they did not match to any of the more abundant adult and cercarial sequences.

The lack of sequence information may well be rectified in the next few months as two major EST sequencing projects (funded by the

states of Sao Paolo and Minas Gerais in Brasil) are currently underway, and together, these will generate an additional 200,000 ESTs. Given that the upper range of estimates of gene number for *S. mansoni* is 20,000, this should greatly increase the chance that any given gene is represented in the EST database. Due to the large size of the genome (270Mb) and the fact that is contains up to 60% repetitive DNA [Simpson, Sher & McCutchan, 1982], full genomic sequencing of *S. mansoni* is still some way off, but some of the first steps have already been taken, such as the construction of a BAC library by the Institut Pasteur, Lille [Le Paslier *et al.*, 2000] and the sequencing of almost 15,000 BAC clone ends. As these sequences become available, it will be possible to re-interrogate the databases using both the PMFs and the peptide sequences.

In the interim, there are a number of possible approaches that could be taken to obtain further sequence information directly. Firstly, the EST sequences which were retrieved by BLAST searching using the peptide sequences could be extended by using a technique such as 5' RACE (rapid amplification of cDNA ends), which would, in theory, enable the full coding sequences of the genes to be obtained. In order to have total confidence that these identifications were correct, translation of the sequences would enable them to be correlated with the results of the peptide mass fingerprinting. Secondly, for the peptide sequences which did not match to any EST, the best approach would probably involve the design of degenerate oligonucleotides for use either as probes in hybridisation studies, or as primers in degenerate PCR, on cDNA libraries. Identified clones could then be sequenced and, if necessary, the sequence extended, again using 5' RACE. Finally, there are a number of the spots which did not yield any peptide sequence data and therefore, peptide sequencing should be attempted again, but either with modifications to the protocols for performing the tryptic digest, or with additional steps in that protocol to eliminate the PEG contamination, e.g., by use of a Q-ToF machine with a capillary liquid

chromatography front end. This would enable the peptides (and the contamination) to be retained on a reverse phase chromatography column. A gradient of increasing acetonitrile concentration should then enable the peptides to be selectively eluted directly from the column into the mass spectrometer. Combined with the greater sensitivity of the new generation of Q-ToF machines would almost certainly allow many more peptides to be observed and sequenced.

In conclusion, I believe that this project has made major progress in the identification and characterisation of the proteins secreted by the *S. mansoni* egg, and that, with the imminent increase in the amount of sequence data, the remaining egg secretions will be identified in the near future.

References

ASAHI, H., OSMAN, A., COOK, R.M., LOVERDE, P.T. & STADECKER, M.J. (2000) *Schistosoma mansoni*: Phosphoenolpyruvate carboxykinase, a novel egg antigen: immunological properties of the recombinant protein and identification of a T-cell epitope. *Infection and Immunity* **68**(6): 3385-3393.

ASCH, H.L. & DRESDEN M.H. (1979) Acidic thiol proteinases activity of *Schistosoma mansoni* egg extracts. *Journal of Parasitology* **65**(4): 543-549.

BJELLQVIST, B., EK, K., RIGHETTI, P.G., GIANAZZA, E., GÖRG, A., WESTERMEIER, R. & POSTEL, W. (1982) Isoelectric focussing in immobilised pH gradients: principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods* **6**: 317-339.

BONETTO, V., BERGMAN, A.C., JORNVALL, H. & SILLARD, R. (1997) C-terminal sequence analysis of peptides and proteins using carboxypeptidases and mass spectrometry after derivatisation of Lys and Cys residues. *Analytical Chemistry* **69**(7): 1315-1319.

BOROS, D.L. & WARREN, K.S. (1970) Delayed hypersensitivity-type granuloma formation and dermal reaction induced and elicited by a soluble factor isolated from *Schistosoma mansoni* eggs. *Journal of Experimental Medicine* **132**(3): 488-507.

CHEN P-D. (1937) The germ cell cycle in the trematode *Paragonimus kellicotti* (Ward). *Transactions of the American Microscopical Society.* **56**, 208-236.

CHENSUE, S. W. & D. L. BOROS (1979). Population dynamics of T and B lymphocytes in the lymphoid organs, circulation, and granulomas of mice infected with Schistosoma mansoni. *American Journal of Tropical Medicine and Hygiene* **28**(2): 291-9.

CHIEN, B.M., MICHAEL, S.M. & LUBMAN, D.M. (1993) Enhancement of resolution in matrix-assisted laser desorption using an ion-trap storage reflectron time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry* **7**(9): 837-843.

COLBURN, A.W., GIANNAKOPULOS, A.E., DERRICK, P.J. & VON RAUMER, M. (2000) A quadratic-field reflectron time-of-flight mass spectrometer incorporating intermediate temporal focusing. *European Journal of Mass Spectrometry* **6**(6): 523-530.

COTTRELL, J. S. (1994). Protein identification by peptide mass fingerprinting. *Peptide Research* **7**(3): 115-24.

CRABTREE, J.E. & WILSON, R.A (1985) *Schistosoma mansoni:* an ultrastructural examination of skin migration in the hamster cheek pouch. *Parasitology* **91**(1): 111-120.

DOENHOFF, M.J., PEARSON, S., DUNNE, D.W., BICKLE, Q., LUCAS, S., BAIN, J., MUSALLAM, R & HASSOUNAH, O. (1981) Immunological control of hepatotoxicity and parasite egg excretion in *Schistosoma mansoni* infections: stage specificity of the reactivity of immune serum in T-cell deprived mice. *Transaction of the Royal Society of Tropical Medicine and Hygiene* **75, 41-53.**

DOENHOFF, M.J., HASSOUNAH, O., MURARE, H., BAIN, J., LUCAS, S. (1986) The schistosome egg granuloma: immunopathology in the cause of host protection or parasite survival? *Transactions of the Royal Society of Tropical Medicine and Hygiene* **80**(4): 503-514.

DORSEY, C.H. & STIREWALT, M.A. (1971) *Schistosoma mansoni*: fine structure of the cercarial acetabular glands. *Experimental Parasitology* **30**(2): 199-214.

DUNNE, D.W., JONES, F.M & DOENHOFF, M.J. (1991) The purification, characterisation, serological activity and hepatotoxic properties of two

cationic glycoproteins (α_1 and ω_1) from *Schistosoma mansoni* eggs. *Parasitology* **103**, 225-236.

EBERL, M., LANGERMANS, J.A., VERVENNE, R.A, NYAME, A.K., CUMMINGS, R.D., THOMAS, A.W., COULSON, P.S. & WILSON, R.A. (2001). Antibodies to glycans dominate the host response to schistosome larvae and eggs: is their role protective or subversive? *Journal of Infectious Disease* **183**(8): 1238-1247.

ECKHARDT A.E., HAYES, C.E. & GOLDSTEIN, I.J. (1976) A sensitive fluorescent method for the detection of glycoproteins in polyacrylamide gels. *Analytical Biochemistry* **73**: 192-197.

EDMAN, P. AND BEGG, G. (1967). A Protein Sequenator. *European Journal of Biochemistry*. **1**: 80-91.

FERRIGE, A.G., SEDDON, M.J. & JARVIS, S. (1991) Maximum-entropy deconvolution in electrospray mass spectrometry. *Rapid Communications in Mass Spectrometry* **5**(8): 374-377.

FINKELMAN, F.D. & MORRIS, S.C. (1999) Development of an assay to measure *in vivo* cytokine production the mouse. *International Immunology* **11**(11): 1811-1818.

FRANCO, G.R., ADAMS, M.D., SOARES, M.B., SIMPSON, A.J., VENTER, J.C. & PENA, S.D. (1995). Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene* **152**(2): 141-7.

GALVANI, M., ROVATTI, L, HAMDAN, M., HERBERT, B. & RIGHETTI, P.G. (2001) Protein alkylation in the presence/absence of thiourea in proteome analysis: A matrix assisted laser desorption/ionisation-time of flight-mass spectrometry investigation. *Electrophoresis* **22**: 2066-2074.

HANG, L.M., WARREN, K.S. AND BOROS, D.L. (1974) *Schistosoma mansoni*: antigenic secretions and the etiology of egg granulomas in mice. *Experimental Parasitology* **35**(2): 288-298.

HAYNES, P.A. & YATES, J.R. (2000) Proteome profiling – pitfalls and progress. *Yeast* **17**: 81-87.

HOCHSTRASSER, D.F., PATCHORNIK, A. & MERRIL, C.R. (1988). Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining. *Analytical Biochemistry* **173**(2): 412-23.

ISHII Y. (1934) Studies on the development of *Fasciolopsis buski*. *Journal of the Medical Association of Formosa* **33**: 349-412.

KARANJA, D.M., COLLEY, D.G., NAHLEN, B.L., OUMA, J.H & SECOR, W.E. (1997) Studies on schistosomiasis in Western Kenya: I. Evidence for immune facilitated secretion of schistosome eggs from patients with *Schistosoma mansoni*. *American Journal of Tropical Medicine and Hygiene* **56**(5): 515-521.

KARAS, M. & HILLENKAMP, F. (1988) Laser desorption ionisation of proteins with molecular masses exceeding 10000 Daltons. *Analytical Chemistry* **60**(20): 2299-2301.

KAUFMANN, R., SPENGLER, B. & LUTZENKIRCHEN, F. (1993) Mass spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass spectrometer using matrix-assisted laser desorption ionization. *Rapid Communications in Mass Spectrometry* **7**(10): 902-910.

KEOUGH, T.W., YOUNGQUIST, R.S. & LACEY, M.P. (1999) A method for high sensitivity sequencing using postsource decay matrix assisted laser desorption ionisation mass spectrometry. *Proceedings of the National Academy of Science of the USA* **96**: 7131-7136.

KLOSE, J. (1975) *Humangenetik* **26**: 231-243.

LE PASLIER, M.C., PIERCE, R.J., MERLIN, F., HIRAI, H., WU, W.J., WILLIAMS, D.J., JOHNSTON, D.A., LOVERDE, P.T. & LE PASLIER, D. (2000) Construction and characterisation of a *Schistosoma mansoni* bacterial artificial chromosome library. *Genomics* **65**(2): 87-94.

