Proteomic analysis of *Entamoeba histolytica*

J. TOLSTRUP[1], E. KRAUSE[2], E. TANNICH[1] *and* I. BRUCHHAUS[1]*

[1] *Bernhard Nocht Institute for Tropical Medicine, 74 Bernhard Nocht Strasse, 20359 Hamburg, Germany*
[2] *Leibniz Institute of Molecular Pharmacology, 10 Robert-Rössle Strasse, 13125 Berlin, Germany*

(*Received 12 June 2006; revised 19 July 2006; accepted 19 July 2006; first published online 11 October 2006*)

SUMMARY

In this study, the proteome of axenically grown *Entamoeba histolytica* parasites was explored by two-dimensional gel electrophoresis (2-DE), employing a practical and effective procedure for the solubilization of *E. histolytica* proteins. Approximately 900 protein species in the pH range between 4 and 7 were detected by Coomassie Blue staining. Ninety-five spots were excised, trypsinated and subjected to mass spectrometry. The resultant data from peptide mass fingerprints were compared with those available in the *E. histolytica* genome and the (non-redundant) National Center for Biotechnology Information (NCBI) databases for the identification and categorization of proteins. Sixty-three of the proteins identified were predicted to relate to the cytoskeleton, surface, glycolysis, RNA/DNA metabolism, the ubiquitin-proteasome system, vesicular trafficking and signal transduction. The present study demonstrates, for the first time, that corresponding genes are indeed expressed in *E. histolytica* and provides a foundation for further proteomic studies of this parasite.

Key words: *Entamoeba*, proteome, mass spectrometry.

INTRODUCTION

Entamoeba histolytica is an important protozoan parasite of humans, that is transmitted *via* the faecal-oral route. This parasite is responsible for approximately 50 million cases of invasive amoebiasis each year (WHO, 1997). *E. histolytica* usually resides in the large intestine, and can persist there for months or years and cause an asymptomatic intestinal infection. However, this amoeba is often invasive and can also penetrate the intestinal epithelium and cause ulcerative colitis or disseminate into other organs, most commonly the liver, leading to abscess formation.

The genome sequence of *E. histolytica* has been available since 2005 (Loftus *et al.* 2005; http://www.genedb.org/genedb/ehistolytica/; http://www.tigr.org/tdb/e2k1/eha1/). This genome contains approximately 10 000 genes and, interestingly, about 30% of them do not currently have any homologues in public gene databases. Analysis of the genomic data provides evidence for lateral gene transfer of bacterial genes into the *E. histolytica* genome (Loftus *et al.* 2005). Also, most of the genes linked to the mitochondrial amino acid biosynthesis pathways are lacking, and the oxidative stress proteins discovered in *E. histolytica* are usually associated with those described for various anaerobic prokaryotes (Loftus *et al.* 2005). Furthermore, the analysis of the *E. histolytica* genome reveals an intriguing redundancy, exemplified by the presence of 30 homologues of the intermediate subunit of the GalGalNAc lectin, multiple new genes encoding pore-forming peptides and approximately 40 cysteine proteinases (Loftus *et al.* 2005).

The availability of the *E. histolytica* genome sequence opens up a range of unique scientific opportunities. For instance, it is now possible to study the transcription and expression profiles of *E. histolytica* genes by microarray and proteomic analyses, respectively. However, currently there are very limited data for most genes expressed under different culture or environmental conditions and in different strains.

To study the transcriptome of *E. histolytica*, MacFarlane *et al.* (2005*a*, *b*) developed a genomic DNA microarray containing approximately 2000 genes. These authors showed that 31–47% of the genes are expressed under 'standard' culture conditions. Cultivation under 'heat shock' conditions induced an upregulation of 17% of the genes, whereas cultivation in the presence of Caco-2 cells had no significant influence on the transcriptional profile (MacFarlane *et al.* 2005*a*, *b*; Shah *et al.* 2005). Recently, this microarray was used for a comparison of the expression levels between different virulent and non-virulent *Entamoeba* species and strains (*E. histolytica* HM-1:IMSS, *E. histolytica* Rahman, and *E. dispar*). Some genes with known roles in pathogenicity and stress response had a decreased expression in non-pathogenic *E. histolytica* (see

* Corresponding author: Bernhard Nocht Institute for Tropical Medicine, 74 Bernhard Nocht Strasse, 20359 Hamburg, Germany. Tel: +49 404 2818 472. Fax: +49 404 2818 512. E-mail: bruchhaus@bni-hamburg.de.

doi:10.1017/S0031182006001442 Printed in the United Kingdom

MacFarlane and Singh, 2006). In another microarray analysis, employing a 'shotgun' genomic DNA microarray, collagen activated and non-activated trophozoites were compared (Debnath *et al*. 2004). Fourteen differentially regulated genes were analysed further by sequencing. Gene products identified in this way were predicted to be involved in cytoskeleton rearrangements, vesicle formation, phagocytosis, protein degradation, stress response and/or ribosomal assembly, and included cysteine proteinases, ribosomal proteins, HSP70 and myosin 1B (Debnath *et al*. 2004).

Differential expression can also be studied at the protein level, an approach that is likely to be of great relevance for understanding regulatory processes or mechanisms. Thus far, 4 proteomic studies of *E. histolytica* have been conducted (Marion *et al*. 2005; Okada *et al*. 2005, 2006; Leitsch *et al*. 2005), of which 3 have explored the protein composition of purified phagosomes. Different proteins involved in uptake-processes were identified, including those involved in signalling, endocytosis, cytoskeleton rearrangement, vesicular trafficking and surface recognition (Marion *et al*. 2005; Okada *et al*. 2005, 2006). A two-dimensional gel electrophoresis (2-DE) protocol was developed for *E. histolytica*, which allows the reproducible display of protein expression profiles for this organism (Leitsch *et al*. 2005). Approximately 1500 protein spots were detected by silver-staining, and 10 landmark spots of the respective Coomassie-stained gels were identified by matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry or protein sequencing (Leitsch *et al*. 2005).

In the present study, a practical and reproducible procedure for the solubilization of *E. histolytica* proteins and their 2-DE analysis is reported. Of the proteins resolved, approximately 100 proteins were subjected to analysis by MALDI-TOF mass spectrometry. The resultant peptide mass fingerprints were analysed by comparison with data available in the *E. histolytica* genome and NCBI databases. The present investigation is considered to pave the way for the construction of 2-DE reference maps for *E. histolytica*.

