

Stress-Dependent Expression of a Polymorphic, Charged Antigen in the Protozoan Parasite *Entamoeba histolytica*

S. Satish,^{1†} Abhijeet A. Bakre,² Sudha Bhattacharya,² and Alok Bhattacharya^{1,3*}

School of Life Sciences,¹ School of Environmental Sciences,² and Bioinformatics Centre,³ Jawaharlal Nehru University, New Delhi-110067, India

Received 21 March 2003/Returned for modification 15 April 2003/Accepted 12 May 2003

We have identified a novel stress inducible gene, Ehssp1 in *Entamoeba histolytica*, the causative agent of amebiasis. Ehssp1 belongs to a polymorphic, multigene family and is present on multiple chromosomes. No homologue of this gene was found in the NCBI database. Sequence alignment of the multiple copies, and genomic PCR data restricted the polymorphism to the central region of the gene. This region contains a polypurine stretch that encodes a domain rich in acidic and basic amino acids. Under normal culture conditions only one copy of this multigene family is expressed, as observed by Northern blot and RT-PCR analysis. The size of this copy of the gene is 1,077 nucleotides, encoding a protein of 359 amino acids. The polymorphic domain in this copy is 64 nucleotides long. However, on exposure of cells to stress conditions such as heat shock or oxidative stress, multiple polymorphic copies of the gene are expressed, suggesting a possible role of this gene in adaptation of cells to stress conditions. The gene copy expressed under normal conditions, and the expression profile of cells under heat stress was identical in two different strains of *E. histolytica* tested. Interestingly, the extent of polymorphism in this gene was very less in *E. dispar*, a nonpathogenic sibling species of *E. histolytica*. Ehssp1 was found to be antigenic in invasive amebiasis patients.

Microbial pathogens encounter highly stressful environments, including alterations in temperature, pH and pO₂ (44), as they enter the host. In addition, they are exposed to the host defense mechanisms, which involve phagocytes. Within these phagocytes, pathogens are confronted with reactive oxygen and nitrogen intermediates and attack by lysosomal enzymes. To protect themselves against the host, pathogens have evolved several strategies including synthesis of heat shock proteins (HSPs) (26, 13). HSPs of pathogens are important antigens and are known to induce very strong humoral and cellular immune responses in infectious diseases caused by bacteria, protozoa, fungi, and nematodes (44). A large number of genes (including HSPs) are induced by heat stress (32). In general, adaptation to one stressor often leads to protection to others, indicating a cross talk between the different pathways involved in the cellular stress response. These stress related genes are expressed at low level in normal cells and their expression is stimulated manyfold under a variety of stress conditions.

The enteric protozoan parasite *E. histolytica* normally resides in the human gut. It is microaerophilic with optimal growth temperature of 35.5°C. Under conditions that are not yet clearly understood, the actively dividing trophozoites in the gut lumen convert into cysts, which are excreted in the feces. Alternatively, at a much lower frequency, the trophozoites invade the intestinal mucosa and may spread to other organs, notably the liver. In this case the trophozoites would encounter higher temperatures, and oxygen tension due to host inflammatory response. Parasite proteins that are triggered to counteract these stress conditions are of great interest, since these

would help us to understand the mechanism of pathogenesis. The ability of *E. histolytica* trophozoites to invade host tissues and to survive outside the protected environment of the intestine is probably accomplished by a strong adaptive response, involving a number of proteins, some of which may be stress induced. *E. histolytica* homologues of some of the known HSPs, such as HSP60 (20, 35) and HSP70 (23, 42) have been identified and partially characterized.

Here we report a novel polymorphic antigen in *E. histolytica* trophozoites, which is encoded by a multigene family and is differentially expressed in response to stress. The polymorphic part of the protein is almost exclusively composed of charged amino acids. The unique protein structure, the high copy number of the gene, and stress-induced regulation of expression of multiple copies point to an important role for this protein in parasite survival within the host.

MATERIALS AND METHODS

Entamoeba strains and culture conditions. All strains used in this study were obtained from the American Type Culture Collection. *E. histolytica* strains HM-1:IMSS clone 6 and HK-9 clone 2, *E. dispar* clone SAW760, were maintained in TYI-S-33 medium (12) at 36°C, in 15-ml glass tubes containing 13 ml of complete medium.

Genomic DNA isolation. DNA was purified essentially as described (6). At the end of the log phase of growth, cells were pooled from 40 tubes after chilling them in ice water for 10 min. A cell pellet was obtained by centrifuging at 275 × g for 7 min at 4°C. Cells (5 × 10⁷ to 10 × 10⁷) were resuspended in 5 ml of buffer (100 mM NaCl, 10 mM EDTA, 10 mM Tris-Cl [pH 8.0]) and lysed by addition of 0.25% sodium dodecyl sulfate (SDS). The resulting suspension was extracted with phenol-chloroform and DNA was collected by ethanol precipitation. It was treated with RNase A (100 µg/ml) followed by proteinase K (100 µg/ml). The suspension was again extracted with phenol: chloroform and ethanol precipitated. DNA was dissolved in TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

Isolation of chromosomal DNA, PFGE, Southern blotting and hybridization. *E. histolytica* chromosomal DNA from cells embedded in agarose blocks was prepared as described previously (3). The pulse conditions used for pulsed-field gel electrophoresis (PFGE) were 70 s for 12 h, 90 s for 6 h; 120 s for 6 h; 150 s

* Corresponding author. Mailing address: School of Life Sciences, Jawaharlal Nehru University, New Delhi-110067, India. Phone: 91-011-26704516. Fax: 91-11-6165886. E-mail: alok0200@mail.jnu.ac.in.

† Present address: Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Mass.