LOCKWOOD, B. C., NORTH, M. J., SCOTT, K. I., BREMNER, A. F. & COOMBS, G. H. (1987) The use of a highly sensitive electrophoretic method to compare the proteinases of trichomonads. *Molecular and Biochemical Parasitology* **24**, 89-95.

MADDISON, D.R., SWOFFORD, D.L. & MADDISON, W. (1997) NEXUS: An extensible file format for systematic information. *Systematic Biology* **46**(4): 590-621.

MANN, M. & WILM, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry* **66**(24): 4390-4399.

MCHUGH, S.M., COULSON, P.S. & WILSON, R.A. (1987) Pathologically induced alterations in the dimensions of the hepatic portal vasculature of mice infected with *Schistosoma mansoni*. *Parasitology* **94**(1): 69-80.

MEDZIHRADSZKY K.F., CAMPBELL, J.M., BALDWIN, M.A., FALICK, A.M., JUHASZ, P., VESTAL, M.L. & BURLINGAME, A.L. (2000) The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Analytical Chemistry* **72**(3): 552-558.

MICHEL, A., GHONEIM, H., RESTO, M., KLINKERT, M.Q. & KUNZ, W. (1995) Sequence, characterisation and localization of a cysteine protease cathepsin L in *Schistosoma mansoni*. *Molecular and Biochemical Parasitology* **73**(1-2): 7-18.

MORRIS, H.R., PAXTON, T., DELL, A., LANGHORN, J., BERG, M., BORDOLI, R.S., HOYES, J. & BATEMAN, R.H. (1996) High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry* **10**(8): 889-896.

NEILL, P.J., SMITH, J.H., DOUGHTY, B.L. & KEMP, M. (1988) The ultrastructure of the schistosome egg. *American Journal of Tropical Medicine and Hygiene* **39**(1): 52-65.

NENE, V., DUNNE, D.W., JOHNSON, K.S., TAYLOR, D.W., CORDINGLEY, J.S. (1986) Sequence and expression of a major egg antigen from *Schistosoma mansoni*. Homologies to heat shock proteins and alpha-crystallins. *Molecular and Biochemical Parasitology*. **21**(2): 179-188.

NEWPORT, G., MCKERROW, J., HEDSTROM, R., CULPEPPER, J., MCGARRIGLE, L. & AGABIAN, N. (1987) Schistosome elastases: biological importance, structure, function and stage-specific expression. *Biochemical Society Symposia* **53**: 115-121.

O'FARRELL, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* **250**: 4007-.

ORTMANN W. (1908) Zur Embryonalentwicklung des Leberegels (*Fasciola hepatica*). *Zoologische Jahrbucher, Abteilung für Anatomie und Ontogie* **26**, 255-292.

PETERSON, W.P. & VON LICHTENBERG, F. (1965) Studies on granuloma formation: IV. In vivo antigenicity of schistosome egg antigen in lung tissue. *Journal of Immunology* **95**(5): 959-965.

REES G. (1940) Studies on the germ cell cycle of the digenetic trematode *Parorchis acanthus* (Nicoll). *Parasitology* **32**: 372-391.

ROLLINSON, D. & SIMPSON, A.J.G. (1988) *The Biology of Schistosomes: from Genes to Latrines.* Academic Press.

SANCHEZ, J. C., HOCHSTRASSER, D. & RABILLOUD, T. (1999). In-gel sample rehydration of immobilized pH gradient. *Methods in Molecular Biology* **112**: 221-5.

SANTONI, V., MOLLOY, M. & RABILLOUD, T. (2000) Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**(6): 1054-1070.

SCHAGGER, H. & VON JAGOW, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry* **166**(2): 368-379.

SCHUBMANN W. (1905) Über die Eibildung und Embroyonalentwicklung von *Fasciola hepatica* L. (*Distomum hepaticum* Retz.). *Zoologische Jahrbucher Abteilung fur Anatomie und Ontogie* **21:** 571-606.

SIMPSON, A.J., SHER, A. & MCCUTCHAN, T.F. (1982) The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequence. *Molecular and Biochemical Parasitology* **6**(2): 125-137.

STADECKER, M. J. (1994). The shrinking schistosomal egg granuloma: how accessory cells control T cell-mediated pathology. *Experimental Parasitology* **79**(2): 198-201.

SWOFFORD, D. L. (1993). PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1. Washington, DC: Smithsonian Institution.

UNLU, M., MORGAN, M.E. & MINDEN, J.S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**(11): 2071-2077.

VAN MARCK, E.A., STOCKER, S., GRIMAUD, J.A., KESTENS, L., GIGASE, P.L. & DEELDER, A.M. (1980) The implantation of Sepharose

beads in mouse livers as an aid to the study of hepatic schistosomal fibrosis. *Experientia* **36**(9): 1116-1118.

VESTAL, M.L., JUHASZ, P. & MARTIN, S.A. (1995) Delayed extraction matrix-assisted laser-desorption time-of-flight mass spectroscopy. *Rapid Communications in Mass Spectrometry* **9**(11): 1044-1050.

VON LICHTENBERG F. & RASLAVICIUS P. (1967) Host response to eggs of *Schistosoma mansoni*. V. Reaction to purified miracidia and egg shells and to viable and heat-killed whole eggs. *Laboratory Investigation* **16**: 892-904.

VON LICHTENBERG, F. (1988) Consequences of infection with schistosomes. In: Rollinson, D. & Simpson, A.J.G. (1988) The biology of schistosomes: from genes to latrines.

WALDROP, F.S. & PUCHTLER, H. (1982) Light microscopic distinction of collagens in hepatic cirrhosis. *Histology* **74**(4): 487-491.

WARREN, K.S., DOMINGO, E.O. & COWAN, R.B. (1967) Granuloma formation around schistosome eggs as a manifestation of delayed hypersensitivity. *American Journal of Pathology* **51**(5): 735-756.

WASINGER, V.C., CORDWELL, S.J., CERPA-POLJAK, A., YAN, J.X., GOOLEY, A.A., WILKINS, M.R., DUNCAN, M.W., HARRIS, R., WILLIAMS, K.L., HUMPHERY-SMITH, I. (1995) Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium* *Electrophoresis* **16**: 1090-1094.

WHO (World Health Organisation) 1998. Schistosomiasis (Factsheet 115). http://www.who.int/inf-fs/en/fact115.html.

WILLIAMS, D.L., ASAHI, H., BOTKIN, D.J. & STADECKER, M.J. (2001) Schistosome infection stimulates host $CD4^+$ T helper cell and B-cell responses against a novel egg antigen thioredoxin peroxidase. *Infection and Immunity* **69**(2): 1134-1141.

WILSON, R.A. (1988) Cercariae to Liver Worms: Development and Migration in the Mammalian Host. In Rollinson, D. & Simpson, A.J.G. (1988) *The Biology of Schistosomes: from Genes to Latrines.*

WILSON, R.A. & COULSON, P.S. (1986) *Schistosoma mansoni*: dynamics of migration through the vascular system of the mouse. *Parasitology* **92**(1): 83-100.

Appendix 1: Peptide Mass Fingerprints of ESP Proteins

The following pages contain the complete set of peptide mass fingerprints from all 21 of the spots on the 2-D gel presented in chapter 4.