MATERIALS AND METHODS

Cultivation of cells and extract preparation

Trophozoites of the *E. histolytica* isolate HM-1:IMSS were cultured axenically in TYI-S-33 medium in plastic tissue culture flasks (Diamond *et al*. 1978). For the experiments, 2×10^6 trophozoites were cultivated for 48 h in 250 ml flasks. The trophozoites were harvested after being chilled on ice for 5 min and sedimented at 430 ***g*** at 4 °C for 5 min. The resultant pellet was washed twice in phosphate-buffered saline (PBS, 6·7 mM $NaHPO_4$, 3·3 mM NaH_2PO_4, 140 mM NaCl, pH 7·2). Immediately prior to cell lysis, the cysteine proteinase inhibitor E-64 was added (final concentration 50 μM) to the cell pellet. Subsequently, the cells were lysed *via* 4 freeze-thaw cycles in CO_2/ethanol. Immediately after the last cycle, E-64 was added (same concentration) to the extract, which was centrifuged for 1 h at 100 000 ***g*** (4 °C). The supernatant (PBS-soluble extract) was stored at −70 °C. The PBS-insoluble fraction was resuspended directly in 600 μl of rehydration buffer prior to storage at −70 °C.

Two-dimensional gel electrophoresis

First dimension IEF was performed at 20 °C using 18 cm immobilized pH gradient (IPG) strips (Amersham Biosciences), applying a linear pH range of 4 to 7. The maximum current setting was 50 μA/strip using the IPGphor Isoelectric Focusing Unit (Amersham Biosciences). The strips were rehydrated for a minimum of 12 h in ceramic strip holders using 100 μl of sample diluted with 250 μl of rehydration buffer (10 M thiourea, 2·5% CHAPS, 0·01% Coomassie Brilliant blue, 0·5% IPG-buffer of the respective pH gradient, and 1 mM dithiothreitol. Approximately 0·5 mg of protein was used. A voltage of 30 V was applied during rehydration (10 h). After rehydration, IEF was carried out under the following conditions: 1000 V for 1 h, 3000 V for 1 h and 8000 V for 12 h. The focusing was terminated between 90 000 and 91 000 Vh. Subsequently, the strips were equilibrated in 50 mM Tris pH 8·8, 6 M urea, 30% glycerol, 2% SDS, 0·03% bromphenol blue supplemented with 65 mM dithiothreitol (15 min) and then in 135 mM iodoacetamide (15 min). After transferring the strips to 12·5% sodium dodecylsulphate (SDS) polyacrylamide gels, they were sealed in place with 0·5% agarose in 25 mM Tris-base, 0·19 M glycine, 0·1% SDS, 0·01% bromphenol blue.

In the second dimension, electrophoresis was performed using 25 mM Tris-base, 0·19 M glycine, 0·1% SDS as the running buffer for 18 h at 10 mA and 4 °C. A pre-stained molecular mass marker (low range, Bio-Rad) was used as a standard. The gels were stained with colloidal Coomassie Brilliant Blue (CBB) staining of proteins (Neuhoff *et al*. 1988). Briefly, gels were incubated in 1·4% phosphoric acid, 8% ammonium sulfate, 20% methanol, 2% Coomassie Brilliant Blue G-250 for 12 h at 22 °C. Destaining was performed in 25% methanol at room temperature for 8 h. Gels were stored at 4 °C in 25% ammonium sulfate.

In-gel digestion of proteins

Protein spots were excised from gels, washed with 50% acetonitrile in 25 mM ammonium bicarbonate,

https://doi.org/10.1017/S0031182006001442 Published online by Cambridge University Press

dehydrated in acetonitrile and dried in a vacuum centrifuge. The gel pieces were rehydrated in 10 μl of 5 mM ammonium bicarbonate, containing 300 ng of trypsin (proteomic grade, Roche Diagnostics, Mannheim, Germany). After 15 min, 10 μl of 5 mM ammonium bicarbonate were added to ensure that the gel pieces were wet during tryptic cleavage (37 °C, overnight). To extract the peptides, 20 μl of 0·3% trifluoroacetic acid (TFA) in acetonitrile were added, and the samples were sonicated for 10 min. The supernatant was removed and dried under vacuum. The peptides were redissolved in 8 μl of 0·1% TFA, 5% acetonitrile in water. Prior to mass spectrometry analysis, the peptides were purified using a reversed-phase minicolumn filled in a micropipette tip (ZipTip C18, Millipore, Bedford, MA, USA), according to the manufacturer's protocol. The peptides were eluted from ZipTips using 2·5 μl of matrix solution (7 mg/ml α-cyano-4-hydroxycinnamic acid (Sigma), dissolved in 60% acetonitrile, 0·12% TFA in water and spotted directly on a target plate and air-dried.

Mass spectrometry

MALDI-mass spectrophotometric measurements were performed using a Voyager-DE STR Bio-Spectrometry Workstation (Perseptive Biosystems, Inc., Framingham, USA) in the reflection mode at an acceleration voltage of 20 kV, 70% grid voltage and a delay of 200 ns. Each spectrum obtained was the mean of 256 laser shots. Mass spectra were calibrated using known autolytic fragments of trypsin as internal standards. The peptide masses were measured as monoisotopic masses. The resultant spectra were processed utilizing the Data Explorer Software (version 3.5.00). Alternatively, an Axima CFR MALDI-TOF spectrometer (Shimadzu Biotech GmbH, Duisburg, Germany) was used. Depending on spectra quality, 200–400 laser shots were accumulated. Control of the spectrometer and spectra processing was carried out employing the Axima CFR software version 2.2.1.

Database analysis

Proteins were identified using the program MS-Fit (http://prospector.ucsf.edu) by comparison with data publicly available in the non-redundant, annotated *E. histolytica* (TIGR; http//www.tigr.org/tdb/e2K1/eha1) and NCBI databases. The parameters for the search were as follows: (i) the modifications on cysteine residues by carboxyamidomethylation were set as obligate, (ii) methionine oxidation was considered as a partial modification and (iii) one was used as the maximum missed tryptic cleavage sites. The monoisotopic masses were considered and the mass tolerance was set at 0·07 Da. To determine the certainty of the analysis, the MOWSE (Molecular Weight Search) score analysis of MS-Fit was used, based on the scoring system described by Pappin *et al.* (1993). For all proteins, the concordance between the theoretical and the experimental isoelectric points and molecular masses as well as a coverage of the protein of more than 20% were essential criteria.