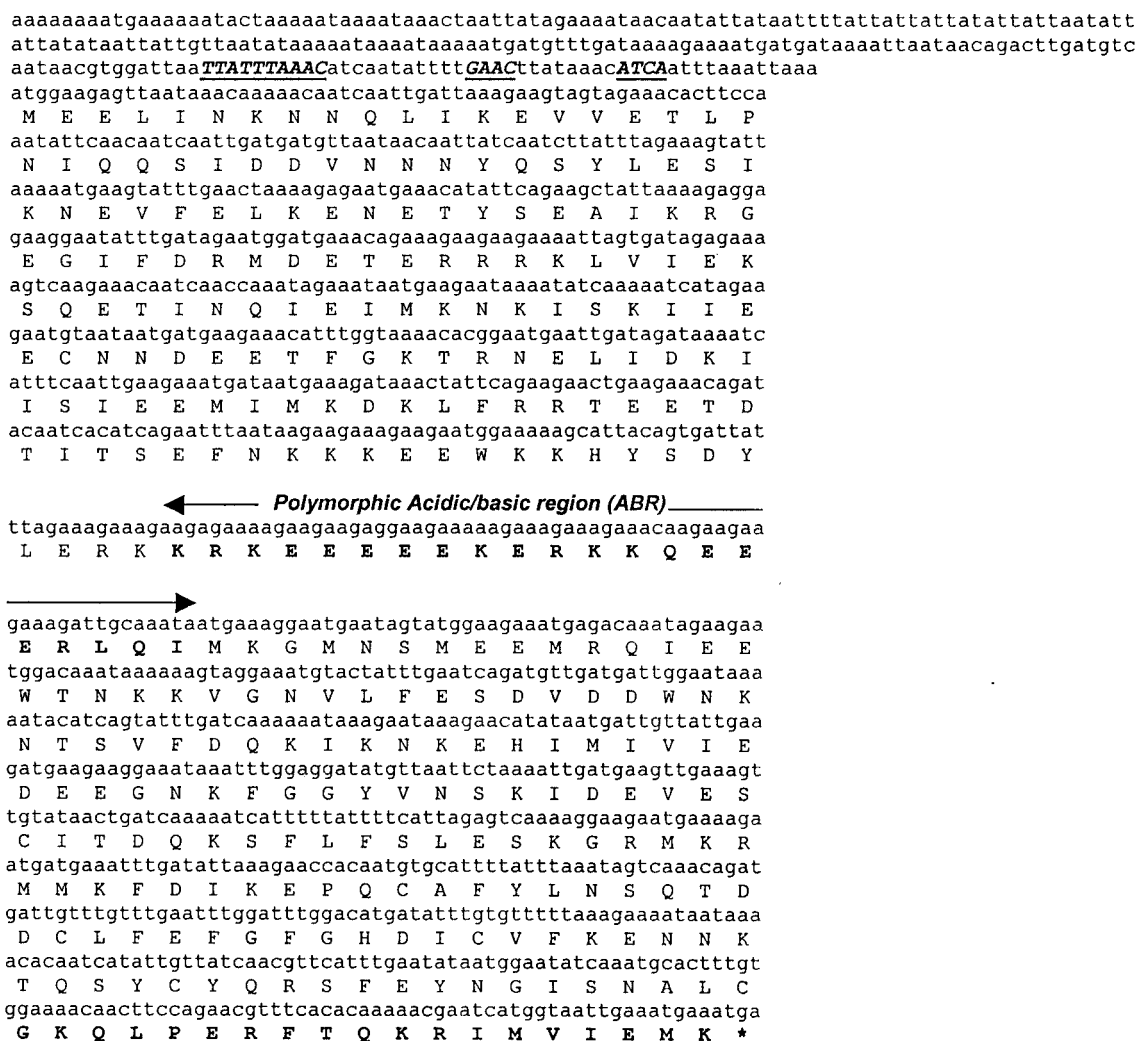


FIG. 1. Ehssp1-5076 gene sequence. Nucleotide sequence along with the amino acid sequence of the transcribed copy of Ehssp1-5076 in strain HM-1:IMSS is shown. The putative promoter consensus sequence, TTATTTAAAC, GAAC, and the transcription initiator ATCA are highlighted. The amino acids in the polymorphic acidic basic region are shown in boldface type. The stop codon is denoted by an asterisk.

for 6 h; and 170 s for 6 h at 5.5 V/cm. *Saccharomyces cerevisiae* (strain YNN6) blocks were used as molecular weight marker.

Genomic DNA (1 µg) was digested with 50 U of restriction enzyme, and separated on 1.0% agarose gel at 4 V/cm. For Southern blotting DNA was transferred to GeneScreen plus (NEN) nylon membranes (28). Blots were hybridized in a solution containing 1% SDS, 1 M NaCl and 3 × 10⁵ cpm ml⁻¹ of DNA probe at 65°C for 16 h. Blots were washed according to manufacturer's instructions and exposed for autoradiography. Radioactive DNA probes were made using the NEBlot random priming kit (NEB, USA).

PCR and RT-PCR. The following primers were used for reverse transcription (RT)-PCR: F2, 5' CGGGTACCATTAAAAATGGAAGAGCTAATTAAC 3'; R2, CGGATCCTATAAATCTTCTTGAAATTAATTTTGTCCATATGTTTCATTCAATTACTATAAT 3'; ABR-F, 5' GATGAAGAAACACTAAGTAAAACA 3'; ABR-R, 5' AACATATCCTCCAAATTTATTCC 3'; HSPF 5' AGGTATGGATCCAAATG 3'; HSPR 5' CTGCTTGTGCCGTTAAATCA 3'; CABPF, 5' ATCTGTTCTAAACATTAATCATAAACT 3'; CABPR, 5' GCGGGCTCCAGTTTAGAGTGAAAATC 3'. The Ehssp1 primers were designed based on the reference sequence (ENTJQ44). Primers were obtained from Microsynth, Switzerland. PCR was performed with 400 ng of *E. histolytica* genomic DNA for 30 cycles.

Total RNA (5 µg) was used in the RT reaction using M-MuLV RT (USB) with Oligo dT primer. The reaction was carried out at 37°C for 1 h followed by inactivation at 95°C for 5 min. Five microliters of this RT mix was used for a regular PCR. No amplicon was observed in the absence of RT enzyme in all

RT-PCR experiments. The PCR products were cloned in pGEMTeasy vector (Promega). DNA sequencing was carried out using dideoxy chain termination method (29).

RNA isolation from normal and stressed cells and Northern blot analysis. Total RNA was purified by guanidinium thiocyanate lysis of cells as described (10). RNA samples (20 µg) were resolved in formaldehyde agarose in formaldehyde gel running buffer [0.1 M MOPS (pH 7.0), 40 mM sodium acetate, 5 mM EDTA (pH 8.0)] and 37% formaldehyde at 3 to 4 V/cm. The RNA was transferred on to GS+ membrane. Hybridization and washing conditions for RNA blots were carried out as per manufacturer's protocol.

In order to isolate total RNA from *Entamoeba* cells after heat stress, mid log phase *Entamoeba* cells (approximately 8 × 10⁵ cells) grown in glass tubes were transferred to a water bath maintained at 42°C. At the various time points, the cells were chilled and RNA was extracted. Cells were exposed to oxidative stress by growing them in 10 ml of complete TYI-S-33 medium in a 50-ml tissue culture flask for 1 h at 36°C with gentle shaking at 40 rpm.

DNA dot blots and copy number estimation. DNA dot blots and copy number estimation were carried out as described (30). Briefly, *E. histolytica* strain HM-1:IMSS genomic DNA and purified Ehssp1 insert DNA were spotted in duplicates in various concentration ranging from 15 to 500 ng. The blot was hybridized using a radiolabeled Ehssp1 probe and the spots were cut out and the amount of radioactivity was measured by scintillation counting. Along with unknown genes the copy number of a few known genes, such as *E. histolytica* calcium binding