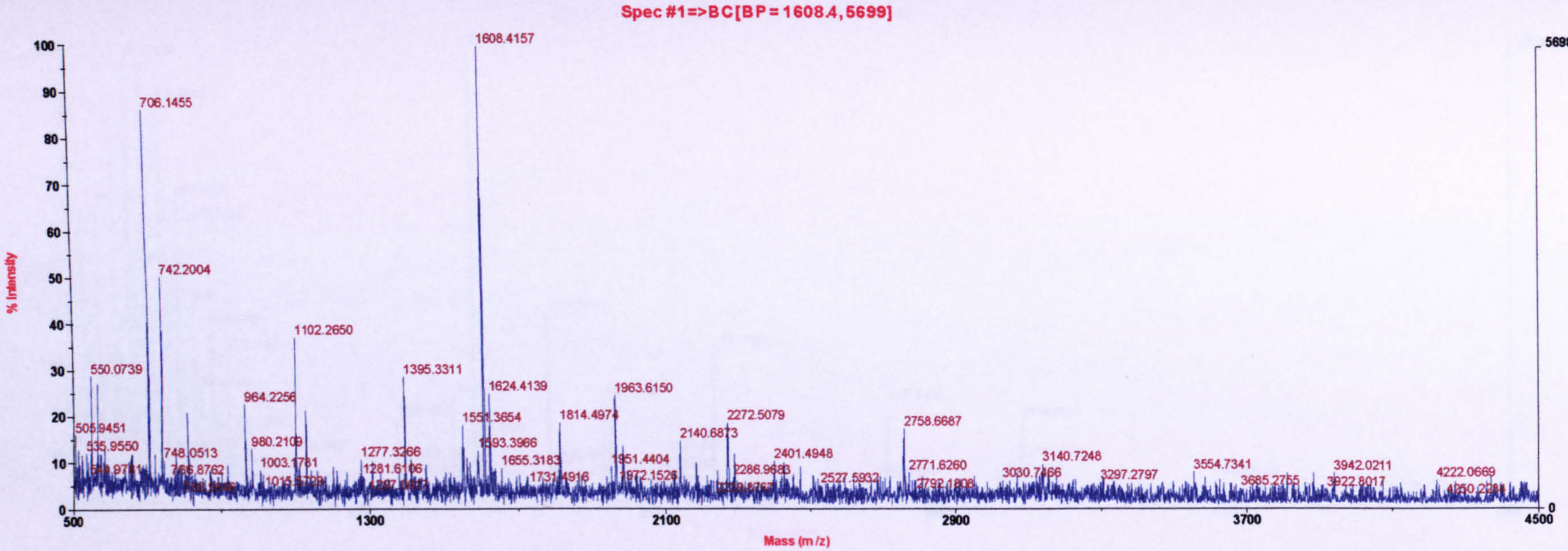


Fig A.1 Peptide mass fingerprint of spot 1

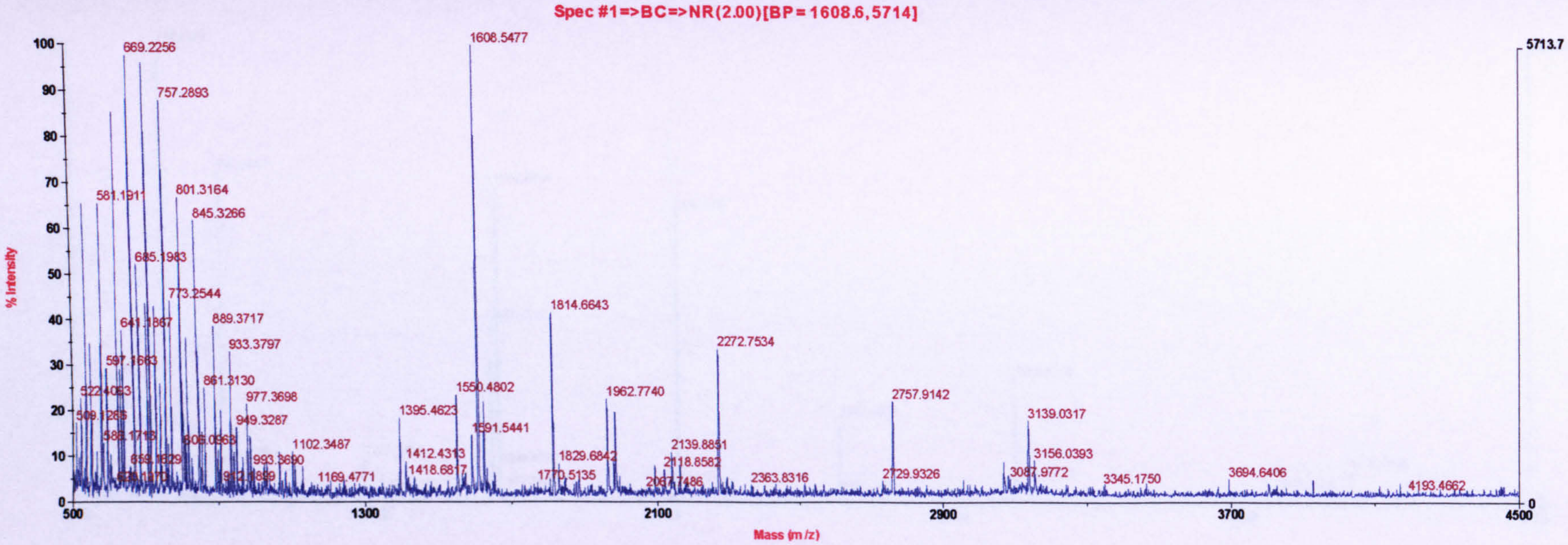


Fig A.2 Peptide mass fingerprint of spot 2

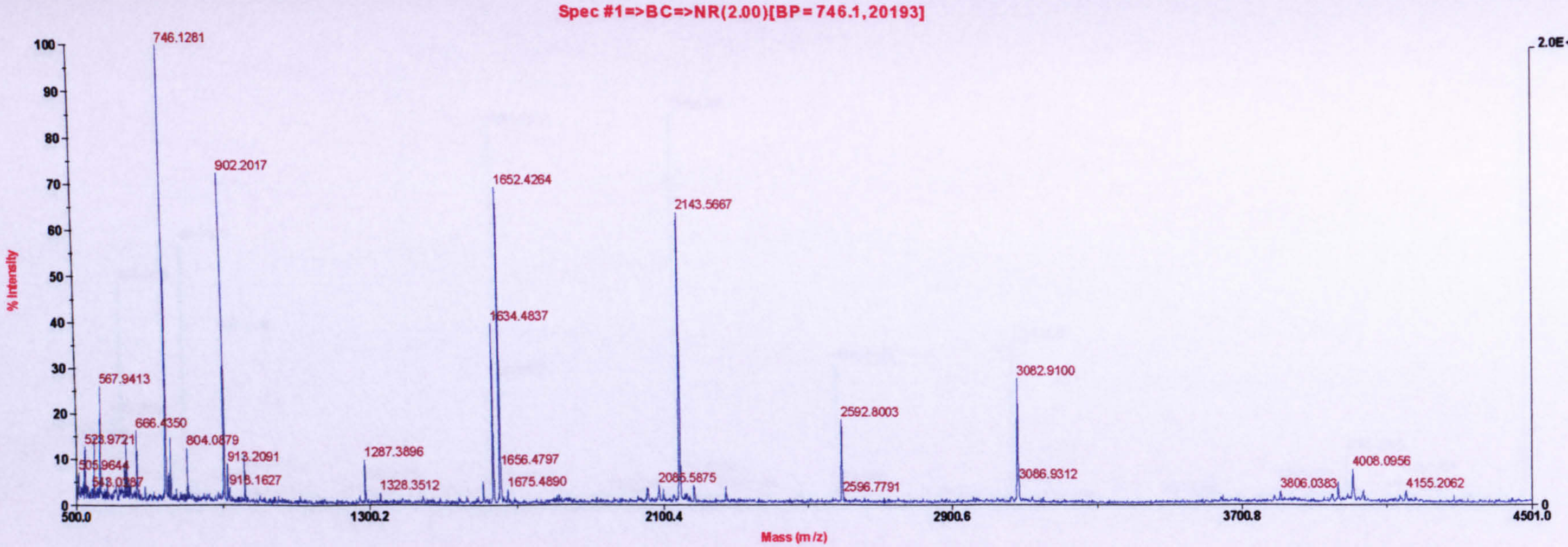


Fig A.3 Peptide mass fingerprint of spot 3

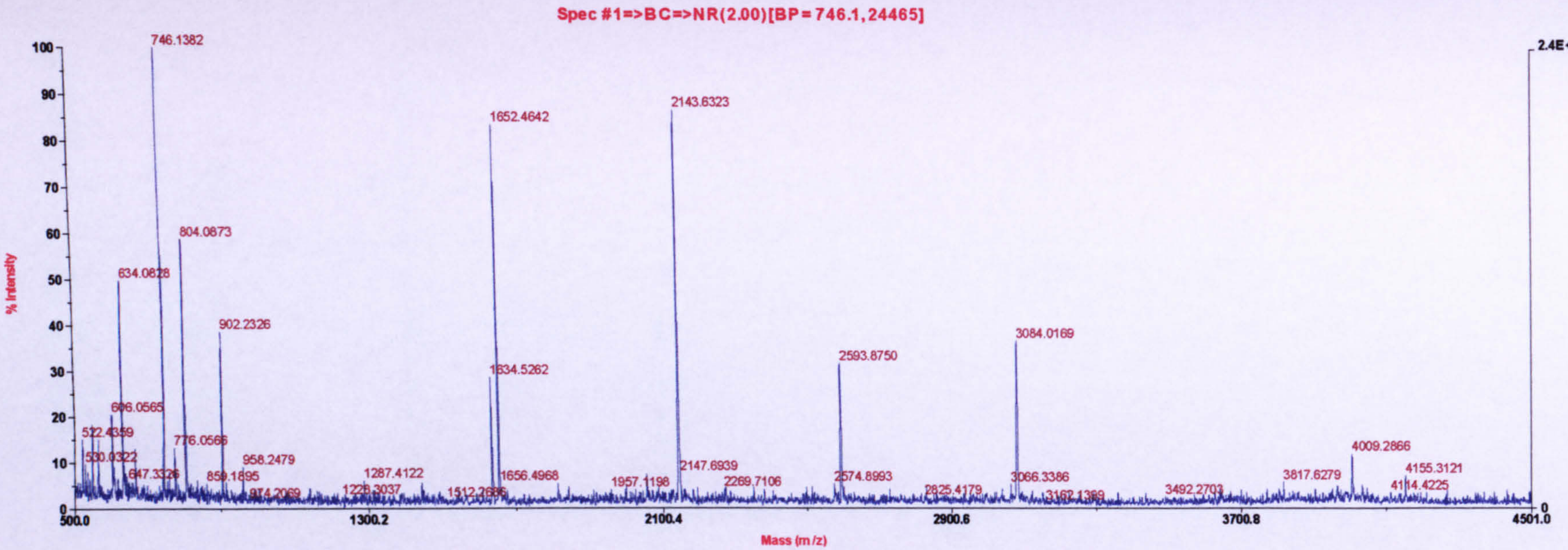


Fig A.4 Peptide mass fingerprint of spot 4