RESULTS AND DISCUSSION

Resolution, sensitivity and reproducibility of 2-DE separations

Entamoeba histolytica contains large amounts of cysteine proteinases, which have to be inhibited rapidly to obtain reproducible and reliable results. Therefore, the presence of the specific cysteine proteinase inhibitor E-64 before and directly after the freeze-thaw cycles of the *E. histolytica* extract was highly beneficial for both resolution and reproducibility in the 2-DE analysis (data not shown). Also, the addition of 10 M thiourea to the rehydration buffer in the solubilization step of both fractions (PBS-soluble proteins as well as the PBS-insoluble proteins) resulted in a greater number of detectable protein species in the 2-DE analysis compared with the solubilization in urea (data not shown). The PBS-soluble as well as the PBS-insoluble protein fractions were subjected to 2-DE analysis. An amount of 500 μg protein per 18 cm IPG strip was shown to be optimal for CBB-stained gels. Representative 2-DE gels of different batches of soluble protein fractions from *E. histolytica* trophozoites are shown in Fig. 1. The extracts were separated over a pH range of 4 to 7, stained with CBB and analysed. The reproducibility of the sample preparation, gel electrophoresis and staining was verified by running sequential replicate gels (at least 5 times) using protein extracts prepared from different cultures of *E. histolytica* trophozoites. Gels run under identical conditions gave almost identical results with respect to spot pattern and relative intensities (Fig. 1). Most of the amoebic proteins separated within a pH range of 4 to 7 have a pI of 5 to 6. A total of approximately 400 CBB-stained proteins could be detected when the PBS-soluble protein extract was separated in a 2-DE using 18 cm IPG strips and a linear pH gradient. In the PBS-insoluble fraction, about 500 protein species could be detected using a Coomassie Blue stain (data not shown).

Protein identification

Using the Coomassie-stained gels of PBS–soluble extracts, 95 protein spots were randomly isolated, in-gel digested, and analysed by MALDI-TOF mass spectrometry. For 90 of these 95 spots, quality peptide mass fingerprint spectra were obtained by

https://doi.org/10.1017/S0031182006001442 Published online by Cambridge University Press

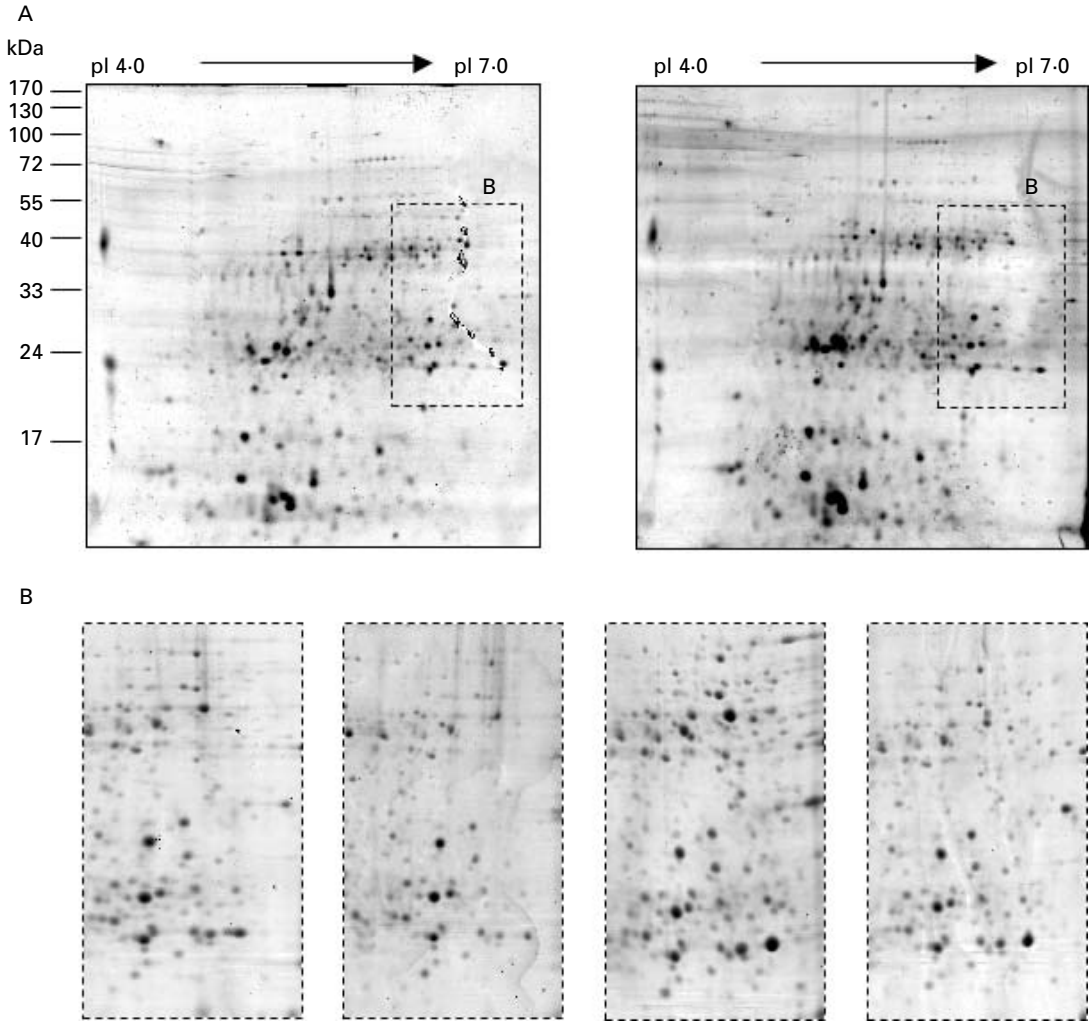

Fig. 1. (A) Representative gels displaying the 2-DE analysis of 2 independent extract preparations from PBS-soluble proteins from *Entamoeba histolytica*. The proteins (500 μg) were separated on linear IPG strips (pH 4–7), followed by 12·5% SDS-PAGE. The gels were stained with Coomassie Blue. Molecular mass standards are indicated on the left. (B) Comparison of 4 enlarged partial 2-DE gels (compared area (B) is marked in (A) showing PBS-soluble proteins derived from 4 independent extract preparations.

MALDI-TOF MS analyses. For 10 spots (# 3, 9, 40, 44, 62, 60, 66, 71, 81, 88; Fig. 2), no protein sequence matches were found in the *E. histolytica* or the NCBI databases. While 63 protein spots related unequivocally to unique proteins (Table 1), for 17 proteins, 2 or more entries were found (Table 2). Multiple entries may result from the presence of 2 or more proteins within 1 protein spot or, alternatively, different proteins may have similar molecular mass fingerprints.

The 63 proteins identified represented 41 different molecules of known function, and 6 proteins were recorded previously as hypothetical proteins in the databases interrogated (Table 1). These proteins could be grouped into 7 categories of functionally related molecules (cytoskeletal proteins, glycolysis, surface-associated proteins, RNA/DNA metabolism, ubiquitin-proteasome pathway, vesicular trafficking and signal transduction) (Table 1).