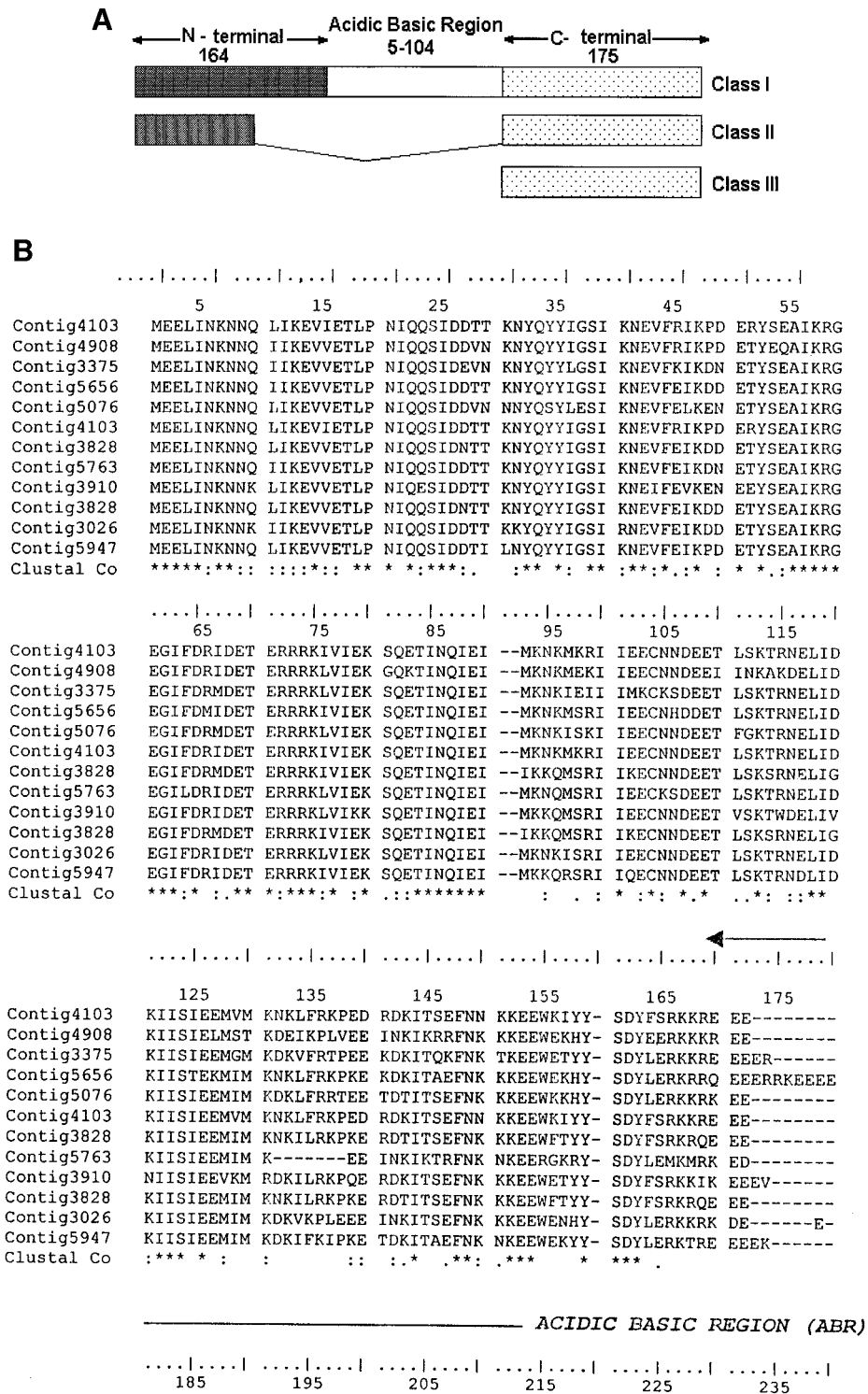


FIG. 2. Multiple alignment of different copies of Ehssp1 proteins showing polymorphism. (A) Schematic representation of the three classes of Ehssp1 proteins. The numbers indicate amino acids in each domain. (B) Multiple alignment of different copies of Ehssp1 protein. The polymorphic ABR is marked by arrow. The contigs used for alignment were obtained from the Sanger *Entamoeba* database.

```

Contig4103 ----- -ERRKQE-----
Contig4908 ----- -Y EEKERR-----
Contig3375 ----- -L----- -QKIN QEKEKRFKK K-NEIEKRLK
Contig5656 RRQQEEERR RRQQEEERR QEEEEERKQE QEKKIQEHEK KIQEYEKKIQ EHEKKIQEQE
Contig5076 ----- -E-EE KERKKQEEER -----LQ
Contig4103 ----- -ERRKQE-----
Contig3828 ----- -RRQQ EERKKQEGER -----LR
Contig5763 ----- -ERRKQE-----
Contig3910 ----- -KKQE EERKKQEQER KIQEHDRKIR ELEKKIQQHE
Contig3828 ----- -RRQQ EERKKQEGER -----LR
Contig3026 ----- -EKER-- -KRQE EERKIQEEER LRRESEDRVR
Contig5947 ----- -K----- -ERKE EERRKEKEKK K--QEGERLR
Clustal Co

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      245      255      265      275      285      295
Contig4103 ----- -QEERLQ RMRGLSTPKH QIKQLEEWTN RKVSNILFDS
Contig4908 ----- -KE-EE EMRKL EEWTN RKVGNILFDS
Contig3375 ----- -RKEMKS-QE EIRQIEEWTN RKVGNILFDS
Contig5656 RKTTEQEKKI QQLEKLRIIK EERRKEERLQ VMKGMST-IE EMLQIEEWTE RKVGNILFDS
Contig5076 ----- -IMKGMNS-ME EMRQIEEWTN KKVGNVLFES
Contig4103 ----- -QEERLQ RMRGLSTPKH QIKQLEEWTN RKVSNILFDS
Contig3828 ----- -IMKGMNT-ME EMLQLEEWTE KRINNILFDS
Contig5763 ----- -QEERLQ RMRGMSTPKY QIEQLEEWTE RKVGNVLFDS
Contig3910 KKIREVEEEEE RRKE----- EERKKQEEKL RMKGMNT-IE EMIQIEEWTE KRIGNILFDS
Contig3828 ----- -IMKGMNT-ME EMLQLEEWTE KRINNILFDS
Contig3026 ----- -R----- -K----- VMK--ET-ME EIQRIEWTN RKVGNILFDS
Contig5947 ----- -IMKGMNT-ME EMIQIEEWTN RRVEDILFDS
Clustal Co
      :*:***: :*: :*:**

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      305      315      325      335      345      355
Contig4103 DIDDWNQGTS VFDQRIMNKE HIIIIIEDEE GNKFGGYVNE KIDKIG---- YCGYINDSKS
Contig4908 NKDDWKVNTS VFDQRIMNKE HIIIIIEDTE GNKFGGYVNS KIDKDE---- --WISDSKS
Contig3375 DIDDWNKNTS VFYQRIMNKE HIIIIIEDED GNKFGGYVNS KIDEDE---- --WINDSKS
Contig5656 DKDDWNKNTS IFDKRIMNKE HLIIIEDEE GNKFGGYVDS KIDKVVDD--- --KIYDSKS
Contig5076 DVDDWNKNTS VFDQKIKNKE HIMIVIEDEE GNKFGGYVNS KIDEVES--- --CITDQKS
Contig4103 DIDDWNQGTS VFDQRIMNKE HIIIIIEDEE GNKFGGYVNE KIDKIG---- YCGYINDSKS
Contig3828 DKDDWNVNTS VFGERIKNKE HIIIIIEDEE GNKFGGYVNS KIDEVES--- --YIKDSKS
Contig5763 DIDDWKEDTS VFGERVLNKE HIIIIIKDSE GNKFGGYVNE KIDKYG---- YD-WINDSKS
Contig3910 DIDNWNINTS IFEERIWNKK HIIIIIKDSE GNKFGGYVNS KIDKVNNGGY- --WINDSQS
Contig3828 DKDDWNVNTS VFGERIKNKE HIIIIIEDEE GNKFGGYVNS KIDEVES--- --YIKDSKS
Contig3026 DIDDWNVSTS VFIQRIINKE HIIIIIEDEE GNKFGGYVNE KIDKVNE--- --CINDSKS
Contig5947 DIDNWNRTS VFNQRIKNKE DIIIIVKDED GNKFGGYVNS KIDKVG---- --WINDSQS
Clustal Co
      :*: * :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      365      375      385      395      405      415
Contig4103 FVFSLESNGR IEGMMKYDIE IPQHAFYLHN QSMNWSFAFG S---DIVVKG ENKKTTSYCS
Contig4908 FVFSLESNER NEGMMKFDIK DTKRAFWLCN QSHDCLFEFG W--RDICIYK ASYKTLSSQC-
Contig3375 FVFSLESNGR MKGMKKFDIK QPEYAFCLFN QTVNCLFAFG FG-HDIFVCK ENDKTSLYC-
Contig5656 FLFSLESKGR MKGMKKFDIK QPEYAFKLYN QSDDRLFVFG YG--DIFVYK ENYKTKSFC-
Contig5076 FLFSLESKGR MKRMMKFDIK EPQCAFYLSN QTDDCLFEFG FG-HDICVFK ENNKTQSYC-
Contig4103 FVFSLESNGR IEGMMKYDIE IPQHAFYLHN QSMNWSFAFG S---DIVVKG ENKKTTSYCS
Contig3828 FVFSLESKGR MKGMKKFDIK QPQHAFHLEN QTLDCLFEEIG ID--DIAIYK ENDKTKSLC-
Contig5763 FVFSLESNGR IEGMMKFDIK QPQAFYLHN QSCDCLFGFG NG-HDIRVCK ENDKTKSWC-
Contig3910 FVFSLESKGR LKGMKKFDIK RPEFAFYLFN QSDDRLFEEFG HGYSIDLVC ENNKTQSYC-
Contig3828 FVFSLESKGR MKGMKKFDIK QPQHAFHLEN QTLDCLFEEIG ID--DIAIYK ENDKTKSLC-
Contig3026 FVFSLESNGR MKVMKKFDIK KPQHAFWLCN QSDNYLFLKG NN--DIVVFK ENNKTISYC-
Contig5947 FVFSLESNGR ---MKKFDIK QPQDAFYLIN QSNDCLEFFG LG--DICVYK ENNKTISHC-
Clustal Co
      *:***: * :*: :*: :*: :*: :*: :*: :*: :*: :* :* :* :* :* :* :* :* :* :* :*

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      425      435      445      455
Contig4103 QQHSYFYDGI ENALCGKQH- ---FTPKRII VIEMK--
Contig4908 EQNSFEYNGI SNALCG---- KYFFTPKQII VIEMK--
Contig3375 SQRSFEYEGI SNALCG---- KQFFTPKRII VIEMK--
Contig5656 KQHSFSYEEI SNALCG-NY- QYRFTPKRFI VIEMK--
Contig5076 YQRSFEYNGI SNALCGQL- PERFTQKRIM VIEMK--