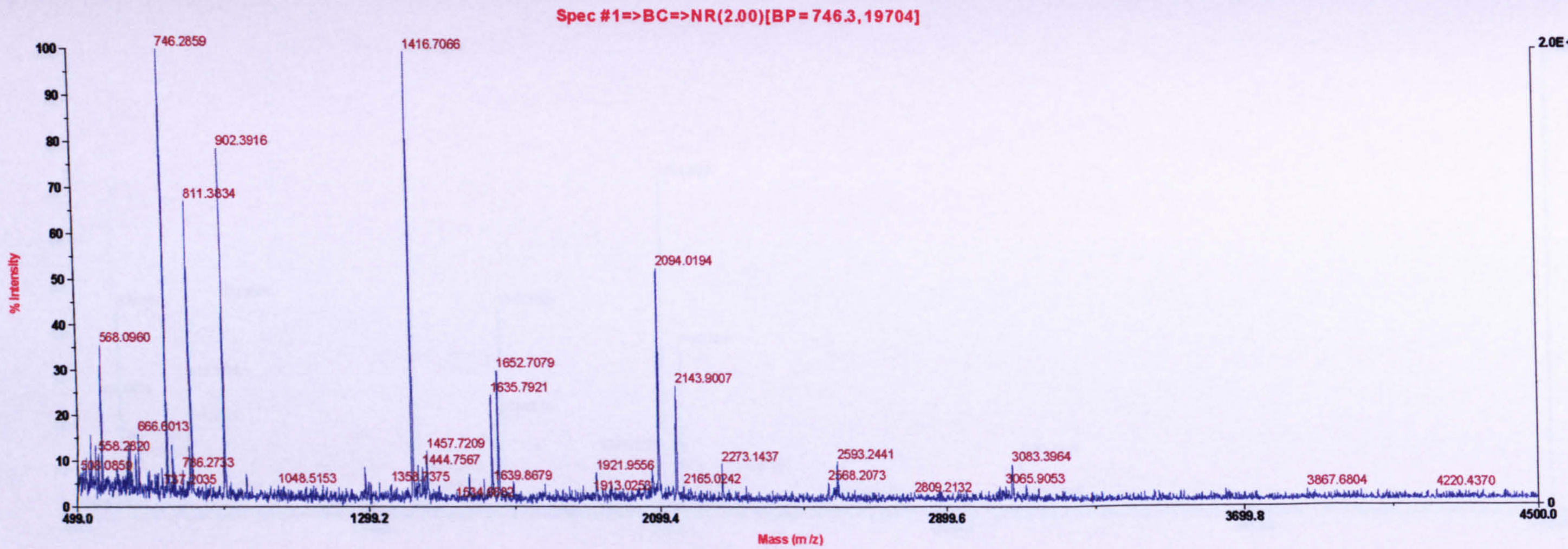


Fig. A.5 Peptide mass fingerprint of spot 5

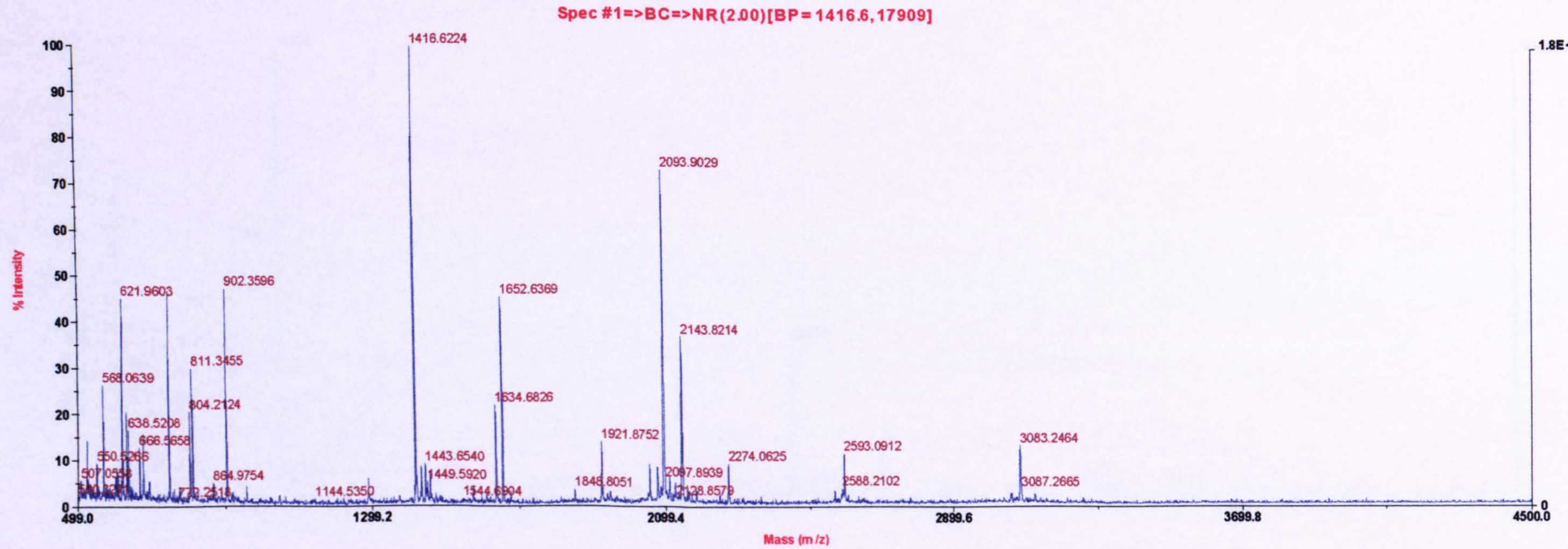


Fig A.6 Peptide mass fingerprint of spot 6

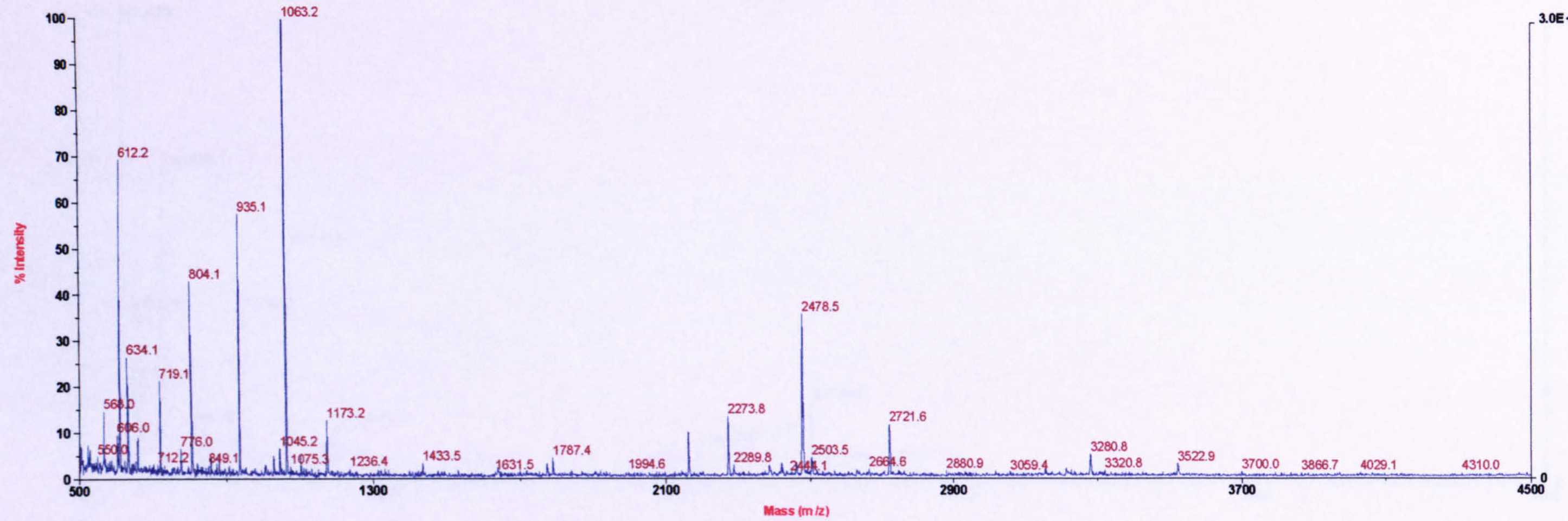


Fig A.7 Peptide mass fingerprint of spot 8.

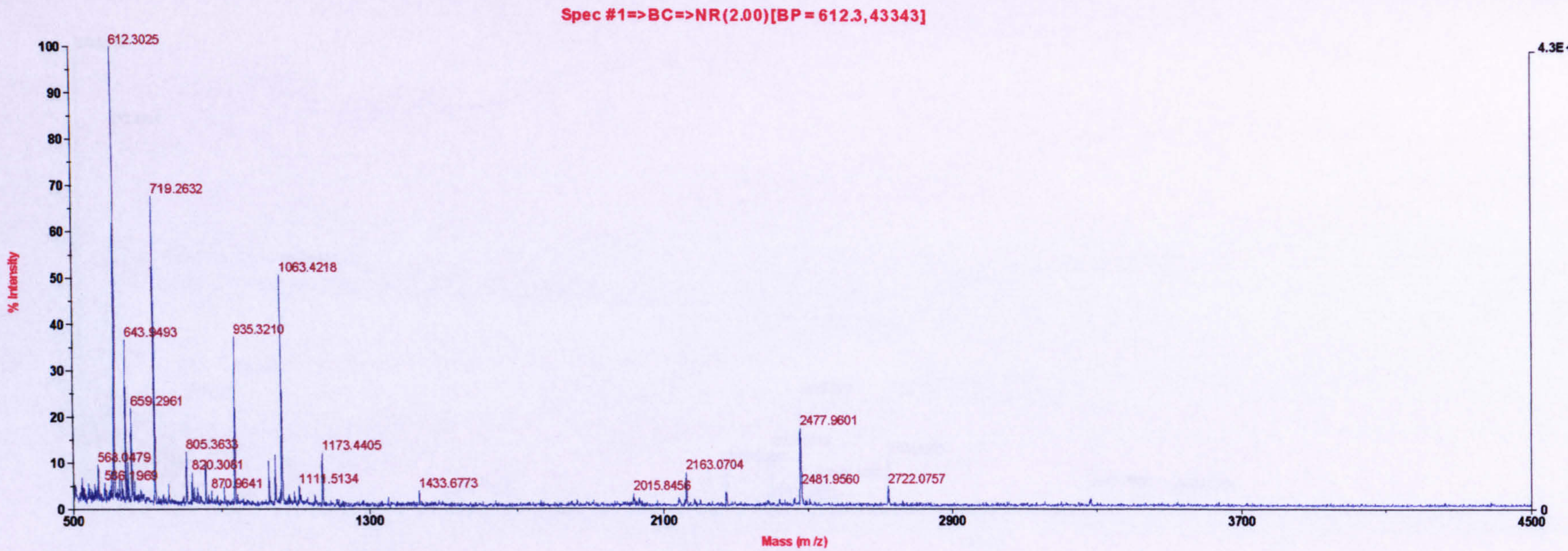


Fig A.8 Peptide mass fingerprint of spot 9.