Proteins associated with the cytoskeleton

The cytoskeleton associated proteins actin, actophorin, myosin calcium-binding light chain, coactosin, actin depolymerization factor and a kinesin-like protein have been reported previously in a phagosome study (Marion *et al*. 2005). Actin depolymerization factors and actophorin are members of the ADF/cofilin protein family (Moon and Dubrin, 1995). These proteins are involved in the turnover of actin filaments in dynamic regions of the cell (Maciver *et al*. 1998; Bamburg, 1999). Coactosin seems to bind the barbed ends of actin filaments, thus sterically hindering the binding of capping proteins (Rohrig *et al*. 1995). Kinesin is a member of the microtubule-based ATPase motor proteins which perform force-generating processes, such as organelle transport and chromosome segregation (Kull *et al*. 1996).

https://doi.org/10.1017/S0031182006001442 Published online by Cambridge University Press

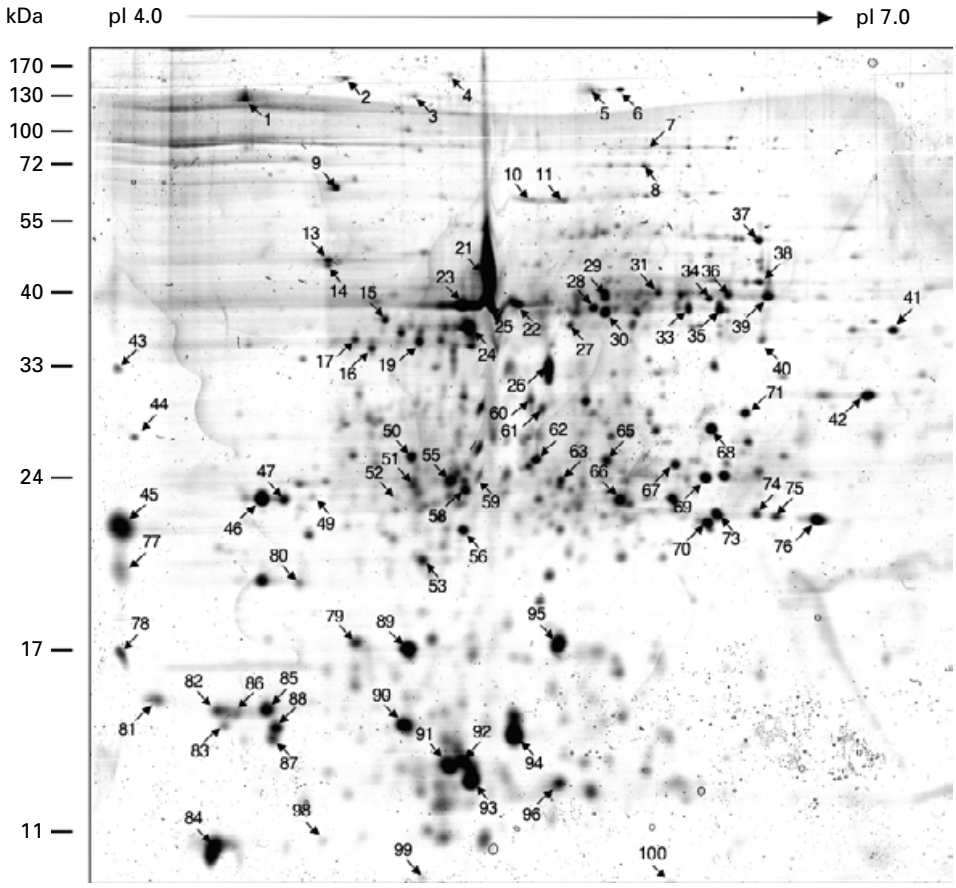

Fig. 2. The 2-DE reference map of proteins from the PBS-soluble extract from *Entamoeba histolytica*. The proteins (500 μg) were separated on linear IPG strips (pH 4–7), followed by 12·5% SDS-PAGE. The gels were stained with Coomassie Blue. Numbers indicate protein spots that were subjected to MALDI-TOF mass spectrometry. The resultant peptide mass fingerprints were analysed and compared with the *E. histolytica* genome and the NCBI databases to predict the identity of the proteins. Spot numbers refer to proteins listed in Tables 1 and 2. Isoelectric point (top) and molecular mass (left) are indicated.

Enzymes involved in glycolysis and glycogen metabolism

Some enzymes of the glycolytic pathway, including aldehyde-alcohol dehydrogenase, NADP-dependent alcohol dehydrogenase, enolase, fructose-1·6-bisphosphate aldolase, pyruvate kinase, pyruvate phosphate dikinase and glycogen phosphorylase were identified. The aldehyde-alcohol dehydrogenase is a bifunctional 95 kDa polypeptide, having both acetaldehyde dehydrogenase and alcohol dehydrogenase activities. Each enzymatic property is located to a different part of the molecule (Bruchhaus and Tannich, 1994; Yang *et al*. 1994; Chen *et al*. 2004). In the present study, a protein with a molecular mass of 38 kDa was detected. All peptide mass fingerprints determined for this protein matched only the aldehyde dehydrogenase domain of the bifunctional enzyme. Thus, the question arises as to whether this molecule is the product of protein degradation during the extraction process or whether the bifunctional protein is proteolytically processed in *E. histolytica*.

Surface-associated proteins

Three putative surface-associated proteins, the 170 kDa subunit of the galactose-inhibitable lectin, a BsA-like leucine rich repeat protein and a cysteine proteinase were identified in the PBS-soluble protein extract. Lectin is one of the most important surface molecules of *E. histolytica*. It is a 260 kDa heterodimer, consisting of the 170 kDa transmembrane (heavy) and a GPI-anchored 35 kDa (light) subunits. These subunits are linked by disulphide bonds (Cheng *et al*. 2001). The functional activity of lectin has been shown to be involved in host cell binding, cytotoxicity, complement resistance and virulence (Mann, 2002; Frederick and Petri, 2005).

Recently, proteins with homology to BspA leucine rich repeat (LRR) proteins were identified (Davis *et al*. 2006). For organisms other than *E. histolytica*, it has been shown that such proteins are surface-associated and that the LRR motif is linked to fibronectin binding (Sharma *et al*. 1998; Kobe and Kajava, 2001). In the genome of *E. histolytica*, more than 80 genes coding for BspA-like proteins have

https://doi.org/10.1017/S0031182006001442 Published online by Cambridge University Press

Table 1. List of *Entamoeba histolytica* proteins analysed by MALDI-TOF mass spectrometry

(The identified proteins can be grouped into 7 functional categories as listed in the table. Accession numbers are from the NCBI protein database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=protein). Apparent pI and molecular weight (MW) were calculated using MacVector 7.2.2 software. T, theoretical; E, experimental.)