```

FIG. 2—Continued

```

Contig4103  QQHSYEDGI ENALCGKQH- ---FTPKRRII VIEMK--
Contig3828  KQHSFEYKGI SKALCGKQF- PERFTPKRRII VIEMIQS
Contig5763  EQCSFEYKGI SNTLCGKLYP D-RFTPKRRII VIEMK--
Contig3910  NQYSFEYEGI SNAKKGKQH- PNHFTPKRRII VIEMK--
Contig3828  KQHSFEYKGI SKALCGKQF- PERFTPKRRII VIEMIQS
Contig3026  KQHSFSYKGI TNALCGKQL- PERFTPKRRII VIEMK--
Contig5947  NQHSFEYKGI ENALCG---- KQHFTPKRRII VIEMK--
Clustal Co  * : : * . *   : : *   * * * : : * * * *
    
```

FIG. 2—Continued

protein (copy number 1) was always estimated. For calculation of copy number the size of haploid genome of *E. histolytica* was taken to be 20 Mb.

Cloning, expression, and partial purification of recombinant Ehssp1 protein. DNA was obtained by RT-PCR amplification of total RNA from exponentially growing *E. histolytica* cells, using the F2/R2 primers. The *Acc65I* and *BamHI* sites present in the primers were utilized to clone the product in pET 30a vector, digested with the same enzymes. The ligated product was used to transform *Escherichia coli*, BL21(DE 3) competent cells to get pETEhssp1.

Cells containing pETEhssp1 were grown to an OD₆₀₀ of 0.6, induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) and harvested after 180 min by centrifugation at 6,000 rpm (Remi tabletop centrifuge) for 5 min. The cell pellet was resuspended in 1× phosphate-buffered saline and lysed by sonication (five full pulses for 60 s, with a gap of 120 s). The pellet containing the Ehssp1 protein in inclusion bodies was collected by centrifugation at 15,000 rpm for 20 min, at 4°C.

SDS-PAGE and Western blot analysis. Protein samples were electrophoresed in SDS-PAGE as described (17). Proteins were transferred from gels onto nitrocellulose membranes (36). The blots were immunostained using pooled amebic patient and normal human sera, at a dilution of 1:400, followed by detection by alkaline phosphatase labeled anti-human secondary antibody. Nonspecific blocking was carried out with 3% bovine serum albumin in phosphate-buffered saline. Nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) in buffer containing 0.1 M sodium bicarbonate and 1 mM MgCl₂ was used for color development.

Sequence analysis. BLAST (9) searches (version 2.0 at www.ncbi.nlm.nih.gov/BLAST at NCBI, USA and WU-BLAST version 2.0 at www.sanger.ac.uk