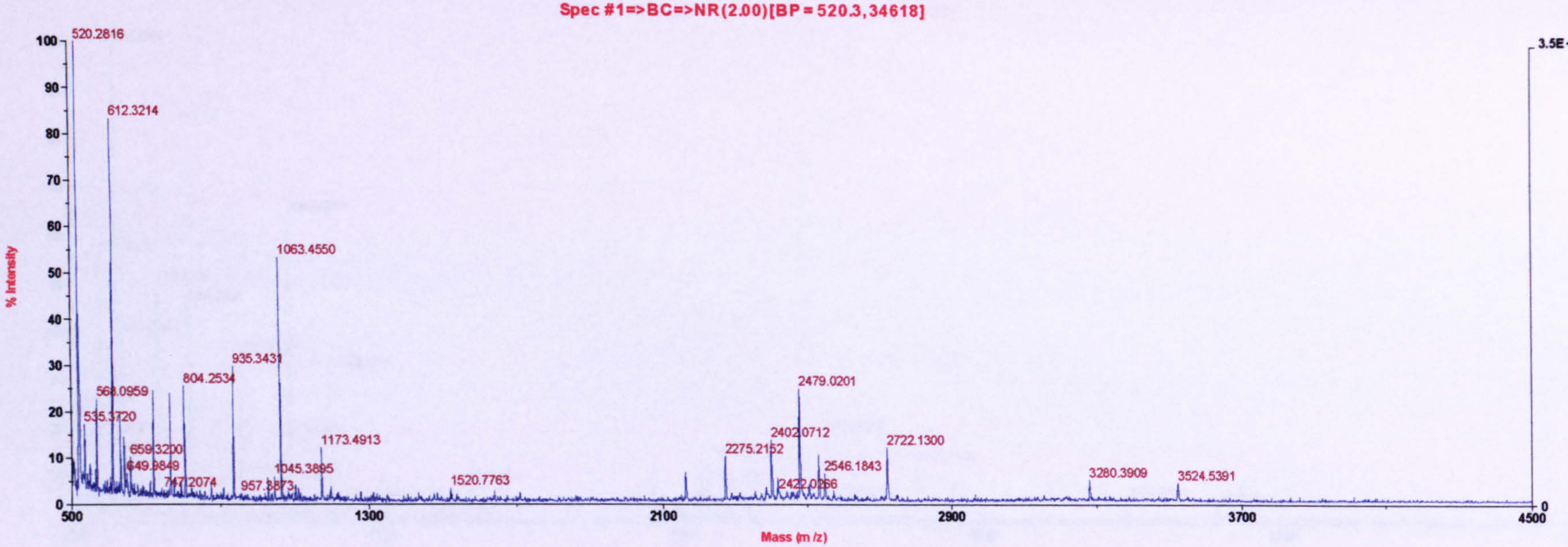


Fig A.9 Peptide mass fingerprint of spot 10a.