Protein identified	Spot no.	Accession no.	MOWSE-score	Matched masses	T/E MW (kDa)	pI T/E
Cytoskeletal proteins						
Actin	21	XP_656738	$2{\cdot}73\times10^{3}$	6/29	41·7/47	5·0/5·7
Actin	22	XP_656738	$2{\cdot}1\times10^{3}$	5/32	41·7/42	5·0/5·7
Actin	24	XP_656738	$6{\cdot}74\times10^{3}$	10/33	41·7/37	5·0/5·6
Actin	25	XP_656738	$2{\cdot}77\times10^{5}$	13/37	42·0/39	5·2/5·8
Actin	23	XP_656610	$7{\cdot}98\times10^{4}$	9/53	33·3/40	5·3/5·5
Actin	26	XP_656610	$1{\cdot}03\times10^{5}$	12/46	33·3/33	5·3/5·8
Actin	37	XP_656610	$3{\cdot}66\times10^{6}$	12/38	33·3/50	5·3/6·5
Actin depolymerization factor	91	XP_654122	$2{\cdot}26\times10^{3}$	4/47	15·2/13	5·4/5·5
Actophorin	93	XP_651689	224	3/20	15·7/13	5·3/5·5
Coactosin	89	XP_650926	$3{\cdot}53\times10^{3}$	5/55	16·1/17	5·0/5·3
Coactosin	90	XP_650926	$3{\cdot}53\times10^{3}$	6/35	16·1/15	5·0/5·3
Kinesin-like protein	82	XP_651294	$2{\cdot}59\times10^{3}$	5/27	64·0/15	6·5/4·7
Myosin calcium-binding light chain	87	XP_657577	$2{\cdot}1\times10^{3}$	4/32	16·9/14	4·7/4·8
Glycolysis						
Aldehyde-alcohol dehydrogenase 2	30	XP_657577	$7{\cdot}99\times10^{5}$	11/57	95·6/38	6·9/6·6
Enolase	31	XP_649161	$2{\cdot}19\times10^{4}$	6/48	47·4/40	5·9/6·2
Enolase	33	XP_649161	$6{\cdot}37\times10^{4}$	10/57	47·4/39	5·9/6·3
Enolase	35	XP_649161	$1{\cdot}03\times10^{5}$	7/57	47·4/39	5·9/6·4
Enolase	39	XP_649161	$9{\cdot}64\times10^{6}$	11/55	47·4/40	5·9/6·5
Fructose-1·6-bisphosphate aldolase	42	XP_650373	$2{\cdot}27\times10^{4}$	8/50	36·2/30	6·1/6·7
Fructose-1·6-bisphosphate aldolase	68	XP_650373	528	6/41	36·2/28	6·1/6·3
Glycogen phosphorylase	5	XP_655667	344	5/19	105/120	6·2/6·0
NADP-dependent alcohol dehydrogenase	41	XP_653507	$2{\cdot}87\times10^{4}$	10/44	39·2/37	6·1/6·7
Pyruvate phosphate dikinase	29	XP_657250	$2{\cdot}58\times10^{4}$	8/52	63/40	5·9/6·0
Pyruvate kinase	75	XP_649089	$2{\cdot}06\times10^{3}$	4/32	36·4/22	6·5/6·5
Surface-associated proteins						
BspA-like leucine rich repeat protein	47	XP_652259	487	4/32	19·7/24	5·0/5·9
Cysteine proteinase (EhCP-C3)	98	XP_655128	$1{\cdot}25\times10^{3}$	4/34	65·5/11	5·5/5·2
Galactose-inhibitable lectin 170 kda subunit	1	XP_655415	$5{\cdot}84\times10^{3}$	9/51	145/120	5·6/4·7
RNA/DNA metabolism						
DEAD/DEAH box helicase	49	XP_649638	$2{\cdot}33\times10^{3}$	4/37	47·7/24	5·7/5·0
DNA polymerase alpha catalytic subunit	6	XP_657373	645	7/22	130/120	8·7/6·1
Similar to signal recognition particle 68 kDa	55	XP_649179	$1{\cdot}1\times10^{3}$	9/35	54·6/24	8·9/5·5
Ubiquitin-proteasome pathway						
70 kDa heat shock protein	10	XP_653218	754	5/35	74·0/63	5·1/5·8
Heat shock protein 70	11	XP_650458	$6{\cdot}03\times10^{3}$	7/30	72·0/63	5·3/5·9
Proteasome alpha subunit	65	XP_65639	$6{\cdot}52\times10^{3}$	5/47	26·9/25	5·9/6·0
Proteasome alpha subunit	43	XP_653086	$1{\cdot}06\times10^{3}$	4/27	28·2/42	5·6/4·1
Ubiquitin	100	XP_650527	$4{\cdot}16\times10^{3}$	6/31	8·7/10	6·6/6·1
Ubiquitin-related protein	99	XP_651112	$1{\cdot}42\times10^{3}$	3/31	8·6/10	5·0/5·3
Ubiquitin-conjugating enzyme	27	XP_657165	$8{\cdot}14\times10^{3}$	6/50	39·8/37	5·0/5·9
Ubiquitin-protein ligase E3 component	4	XP_650859	$3{\cdot}4\times10^{3}$	7/25	160/130	5·8/5·5
Vesicular trafficking						
Rab GDP dissociation inhibitor	28	XP_655898	$3{\cdot}13\times10^{4}$	7/55	55·0/38	5·4/6·0
Rab family GTPase RabK1	76	XP_652298	$1{\cdot}18\times10^{3}$	4/43	24·4/22	5·0/6·6
Rab family GTPase RabK1	83	XP_652298	355	3/32	24·4/15	5·0/4·5
Rab family GTPase RabF4	59	XP_654217	364	3/24	23·7/24	6·6/5·6
Rab family GTPase RabX28	53	XP_650164	986	5/27	21·7/21	5·6/5·4
Rab family GTPase RabX28	74	XP_650164	341	4/30	21·7/22	5·6/6·4
Rho family GTPase	56	XP_650536	447	2/20	22·5/22	5·8/5·5
Rho GDP exchange inhibitor	95	XP_654522	$1{\cdot}93\times10^{7}$	12/39	20·0/17	5·5/5·9
Signal transduction						
Protein kinase	2	XP_655027	72·7	4/22	138/130	5·7/5·2
Serine/threonine phosphatase	70	XP_656538	616	5/31	46·8/23	8·3/6·3

https://doi.org/10.1017/S0031182006001442 Published online by Cambridge University Press

Table 1 (*cont.*)