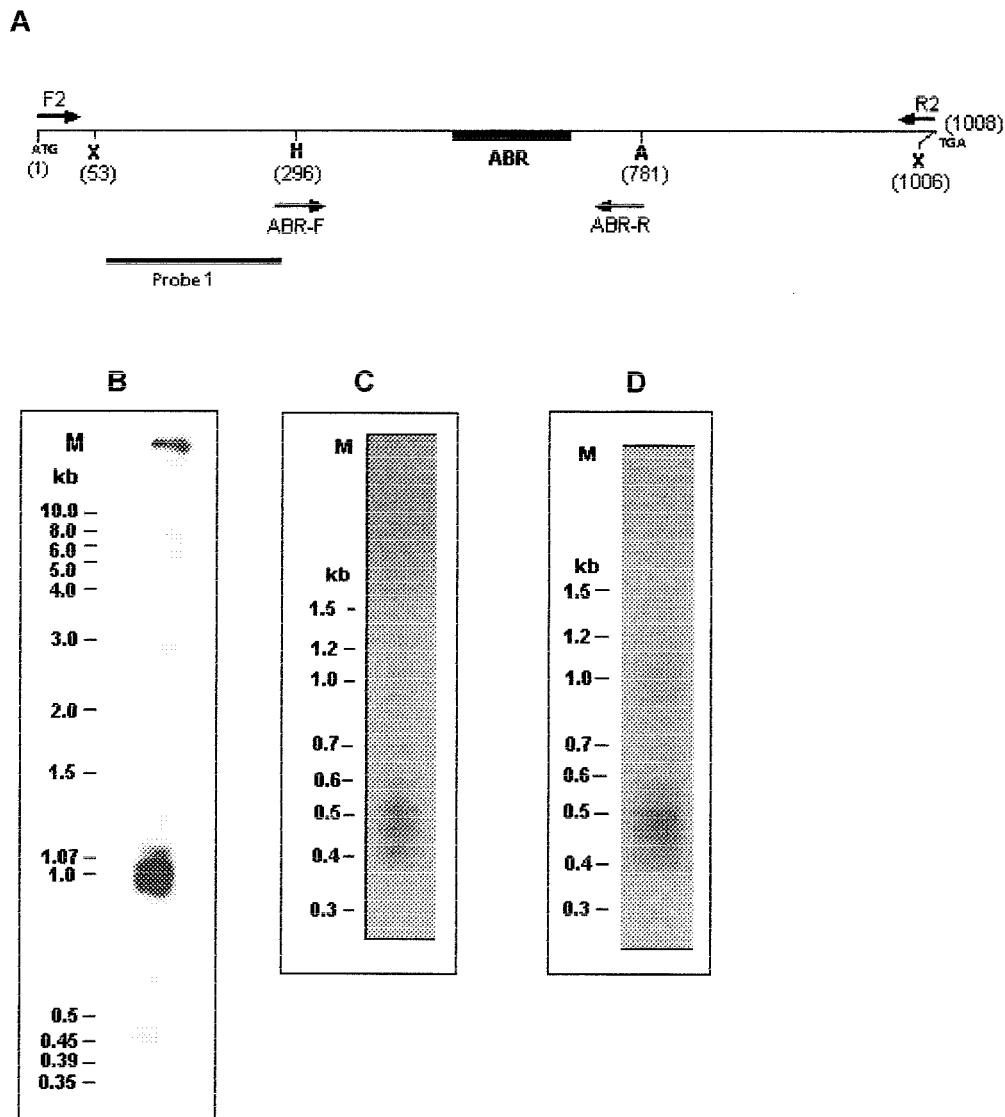


FIG. 3. Polymorphism in Ehssp1. (A) Schematic representation of Ehssp1-SSE58 showing the location of PCR primers and DNA probe. Numbers indicate nucleotide positions. Abbreviations: X, *XmnI*; H, *HinfI*. (B) Southern blot analysis. Genomic DNA of strain HM-1:IMSS was digested with *XmnI*, electrophoresed in 1% agarose gel at 4V/cm and transferred to nylon membrane. The blot was hybridized with probe 1. (C) PCR amplification of genomic DNA of strain HM-1:IMSS using ABR-F-ABR-R primer set (D) PCR amplification of genomic DNA of strain HK-9 using ABR-F-ABR-R primer set. The PCR products were electrophoresed in a 1.2% agarose gel at 4V/cm. M, molecular weight markers.

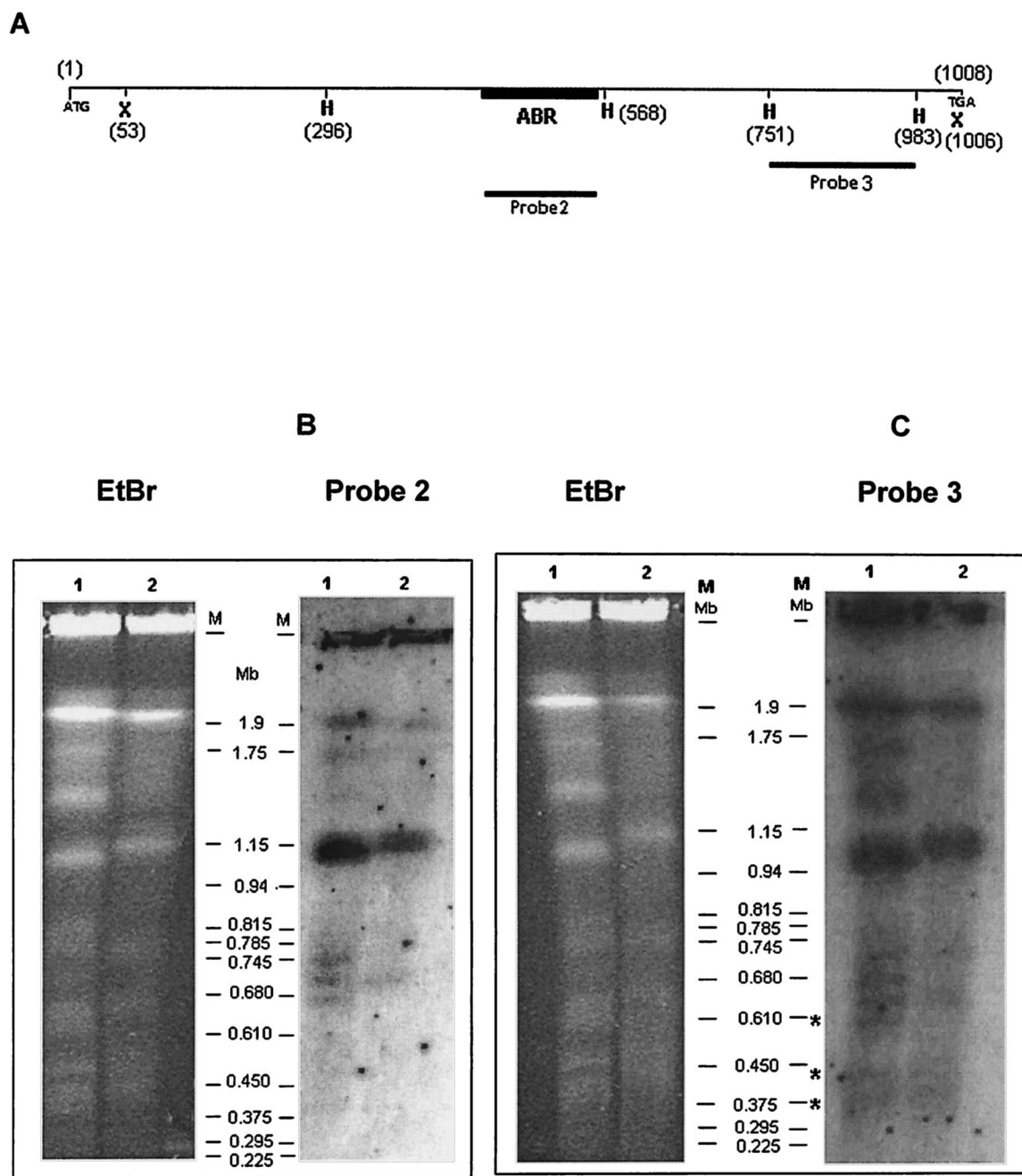


FIG. 4. Chromosomal localization of Ehssp1. (A) Schematic representation of Ehssp1-SSE58 showing the position of DNA probes. The probes used for hybridization were derived from SSE58. Numbers indicate nucleotide positions. Abbreviations: X, *Xmn*I; H, *Hinf*I. (B and C) PFGE-separated chromosomes of strain HK-9 (lane 1) and HM-1:IMSS (lane 2) stained with ethidium bromide (EtBr) and hybridized with probe 2 (B) and probe 3 (C). The bands of strain HK-9 that differentially hybridized with probe 3 are shown by an asterisk. The PFGE run conditions are mentioned in Materials and Methods. Autoradiographic exposure of Southern blots was for 3 days.