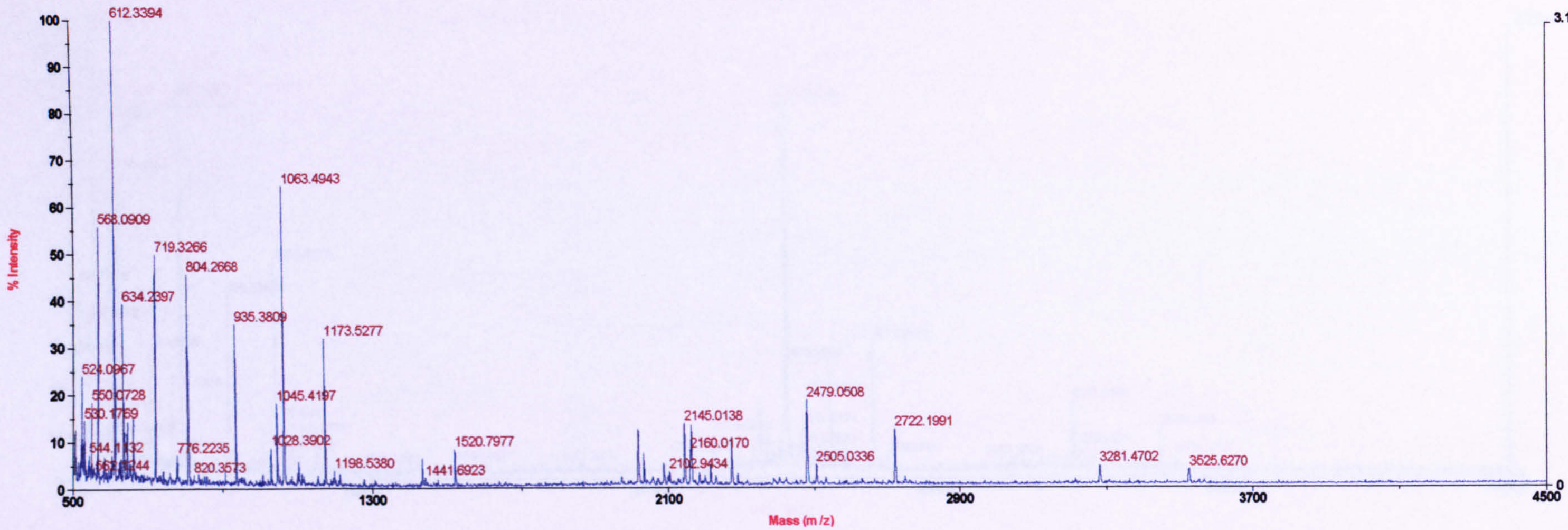


Fig A.10 Peptide mass fingerprint of spot 10b

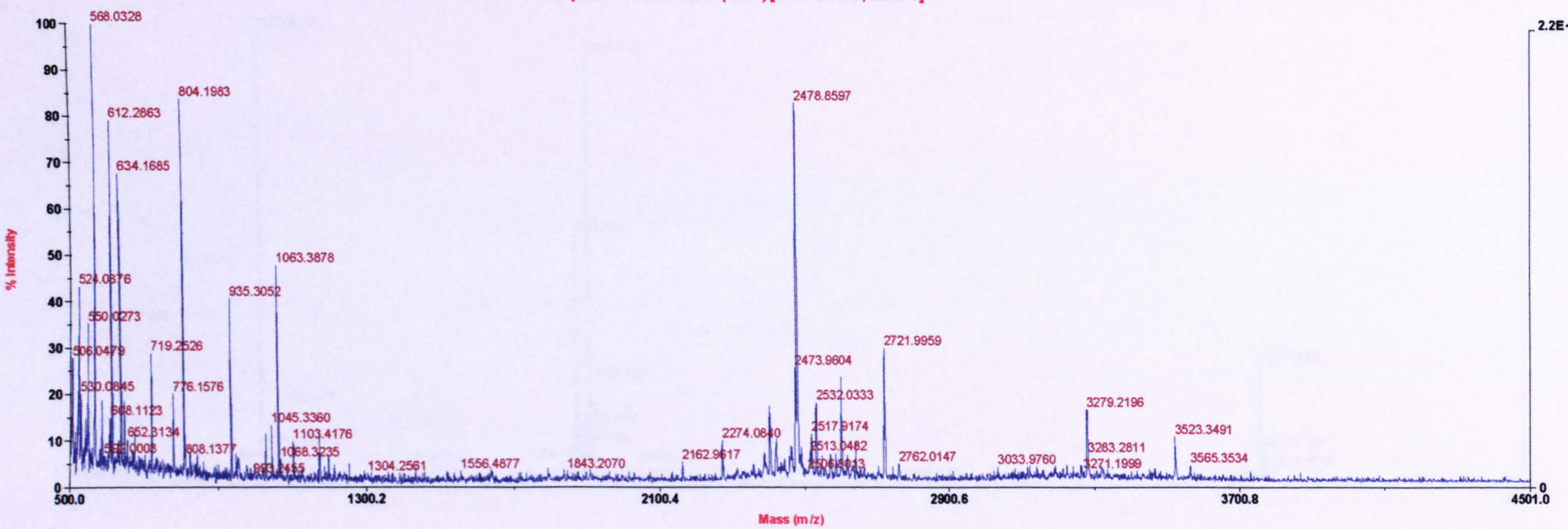


Fig A.11 Peptide mass fingerprint of spot 11

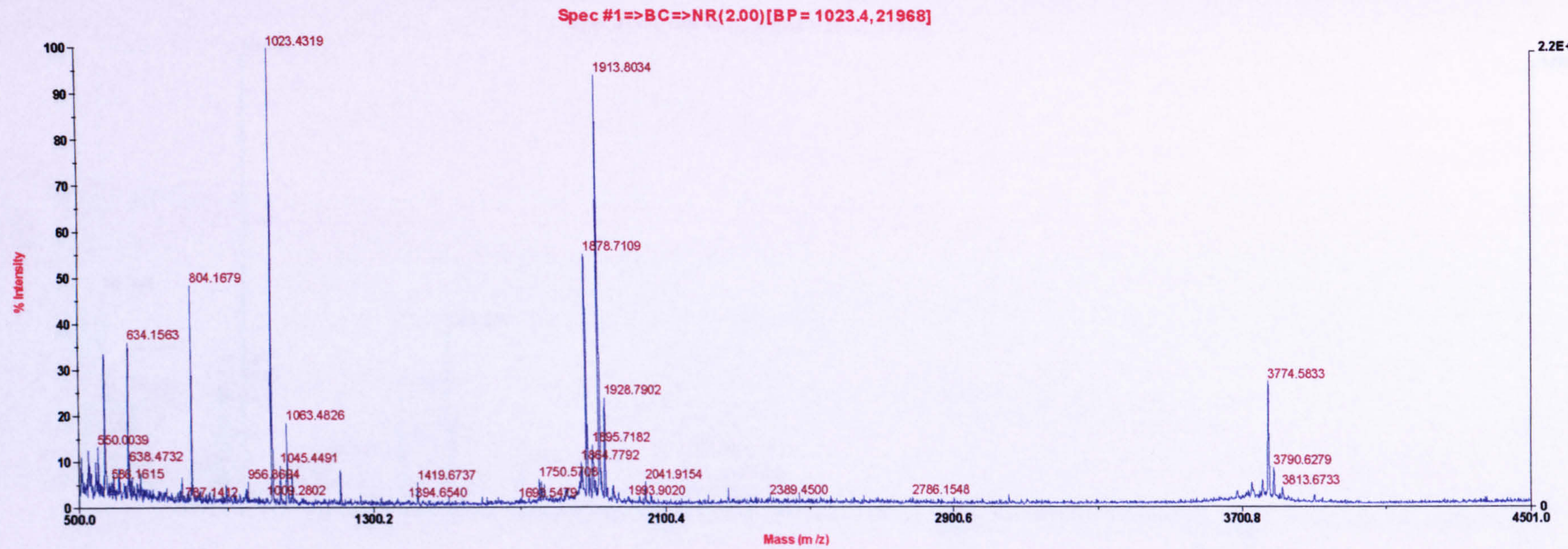


Fig A.12 Peptide mass fingerprint of spot 12

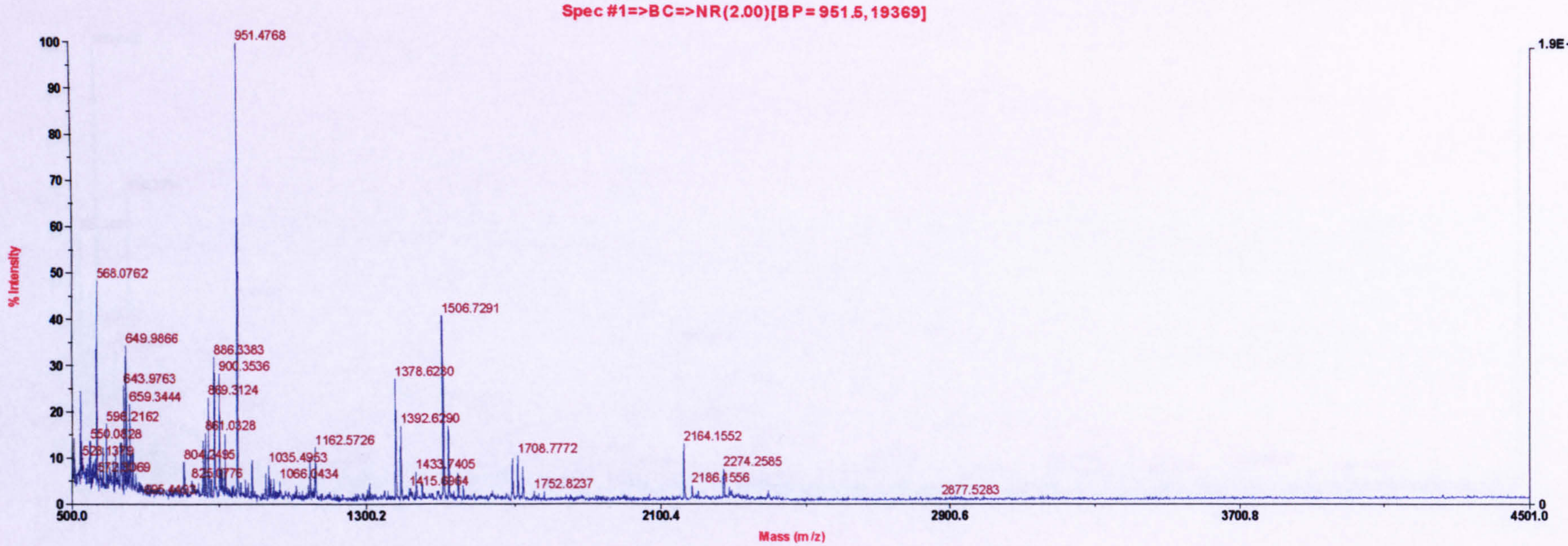


Fig A.13 Peptide mass fingerprint of spot 13

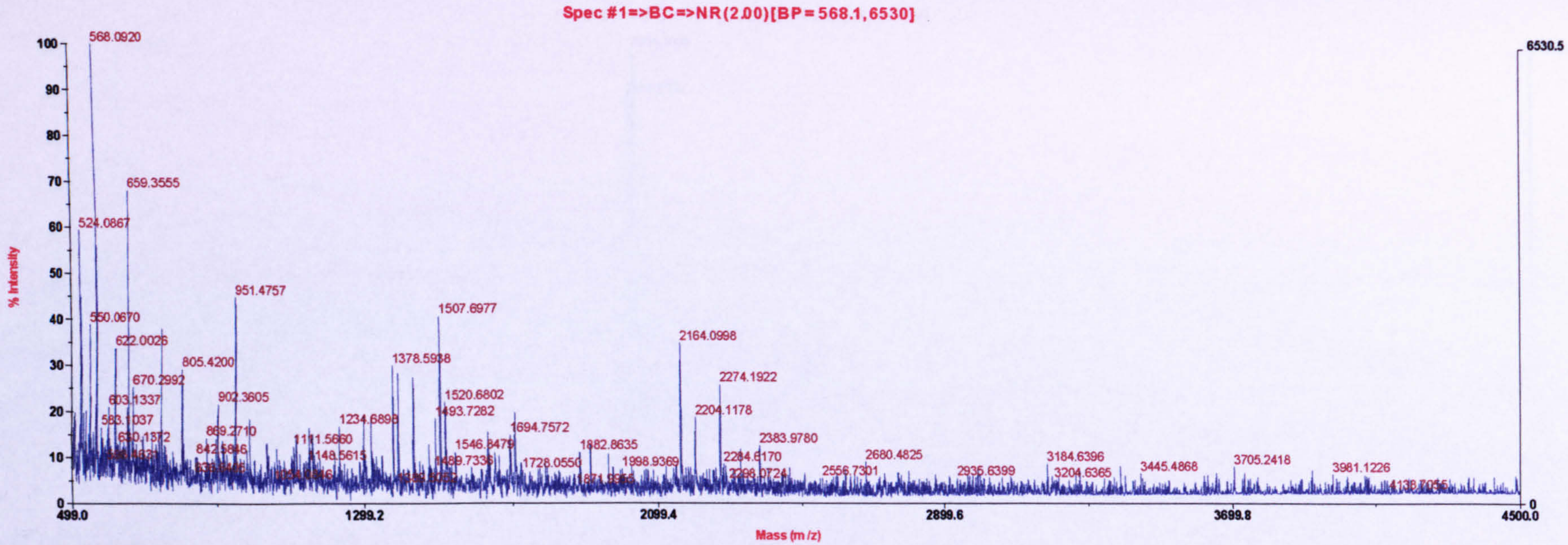


Fig A.14 Peptide mass fingerprint of spot 14

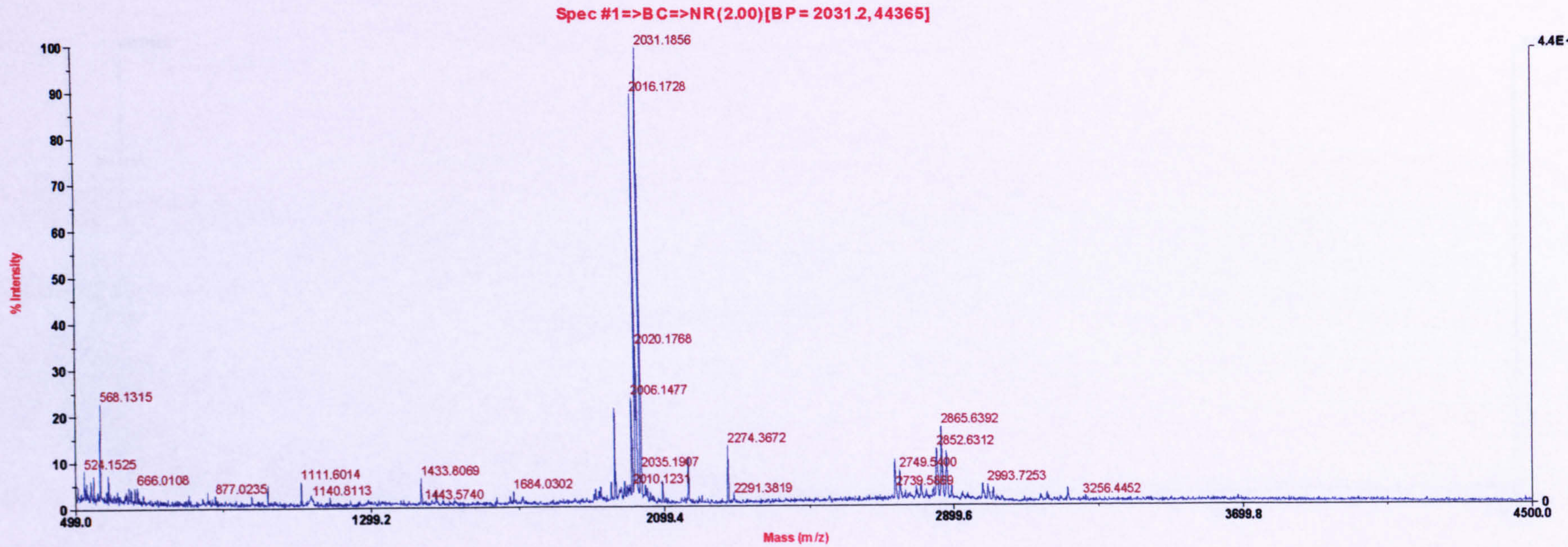


Fig A.15 Peptide mass fingerprint of spot 15

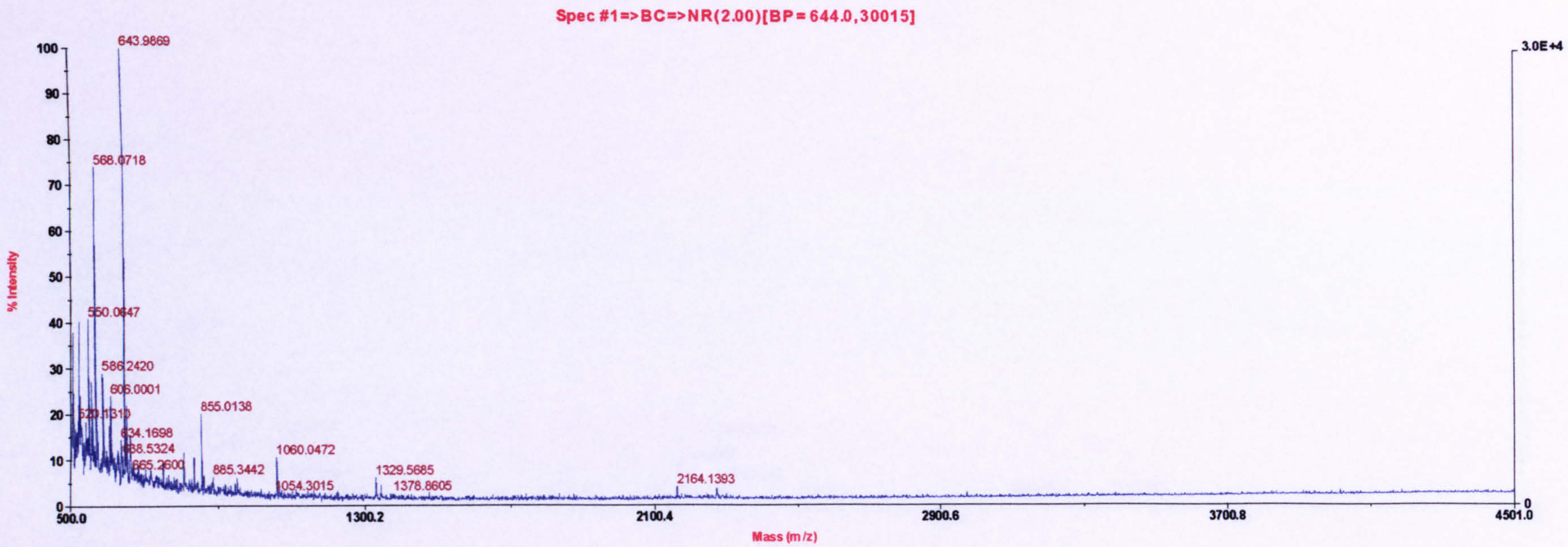


Fig A.16 Peptide mass fingerprint of spot 17

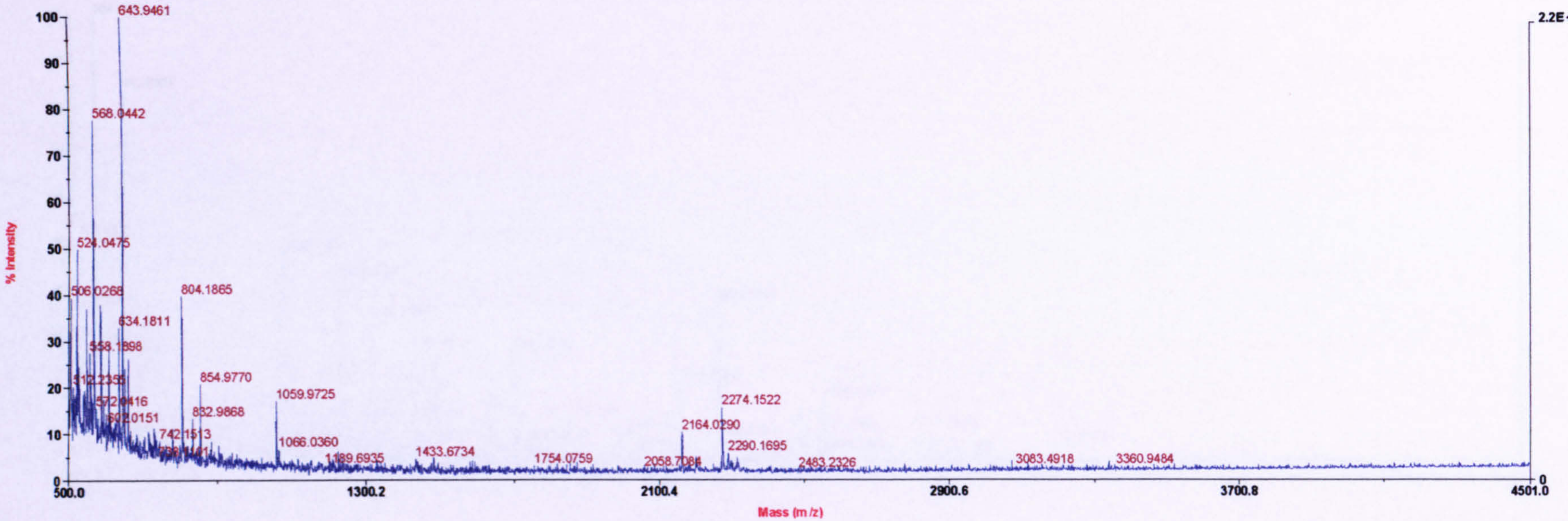


Fig A.17 Peptide mass fingerprint of spot 18

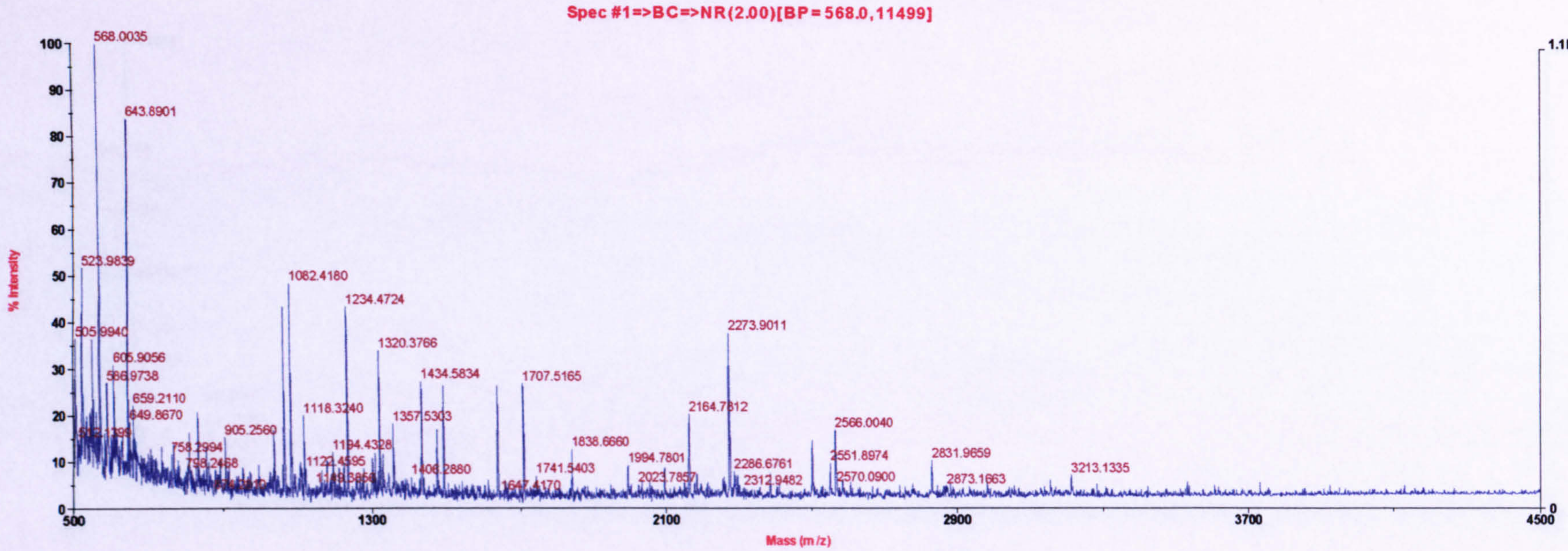


Fig A.18 Peptide mass fingerprint of spot 19

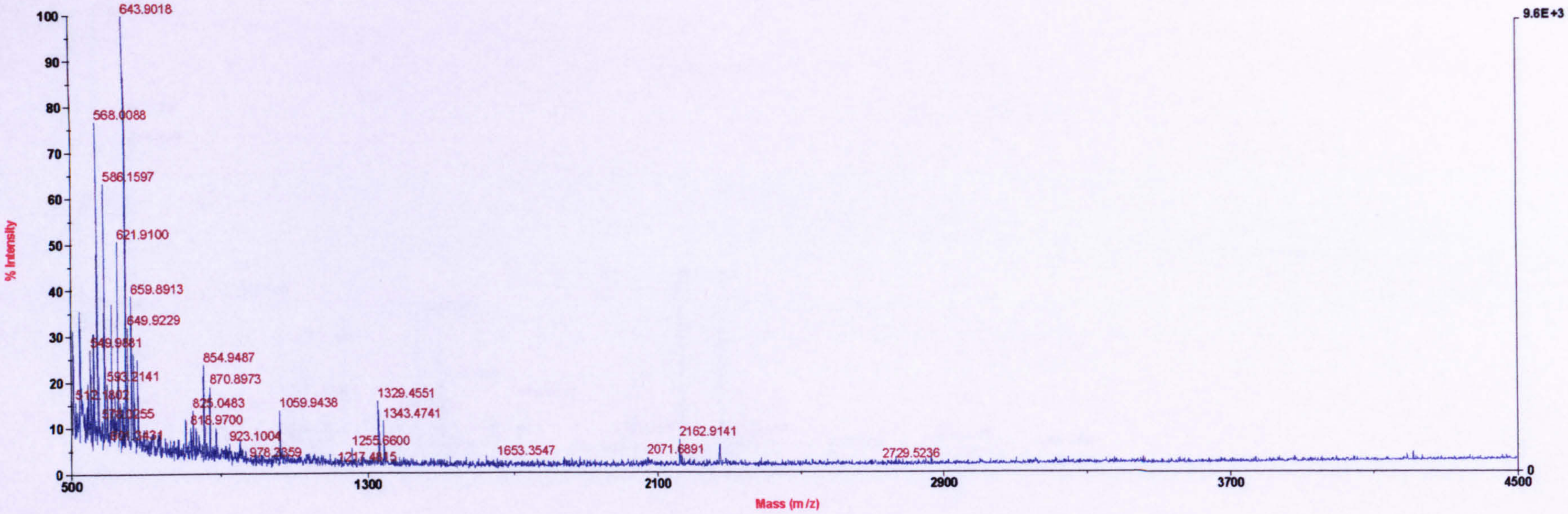


Fig A.19 Peptide mass fingerprint of spot 20

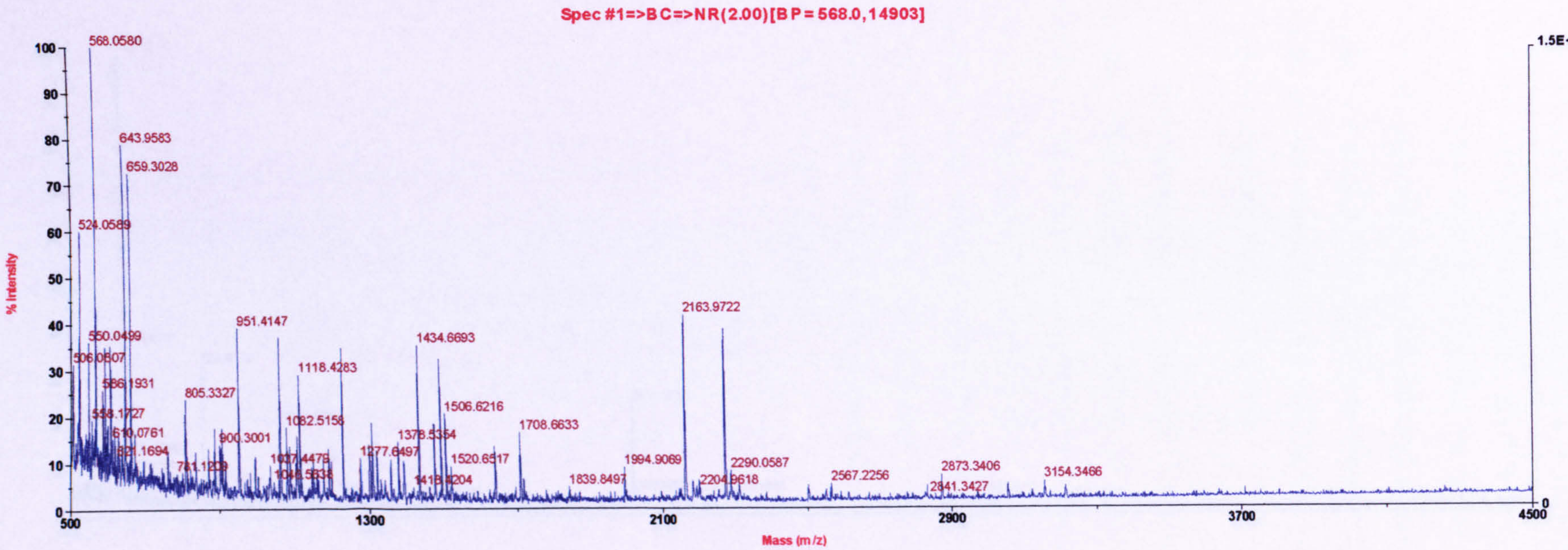


Fig A.20 Peptide mass fingerprint of spot 21

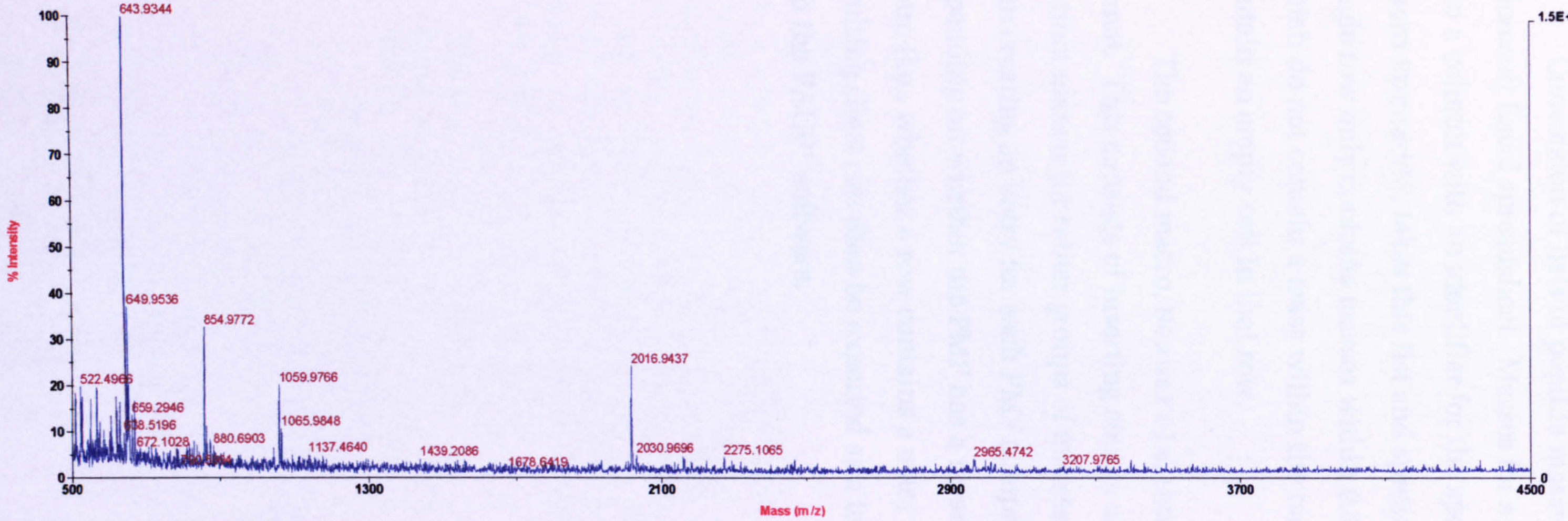


Fig A.21 Peptide mass fingerprint of spot 22

Appendix 2: Excel Macros to Convert Lists of Peptide Mass to NEXUS Format

Conversion of lists of peptide masses is performed using a Microsoft Excel spreadsheet. Masses for a particular gel spot are entered into a column with an identifier for the spot in the first row. The first macro `SpreadMS`, takes this list and spreads out the masses, so that any single row only contains masses within 0.5Da of each other; PMFs which do not contain a mass within the range for a particular row contain an empty cell in that row.

The second macro, `NexusFile`, then converts this list to NEXUS format. This consists of inserting counts of the numbers of PMFs and distinct masses (or rather groups of masses with fall within 0.5Da) and then creating an entry for each PMF comprising a string of 1's and 0's depending on whether the PMF has a mass within a particular mass group (i.e., whether a row contains a mass value or is empty). The resulting sheet can then be exported as a text file and loaded directly into the PAUP* software.

```
Sub SpreadMS()

    Dim NumCols As Integer
    Dim Pos(50) As Integer
    Dim I As Integer