Protein identified	Spot no.	Accession no.	MOWSE-score	Matched masses	T/E MW (kDa)	pI T/E
Other proteins						
Hypothetical protein with bipartite nuclear localization signal	15	XP_654489	659	3/25	33·0/38	6·2/5·2
Hypothetical protein with bipartite nuclear localization signal	50	XP_657008	$1{\cdot}98 \times 10^3$	8/56	37·4/26	9·1/5·3
Hypothetical protein	58	XP_653545	$7{\cdot}77 \times 10^3$	4/47	21·2/24	5·8/5·5
Hypothetical protein	63	XP_652924	$4{\cdot}7 \times 10^3$	8/30	29·3/24	9·5/5·9
Hypothetical protein with Lipid-binding START	73	XP_649606	884	3/33	23·4/22	5·9/6·3
Hypothetical protein with Lipid-binding START	67	XP_649520	$5{\cdot}12 \times 10^4$	7/45	25·5/25	5·8/6·2
14-3-3 protein 3	80	XP_654465	$5{\cdot}34 \times 10^4$	6/35	27·5/20	4·8/4·9
Calreticulin	14	XP_655241	$7{\cdot}09 \times 10^3$	8/54	45·0/48	4·7/5·1
60S ribosomal protein L27	92	XP_654328	$1{\cdot}15 \times 10^3$	5/50	15·8/13	10·4/5·5
Short-chain dehydrogenases/ reductase SDR	16	XP_649179	$1{\cdot}02 \times 10^3$	5/32	35·6/33	8·1/5·2
Skp1 protein	86	XP_652495	$1{\cdot}7 \times 10^3$	4/27	18·2/15	4·6/4·8
TPR repeat protein	52	XP_652329	707	4/37	35·2/24	5·4/5·3
WD-40 repeat and SOF1 domain protein	19	XP_652155	$1{\cdot}39 \times 10^3$	3/32	34·9/36	5·3/5·3
WD repeat and SOF1 domain protein	17	XP_655317	532	4/31	52·4/38	8·8/5·2
Zinc finger protein	7	XP_651863	538	7/31	93·4/95	5·4/6·2

been identified (http://www.tigr.org/tdb/e2k1/eha1/). Further characterization of one family member confirmed the surface localization for *E. histolytica*, although no transmembrane domain has been predicted (Davis *et al.* 2006). The BspA-like protein identified herein has a molecular mass of approximately 20 kDa, which is in marked contrast to a previous study (Davis *et al.* 2006) describing a protein of about 60 kDa. These proteins have a sequence identity of around 50%.

In addition, a cysteine proteinase was identified which contains an N-terminal, putative transmembrane domain. As for the acetaldehyde/alcohol dehydrogenase identified, only a portion of 11 kDa was detected, possibly as a result of protein degradation.

Proteins of the ubiquitin-proteasome system

Entamoeba contains a ubiquitin-proteasome system. Some proteins involved in this system, including 2 proteasome alpha subunits, HSP70, ubiquitin, an ubiquitin-related protein, an ubiquitin-conjugating enzyme and an ubiquitin-protein ligase E3 component, were discovered using the present proteomic approach.

Signalling proteins and intracellular trafficking

Small GTP-binding proteins are common to all eukaryotes. They are involved in important processes, such as intracellular trafficking, cell proliferation and cytoskeleton rearrangement. The superfamily of GTP-binding proteins is classified into 5 families (*i.e.* Ras, Rho/Rac, Rab, Sar/Arf and Ran). Ninety-one putative Rab proteins have been identified in the *E. histolytica* genome database (Saito-Nakano *et al.* 2005). In the present investigation, molecules RabK1, RabF4, and RabX28 were identified. For all 3 of these molecules, no obvious homologues were detected in other organisms for which data are available. In addition to the 3 Ras family members, 1 Rho family member was discovered. It is known that about 30 Rho proteins are represented in the *E. histolytica* genome database (http://www.tigr.org/tdb/e2k1/eha1/). The activity of Ras and Rho proteins is modulated by 3 classes of proteins, namely GTPase activating protein (GAP), GDP-dissociation stimulator (GDS) and GDP-dissociation inhibitor (GDI) (Takai *et al.* 1994). This proteomic approach revealed 1 Rab and 1 Rho GDP dissociation/exchange inhibitor. Based on the *E. histolytica* genome (http://www.tigr.org/tdb/e2k1/eha1/), 3 genes of the Rab dissociation inhibitor but only 1 gene of the Rho GDP exchange inhibitor are known. In a recent proteomic study of the phagosome of *E. histolytica*, several different small GTPases (e.g. Rab1, 5, 7A, 7b, 7C, 7E, 8, 11B, 11C, X17) and other proteins involved in vesicular trafficking were identified (Okada *et al.* 2006). Surprisingly, none of these proteins were recovered in the present investigation.

Proteins involved in signal transduction

Herein 1 kinase and 1 phosphatase, proposed to be involved in signal transduction, were identified. The kinase has homology to serine/threonine kinases and belongs to a large family of more than 300

https://doi.org/10.1017/S0031182006001442 Published online by Cambridge University Press

Table 2. List of *Entamoeba histolytica* proteins analysed by MALDI-TOF mass spectrometry with two or more entries in the database

(Accession numbers are from the NCBI protein database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=protein). Apparent pI and molecular weight (MW) were calculated using MacVector 7.2.2 software. T, theoretical; E, experimental.)