/Projects/E_histolytica/blast_server.shtml *Entamoeba* BLAST server, Sanger Centre, Hinton, Cambridge, United Kingdom), respectively, were used for finding homologues of SSE58 using the *Entamoeba* genome sequences deposited regularly in the GSS division of GenBank. The sequencing effort is part of the International *Entamoeba* Genome Sequencing Project and is supported by award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. *Entamoeba* genome sequence contig databases from TIGR and the Sanger Centre were also used for analysis. Alignment of multiple sequences was carried out using ClustalW.

Nucleotide sequence accession numbers. The nucleotide sequences mentioned in the paper have been deposited in the GenBank database with accession numbers, AF322253, AY278996, AY281154, and AY281155.

RESULTS

Identification and sequence analysis of Ehssp1. In a screen for repetitive DNA, a 2.57-kb sequence (SSE58; accession

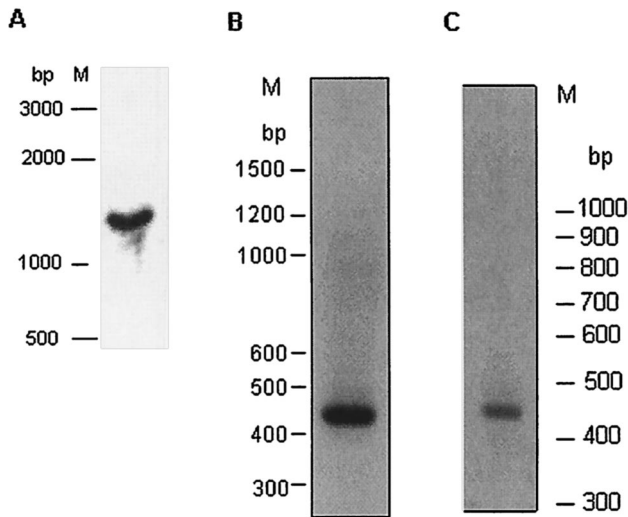


FIG. 5. Analysis of Ehssp1 transcript. (A) Northern blot analysis. Total HM-1:IMSS RNA (30 µg) was electrophoresed in 1.2% formaldehyde agarose gel at 4 V/cm, and transferred to nylon membrane. The blot was hybridized with probe 3 (shown in Fig. 3). M, RNA molecular weight marker. RT-PCR analysis of (B) HM-1:IMSS and (C) HK9 RNA using ABR-F-ABR-R primers (Fig. 3). RT-PCR products were electrophoresed on 1.2% agarose gel at 4 V/cm. M, DNA molecular weight markers.

number AF322253) was identified from the *E. histolytica* genomic library. It contained a partial open reading frame (ORF) towards one end. The ORF sequence was used to search the Sanger *Entamoeba* genome database for homologous sequences. Forty-five different contigs from the database that contained the ORF were aligned to assign the start and stop codons for the ORF. The nucleotide sequence of the full-length ORF, was named Ehssp1 (*E. histolytica* stress-sensitive protein family one).

The sequences, upstream and downstream of the Ehssp1 gene were analyzed from a contig (contig 5076, Sanger *Entamoeba* sequence database) containing the full-length Ehssp1 ORF, for the presence of cis regulatory elements (Fig. 1). The TATA element, GAAC sequence and Initiator sequence described for several *Entamoeba* genes (7, 27) could be identified upstream of the putative start codon. Sequences further up-

stream were particularly rich in A+T (84%) compared to 78% A+T in other intergenic regions of *E. histolytica* (5).

Multiple alignment of different copies of Ehssp1 have length polymorphism, which is mainly restricted to a purine-rich stretch towards the middle of the ORF. In most copies (29 out of 45), a continuous reading frame was maintained and the polymorphism did not lead to frame shift mutations.

Ehssp1 protein. BLASTP analysis failed to show any homologue of this gene in the nonredundant database. The encoded protein could be divided into three regions, with a central polymorphic domain (containing 5 to 104 amino acids) separating the N-terminal domain (164 amino acids in most copies with a range of 157 to 167 amino acids) from the C-terminal domain (175 amino acids in most copies with a range of 165 to 181 amino acids) and there were three distinct classes of Ehssp1 that could be seen by sequence analysis (Fig. 2A). Multiple alignment of 21 contigs containing the full-length ORF showed that the N- and C-terminal domains of the various copies were 86% and 74% identical, respectively, at the amino acid level. The polymorphic domain, encoded by a polypurine stretch, was predominantly composed of acidic/basic amino acids (glutamate, arginine and lysine) (Fig. 2B).

The protein encoded by the different gene copies had a theoretically calculated pI range from 5.12 to 8.53 (depending on the length and composition of the polymorphic region). Perfect repetitive motifs, as found in the acidic-basic repeat antigen (ABRA) of *Plasmodium falciparum* (41), were absent in the polymorphic stretch of this protein. The protein was highly hydrophilic with a small hydrophobic stretch towards the C terminus. From amino acid sequence, the protein seemed to lack any trans-membrane helices.

Confirmation of polymorphism in Ehssp1 by genomic Southern and PCR. From sequence analysis of contigs it was predicted that ~90% of Ehssp1 gene copies would fall in the size range of 0.95 to 1.15 kb. This was demonstrated by Southern hybridization. DNA was digested with *Xmn*I, which cuts near the ends of the Ehssp1 gene (Fig. 3A). Hybridization of the Southern blot with Ehssp1 probe showed a broad band with a size range of 0.9 to 1.07 kb (Fig. 3B). Genomic DNA was PCR amplified using primers flanking the ABR (ABR-F and ABR-R). This primer set amplified a cluster of three bands in the size range of 400 to 550 bp and one faint band of 574 bp

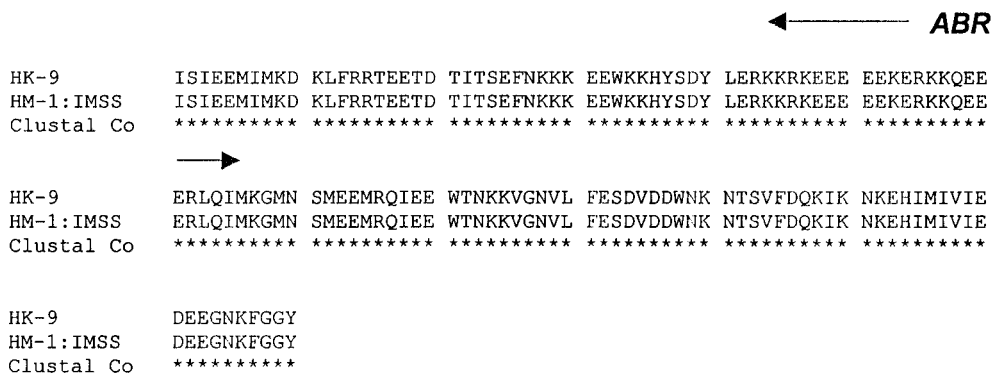


FIG. 6. Alignment of the Ehssp1 protein expressed by exponentially growing cells of the *E. histolytica* strains HK-9 and HM-1:IMSS. The ABR is indicated with an arrow.