    Dim OutRow As Integer
    Dim MinVal As Double
    Dim Done As Boolean

' Work out the number of PMFs by counting the columns
' which have a name in row 1.

    NumCols = 1
    While Sheet1.Cells(1, NumCols) <> ""
        Sheet2.Cells(1, NumCols) = _
                    Sheet1.Cells(1, NumCols)
        NumCols = NumCols + 1
    Wend
    NumCols = NumCols - 1

' The Pos array holds the current mass we are
' considering for each column, so start by setting
' it to the row number of the first mass (2).

    For I = 1 To NumCols
        Pos(I) = 2
    Next

' Now spread out the masses from sheet1 and put
' them into sheet two.  This is done by finding the
' minimum current mass and then finding all the other
' current masses which fall within 0.5Da.  These are
' then written to a new row in sheet two, and the
' current mass  pointer in the Pos array updated.
' This is repeated until no masses remain from all of
' the PMFs

    OutRow = 2
    Done = False
    While Not Done
        MinVal = 9999999.99
    For I = 1 To NumCols
        If Sheet1.Cells(Pos(I), I) <> "" Then
            If Sheet1.Cells(Pos(I), I) < MinVal _
            Then
                MinVal = Sheet1.Cells(Pos(I), I)
                MinCol = I
            End If
        End If
    Next
```

```
            Sheet2.Cells(OutRow, MinCol) = _
                    Sheet1.Cells(Pos(MinCol), MinCol)
            Pos(MinCol) = Pos(MinCol) + 1
            For I = 1 To NumCols
                If Abs(Sheet1.Cells(Pos(I), I) - MinVal) _
                        <  0.5 Then
                    Sheet2.Cells(OutRow, I) = _
                            Sheet1.Cells(Pos(I), I)
                    Pos(I) = Pos(I) + 1
                End If
            Next
            OutRow = OutRow + 1
            Done = True
            For I = 1 To NumCols
                If Sheet1.Cells(Pos(I), I) <> "" Then
                    Done = False
                End If
            Next
            DoEvents
        Wend
    End Sub
```

```
Sub NexusFile()

Dim I As Integer
Dim J As Integer
Dim NumSpecies As Integer
Dim NumCharacters As Integer
Dim OutLine As String

    Sheet4.Range("A1", "Z3000").Clear