Spot no.	Protein identified	Accession no.	Mowse score	Matches masses	T/E MW (kDa)	T/E pI
8	Hypothetical protein	XP_657547	$1{\cdot}21 \times 10^4$	6/46	88·0/72	8·7/6·2
	Hypothetical protein	XP_650531	$3{\cdot}87 \times 10^3$	7/46	94·3/72	8·8/6·2
	Protein kinase	XP_653371	$1{\cdot}86 \times 10^3$	5/46	69·9/72	8·4/6·2
	Damaged DNA binding protein	XP_654084	$1{\cdot}67 \times 10^3$	8/46	113/72	6·2/6·2
	DNA-directed RNA polymerase subunit	XP_653488	$1{\cdot}03 \times 10^3$	9/46	132/72	8·7/6·2
	Hypothetical protein	XP_654308	$1{\cdot}02 \times 10^3$	6/46	86·0/72	6·1/6·2
13	Hypothetical protein	XP_648157	$1{\cdot}4 \times 10^3$	7/39	65/49	6·9/5·1
	Zinc finger protein	XP_651208	846	9/39	110/49	8·6/5·1
	Hypothetical protein	XP_648048	671	7/39	75/49	5·8/5·1
	90 kDa heat shock protein	XP_64996	646	7/39	81/49	5·0/5·1
34	Hypothetical protein	XP_654342	$8{\cdot}92 \times 10^3$	8/49	73·0/40	6·0/6·4
	Protein kinase	XP_652434	$8{\cdot}61 \times 10^3$	9/49	63/40	8·7/6·4
36	Hypothetical protein	XP_656878	$1{\cdot}4 \times 10^4$	10/55	48/40	5·7/6·4
	Conserved hypothetical protein	XP_649508	$4{\cdot}79 \times 10^3$	10/55	94/40	5·3/6·4
	Hypothetical protein	XP_656156	$4{\cdot}02 \times 10^3$	7/55	81·9/40	4·9/6·4
38	Oligopeptidase A	XP_649600	$3{\cdot}79 \times 10^3$	7/55	83·7/44	5·6/6·5
	Hypothetical protein	XP_657181	$2{\cdot}55 \times 10^3$	7/55	91·6/44	8·8/6·5
	Hypothetical protein	XP_654342	$2{\cdot}54 \times 10^3$	6/55	73/44	6·0/6·5
	RNA-binding protein	XP_654873	$2{\cdot}18 \times 10^3$	6/51	42·5/23	8·6/5·2
45	Hypothetical protein	XP_651427	$3{\cdot}08 \times 10^3$	6/51	48·4/23	7·1/5·2
	Hypothetical protein	XP_654302	2000	6/51	43·9/23	6·2/5·2
	Hypothetical protein	XP_657047	$1{\cdot}3 \times 10^3$	5/39	29·7/23	5·6/5·2
	GTP-binding protein beta chain	XP_657050	669	5/51	35·0/23	7·1/5·2
46	Phosphofructokinase	XP_653373	$8{\cdot}13 \times 10^3$	6/49	47·7/23	6·9/4·7
	Hypothetical protein	XP_653264	$3{\cdot}6 \times 10^3$	4/49	27·4/23	9·0/4·7
	Syntaxin	XP_656700	$1{\cdot}32 \times 10^3$	4/49	25·0/23	6·4/4·7
51	Hypothetical protein	XP_649289	213	3/34	19·6/25	4·8/5·3
	Hypothetical protein	XP_655821	735	3/34	36·8/25	5·4/5·3
	Hypothetical protein	XP_651686	542	3/34	28·3/25	5·2/5·3
61	Diacylglycerol kinase	XP_651926	$2{\cdot}08 \times 10^4$	6/58	59·9/29	7·4/5·9
	Proteasome regulatory subunit	XP_655570	$1{\cdot}03 \times 10^4$	7/58	54·3/29	8·0/5·9
69	Oligopeptidase A	XP_649600	$1{\cdot}08 \times 10^4$	6/44	83·8/24	5·6/6·3
	Centromere/microtubule binding protein cbf5	XP_649989	$2{\cdot}97 \times 10^3$	6/44	54·1/24	8·4/6·3
	Proteasome alpha subunit	XP_653086	$1{\cdot}97 \times 10^3$	4/44	28·1/24	5·6/6·3
77	Proteinphosphatase 2C epsilon	XP_656538	781	4/33	46·7/21	8·3/5·2
	Hypothetical protein	XP_651600	623	5/33	40·9/21	7·4/5·2
78	Hexokinase	XP_655965	$3{\cdot}62 \times 10^3$	7/51	49·4/18	5·1/5·2
	Glucosamine 6-phosphate N-acetyltransferase	XP_655194	$1{\cdot}4 \times 10^3$	3/25	17·7/18	6·8/5·2
79	WD repeat protein	XP_656135	$5{\cdot}12 \times 10^3$	4/28	73·8/20	7·1/5·7
	Cysteine proteinase 3	XP_653254	$1{\cdot}97 \times 10^3$	3/28	33·9/20	6·9/5·7
	Proteinphosphatase 2C epsilon	XP_656538	$1{\cdot}18 \times 10^3$	4/28	46·7/20	8·3/5·7
84	Hypothetical protein	XP_650424	926	2/39	9·2/10	5·1/4·5
	Hypothetical protein	XP_648700	487	4/34	17·2/10	9·8/4·5
85	Zinc finger protein	XP_655068	$2{\cdot}17 \times 10^4$	5/29	45·4/16	8·5/4·9
	Hypothetical protein	XP_649407	$8{\cdot}85 \times 10^3$	4/29	63·9/16	5·3/4·9
	Hypothetical protein	XP_653196	$2{\cdot}27 \times 10^3$	3/29	53·4/16	8·3/4·9
94	Ubiquitin conjugating enzyme	XP_657165	$8{\cdot}03 \times 10^3$	4/35	39·8/14	5·0/5·7
	DNA topoisomerase	XP_654689	$2{\cdot}02 \times 10^3$	5/38	33·9/14	9·5/5·7
96	Rab family GTPase	XP_654968	392	3/25	21·8/12	6·0/5·8
	Nucleotide biding protein	XP_650593	303	3/25	35·9/12	6·1/5·8
	Hypothetical protein	XP_655441	205	3/25	15·5/12	4·4/5·8

https://doi.org/10.1017/S0031182006001442 Published online by Cambridge University Press

members represented in the *E. histolytica* genome. For the serine/threonine phosphatases, genes of 66 members exist within the *E. histolytica* genome (http://www.genedb.org/genedb/ehistolytica/).

Proteins involved in RNA/DNA metabolism

Three proteins predicted to be involved in RNA or DNA metabolism, including a helicase, DNA polymerase and a putative signal recognition particle, were identified.

In conclusion, the combined use of the present proteomic approach together with comparative analyses in publicly available genome sequence data for *E. histolytica* and other organisms has identified a range of *E. histolytica* proteins which had not been reported previously to be expressed in this parasite. Clearly, the present proteomic analysis provides a foundation for the identification of novel proteins from *E. histolytica* which are crucial for the development, survival and pathogenicity of this parasite, and should also lead to a better understanding of the molecular biology of *E. histolytica* and its relationship with its host.

This work was supported by the Deutsche Forschungsgemeinschaft (DFG), BR 1744/7-1. We would like to thank T. W. Gilberger and T. Roeder for critical reading of the manuscript.

REFERENCES

Bamburg, J. R. (1999). Proteins of the ADF/cofilin family: essential regulators of actin dynamics. *Annual Review of Cellular Developmental Biology* **15**, 185–230.

Bruchhaus, I. and Tannich, E. (1994). Purification and molecular characterization of the NAD(+)-dependent acetaldehyde/alcohol dehydrogenase from *Entamoeba histolytica*. *The Biochemical Journal* **303**, 743–748.

Chen, M., Li, E. and Stanley, S. L. Jr. (2004). Structural analysis of the acetaldehyde dehydrogenase activity of *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2), a member of the ADHE enzyme family. *Molecular and Biochemical Parasitology* **137**, 201–205.

Cheng, X. J., Hughes, M. A., Huston, C. D., Loftus, B., Gilchrist, C. A., Lockhart, L. A., Ghosh, S., Miller-Sims, V., Mann, B. J., Petri, W. A. Jr. and Tachibana, H. (2001). Intermediate subunit of the Gal/GalNAc lectin of *Entamoeba histolytica* is a member of a gene family containing multiple CXXC sequence motifs. *Infection and Immunity* **69**, 5892–5898.