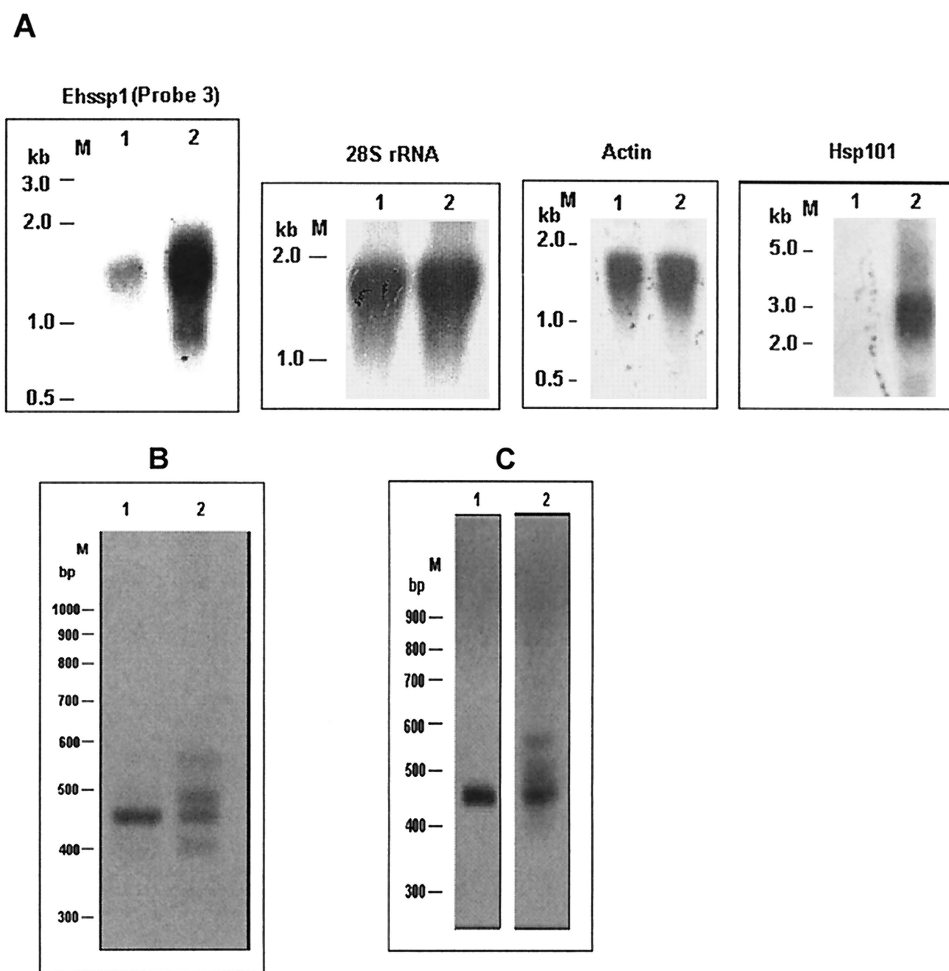


FIG. 7. Transcription of *Ehssp1* under stress conditions. (A) Northern blot analysis of RNA from HM-1:IMSS cells grown at 36°C (lane 1); and transferred to 42°C for 1 h (lane 2). Various gene probes were used as indicated on top of each panel. Actin and 28S rDNA probes also served as loading controls. Hsp101 was used as positive control for heat shock treatment. (B) RT-PCR using the ABR-F-ABR-R primer set (Fig. 3), was carried out with total RNA from normal cells (lane 1) and heat treated cells (lane 2). (C) same as panel B except that RNA was from strain HK-9.

(Fig. 3C). The bands were diffuse, suggesting size heterogeneity. The PCR amplification pattern for strain HK-9 was very similar to that of strain HM-1:IMSS (Fig. 3D), suggesting that the organization of this gene family is similar in different strains of *E. histolytica*.

Chromosomal location of *Ehssp1* gene. Southern blots of PFGE-separated chromosomes of *E. histolytica*, strains HM-1:IMSS and HK-9 were hybridized with probes 2 (Fig. 4B) and 3 (Fig. 4C) derived from SSE58. Both probes hybridized with a number of bands, of which the most intense was a 1.1 Mb band in either strain. Probe 3 gave a strong signal with a 1.9 Mb band also which hybridized poorly with probe 2. In addition, some fainter bands were more in number in HK-9. Three of these (marked by asterisks in Fig. 4C) gave no detectable signal with probe 2 (confirmed by densitometric scanning). The bands that differentially hybridized with probe 3 could contain copies of the gene with very short ABRs. The number of copies of *Ehssp1* per haploid genome was estimated to be 306 using a quantitative hybridization method as described before (30). The copy number of *E. histolytica* calcium binding protein was

found to be one using the same method. This protein is known to be present as a single copy gene in *E. histolytica* genome (43). Many of the *Ehssp1* copies did not contain full-length coding region and had truncation mainly at the 5'-end (data not shown here). Thus, the gene is highly repetitive and present on multiple chromosomes.

Transcription of *Ehssp1*. Transcription was analyzed by northern hybridization and RT-PCR. Northern hybridization (Fig. 5A) gave a single band of 1.4 kb with probe 3 (shown in Fig. 4).

RT-PCR analysis of total RNA was carried out with the primer set ABR-F-ABR-R (shown in Fig. 3). A discrete band of 450 bp was observed (Fig. 5B), which indicates that only one or a few copies of the polymorphic gene are transcribed. Similarly RT-PCR with ABR-F-ABR-R primers displayed a single band of 450 bp (Fig. 5C). In order to identify the transcribed genomic copy, the RT-PCR product from strain HM-1:IMSS was cloned and five randomly picked clones were sequenced. The sequences of all five were identical, suggesting that only a single copy of the *Ehssp1* gene family is expressed in exponen-

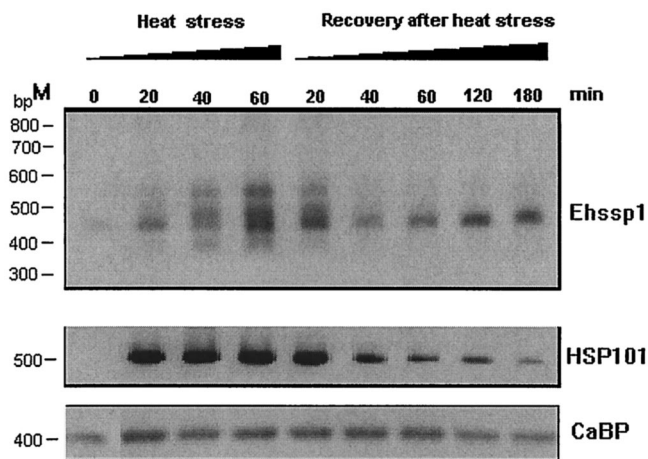


FIG. 8. Time course of reversible induction of Ehssp1 genes during heat stress. HM-1:IMSS cells growing at 36°C were transferred to 42°C. At the time points indicated, growth was stopped and RNA was isolated. To study the recovery from heat stress, cells grown at 42°C for 60 min were reverted to 36°C and RNA was isolated at different time intervals. RT-PCR was carried out using ABR-F-ABR-R primers. As controls, RT-PCR was carried out with Ehssp101 primers (HSP-F/HSP-R, middle panel) and EhCaBP primers (CABP-F/CABP-R, bottom panel). The PCR products were electrophoresed in 1.2% agarose gel at 4 V/cm.