' Count the number of "species" (i.e., PMFs) by
' counting their names (in row 1)

    NumSpecies = 0
    While Sheet2.Cells(1, NumSpecies + 1) <> ""
        NumSpecies = NumSpecies + 1
    Wend

' Count the number of mass groups (characters) by
' finding the first empty row

    NumCharacters = 0
    While Application.WorksheetFunction.Sum( _
     Sheet2.Range(Sheet2.Cells(NumCharacters + 2, 1),
     Sheet2.Cells(NumCharacters + 2, NumSpecies)))
        NumCharacters = NumCharacters + 1
    Wend

' Now write the header information into the file

    Sheet4.Cells(1, 1) = "#NEXUS"
    Sheet4.Cells(2, 1) = ""
    Sheet4.Cells(3, 1) = "Begin Taxa;"
    Sheet4.Cells(4, 1) = "    Dimensions NTax = " & _
                         Format(NumSpecies) & ";"

    OutLine = "    TaxLabels"
    For I = 1 To NumSpecies
        OutLine = OutLine & " " & Sheet2.Cells(1, I)
    Next
    OutLine = OutLine & ";"
    Sheet4.Cells(5, 1) = OutLine
    Sheet4.Cells(6, 1) = "End;"
    Sheet4.Cells(7, 1) = ""
    Sheet4.Cells(8, 1) = "Begin Characters;"
    Sheet4.Cells(9, 1) = "    Dimensions NChar = " & _
                       Format(NumCharacters) & ";"
    Sheet4.Cells(10, 1) = "    Format Datatype = " & _
                        " Standard;"
    Sheet4.Cells(11, 1) = "    Matrix"
```

```
' For each PMF construct a line with 0's and 1's to
' indicate whether a cell is empty or not.

    For I = 1 To NumSpecies
        OutLine = "    " & Sheet2.Cells(1, I) & " "
        For J = 1 To NumCharacters
            If Sheet2.Cells(J + 1, I) = "" Then
                OutLine = OutLine & "0"
            Else
                OutLine = OutLine & "1"
            End If
        Next
        Sheet4.Cells(I + 11, 1) = OutLine
        DoEvents
    Next

' Write out the footer information for PAUP*

    Sheet4.Cells(NumSpecies + 11, 1) = _
            Sheet4.Cells(NumSpecies + 11, 1) & ";"
    Sheet4.Cells(NumSpecies + 12, 1) = "End;"
    Sheet4.Cells(NumSpecies + 13, 1) = ""
    Sheet4.Cells(NumSpecies + 14, 1) = _
                    "Begin Assumptions;"
    Sheet4.Cells(NumSpecies + 15, 1) = _
            "    Options DefType = DOLLO;"
    Sheet4.Cells(NumSpecies + 16, 1) = "End;"
End Sub
```