Davis, P. H., Zhang, Z., Chen, M., Zhang, X., Chakraborty, S. and Stanley, S. L. Jr. (2006). Identification of a family of BspA like surface proteins of *Entamoeba histolytica* with novel leucine rich repeats. *Molecular and Biochemical Parasitology* **145**, 111–116.

Debnath, A., Das, P., Sajid, M. and McKerrow, J. H. (2004). Identification of genomic responses to collagen binding by trophozoites of *Entamoeba histolytica*. *Journal of Infectious Diseases* **190**, 448–457.

Diamond, L. S., Harlow, D. R. and Cunnick, C. C. (1978). A new medium for axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **72**, 431–432.

Frederick, J. R. and Petri, W. A. Jr. (2005). Roles for the galactose-/N-acetylgalactosamine-binding lectin of *Entamoeba* in parasite virulence and differentiation. *Glycobiology* **15**, 53R–59R.

Kobe, B. and Kajava, A. V. (2001). The leucine-rich repeat as a protein recognition motif. *Current Opinion of Structural Biolology* **11**, 725–732.

Kull, F. J., Sablin, E. P., Lau, R., Fletterick, R. J. and Vale, R. D. (1996). Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature, London* **380**, 550–555.

Leitsch, D., Radauer, C., Paschinger, K., Wilson, I. B., Breiteneder, H., Scheiner, O. and Duchene, M. (2005). *Entamoeba histolytica*: analysis of the trophozoite proteome by two-dimensional polyacrylamide gel electrophoresis. *Experimental Parasitology* **110**, 191–195.

Loftus, B., Anderson, I., Davies, R., Alsmark, U. C., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R. P., Mann, B. J., Nozaki, T., Suh, B., Pop, M., Duchene, M., Ackers, J., Tannich, E., Leippe, M., Hofer, M., Bruchhaus, I., Willhoeft, U., Bhattacharya, A., Chillingworth, T., Churcher, C., Hance, Z., Harris, B., Harris, D., Jagels, K., Moule, S., Mungall, K., Ormond, D., Squares, R., Whitehead, S., Quail, M. A., Rabbinowitsch, E., Norbertczak, H., Price, C., Wang, Z., Guillen, N., Gilchrist, C., Stroup, S. E., Bhattacharya, S., Lohia, A., Foster, P. G., Sicheritz-Ponten, T., Weber, C., Singh, U., Mukherjee, C., El-Sayed, N. M., Petri, W. A. Jr., Clark, C. G., Embley, T. M., Barrell, B., Fraser, C. M. and Hall, N. (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature, London* **433**, 865–868.

MacFarlane, R., Bhattacharya, D. and Singh, U. (2005*a*). Genomic DNA microarrays for *Entamoeba histolytica*: applications for use in expression profiling and strain genotyping. *Experimental Parasitology* **110**, 196–202.

MacFarlane, R. C., Shah, P. H. and Singh, U. (2005*b*). Transcriptional profiling of *Entamoeba histolytica* trophozoites. *International Journal for Parasitology* **35**, 533–542.

MacFarlane, R. C. and Singh, U. (2006). Identification of differentially expressed genes in virulent and nonvirulent *Entamoeba* species: potential implications for amebic pathogenesis. *Infection and Immunity* **74**, 340–351.

Maciver, S. K., Pope, B. J., Whytock, S. and Weeds, A. G. (1998). The effect of two actin depolymerizing factors (ADF/cofilins) on actin filament turnover: pH sensitivity of F-actin binding by human ADF, but not of *Acanthamoeba* actophorin. *European Journal of Biochemistry* **256**, 388–397.

Mann, B. J. (2002). Structure and function of the *Entamoeba histolytica* Gal/GalNAc lectin. *International Review of Cytology* **216**, 59–80.

Marion, S., Laurent, C. and Guillen, N. (2005). Signalization and cytoskeleton activity through myosin IB during the early steps of phagocytosis in *Entamoeba histolytica*: a proteomic approach. *Cellular Microbiology* **7**, 1504–1518.

Moon, A. and Drubin, D. G. (1995). The ADF/Cofilin proteins: stimulus-responsive modulators of actin dynamics. *Molecular Biology of the Cell* **6**, 1423–1431.

Neuhoff, V., Arold, N., Taube, D. and Ehrhardt, W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**, 255–262.

Okada, M., Huston, C. D., Mann, B. J., Petri, W. A. Jr., Kita, K. and Nozaki, T. (2005). Proteomic analysis of phagocytosis in the enteric protozoan parasite *Entamoeba histolytica*. *Eukaryotic Cell* **4**, 827–831.

Okada, M., Huston, C. D., Oue, M., Mann, B. J., Petri, W. A. Jr., Kita, K. and Nozaki, T. (2006). Kinetics and strain variation of phagosome proteins of *Entamoeba histolytica* by proteomic analysis. *Molecular and Biochemical Parasitology* **145**, 171–183.

Pappin, D. J., Hojrup, P. and Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* **3**, 327–332.

Rohrig, U., Gerisch, G., Morozova, L., Schleicher, M. and Wegner, A. (1995). Coactosin interferes with the capping of actin filaments. *FEBS Letters* **374**, 284–286.

Saito-Nakano, Y., Loftus, B. J., Hall, N. and Nozaki, T. (2005). The diversity of Rab GTPases in *Entamoeba histolytica*. *Experimental Parasitology* **110**, 244–252.

Shah, P. H., MacFarlane, R. C., Bhattacharya, D., Matese, J. C., Demeter, J., Stroup, S. E. and Singh, U. (2005). Comparative genomic hybridizations of *Entamoeba* strains reveal unique genetic fingerprints that correlate with virulence. *Eukaryotic Cell* **4**, 504–515.

Sharma, A., Sojar, H. T., Glurich, I., Honma, K., Kuramitsu, H. K. and Genco, R. J. (1998). Cloning, expression, and sequencing of a cell surface antigen containing a leucine-rich repeat motif from *Bacteroides forsythus* ATCC 43037. *Infection and Immunity* **66**, 5703–5710.

Takai, Y., Kaibuchi, K., Sasaki, T., Tanaka, K., Shirataki, H. and Nakanishi, H. (1994). Rho small G protein and cytoskeletal control. *Princess Takamatsu Symposium* **24**, 338–350.

World Health Organization (1997). WHO/PanAmerican Health Organization/UNESCO Bulletin. *World Health Organization* **72**, 97–100.

Yang, W., Li, E., Kairong, T. and Stanley, S. L. Jr. (1994). *Entamoeba histolytica* has an alcohol dehydrogenase homologous to the multifunctional adhE gene product of *Escherichia coli*. *Molecular and Biochemical Parasitology* **64**, 253–260.

https://doi.org/10.1017/S0031182006001442 Published online by Cambridge University Press