tially growing trophozoites. The sequence of the expressed copy was used to search the *Entamoeba* contig database at the Sanger Centre, and a contig with 100% identity was obtained (contig 5076). The complete sequence identity of the RT-PCR product with the ORF in contig 5076, shows that the Ehssp1 gene has no introns. The expressed copy was 1,077 nucleotides long, encoding a protein of 359 amino acids. The sequence of the expressed copy along with the flanking sequences is shown

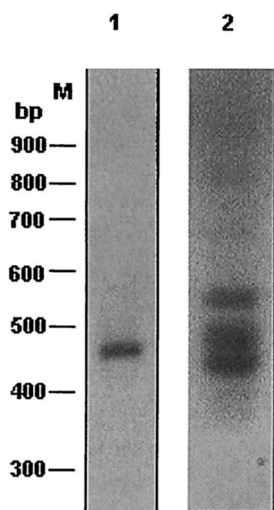


FIG. 9. Effect of oxidative stress on the expression of multiple copies of Ehssp1. HM-1:IMSS cells were exposed to oxidative stress for 1 h as described in Materials and Methods. Total RNA was used for RT-PCR analysis with ABR-F-ABR-R primers. The PCR products were resolved in 1.2% agarose gel at 4 V/cm. Lane 1, normal cells; lane 2, cells after oxidative stress.

above. The amino acid sequence of the protein encoded by the RT-PCR product of strain HK-9 (using ABR-F-ABR-R primers) showed 100% identity in the ABR domain (Fig. 6).

Expression of Ehssp1 under conditions of growth-stress. Since only a single copy of the polymorphic Ehssp1 gene was expressed in exponentially growing cells, we decided to check for conditions under which other copies may be expressed. Various conditions of growth-stress were tried for this purpose. *E. histolytica* cells (normally grown at 36°C) were subjected to heat stress at 42°C for 1 h and expression of *Ehssp1* was analyzed by Northern blot analysis and RT-PCR. Northern hybridization showed a significant increase (4.7-fold) in expression of *Ehssp1* after heat shock, and the RNA band was very broad compared to that from normal cells (Fig. 7A). The expression of a putative heat shock protein hsp101 was checked under the same conditions to confirm the heat shock treatment (Fig. 7A). The hsp101 probe hybridized only with the RNA from heat-shocked cells. The broad band observed in Northern blot analysis using Ehssp1 gene probe could be due to abnormal addition of poly(A) tail after heat shock, as reported for some heat shock genes (21, 24). Alternately, it could be due to expression of other copies of Ehssp1. To resolve this, RT-PCR was performed using the ABR-F and ABR-R primers flanking the polymorphic region. Multiple bands were obtained with RNA from heat-treated cells (Fig. 7B, lane 2) compared to RNA from normal cells (lane 1). A similar RT-PCR profile was observed for the *E. histolytica* strain HK-9 (Fig. 7C). Therefore, it appears that multiple polymorphic copies of *Ehssp1*, which were silent in exponential cells are transcribed during heat stress. This was confirmed by cloning the RT-PCR product and determining the nucleotide sequence of four clones. Multiple alignment of these clones showed that each clone had a distinct sequence (contigs 5247, 533, and 2483) and the polymorphism was very pronounced in the ABR (data not shown here). The sequence of one of these clones (hs4) was identical with the copy expressed in exponentially growing cells indicating that this copy is also expressed under heat stress.

RT-PCR was also used to study the kinetics of induction of silent copies of Ehssp1 during heat shock. The results showed that within 20 min of heat shock the expression of heat-inducible copies could be discerned (Fig. 8). The level of expression increased for 60 min. Heat treatment beyond 60 min was lethal for the cells. Compared with that of Ehssp1, the expression of HSP101 reached a maximum within 20 min of heat shock. When cells were allowed to recover after a 60-min heat shock, the expression of the heat-inducible copies of Ehssp1 declined, with a concomitant increase in expression of a single 450-bp band (Fig. 8), which corresponds to the copy expressed in exponential cells (Fig. 7B, lane 1). This was confirmed by sequencing two independent clones of the RT-PCR product of the RNA expressed after recovery from heat stress.

Entamoeba is microaerophilic, and excess oxygen is toxic to this organism. Therefore, RT-PCR using ABR-F and ABR-R primers was done to examine the effect of oxidative stress on the expression of Ehssp1. The results showed multiple bands similar to those obtained under heat shock conditions (Fig. 9).

Organization of Ehssp1 in *E. dispar*. Of all *Entamoeba* species that dwell in the human gut, *E. histolytica* is the only one that is pathogenic. Of particular interest is *E. dispar*, which is a

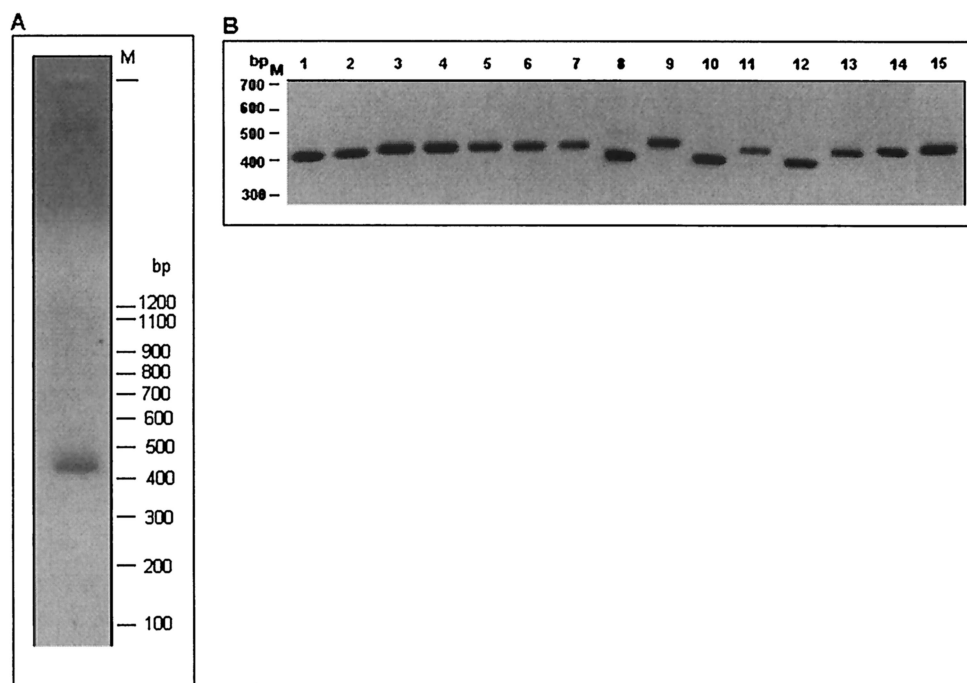


FIG. 10. Ehssp1 homologue in *E. dispar*. (A) PCR amplification of *E. dispar* genomic DNA using ABR-F-ABR-R primers. The PCR product was electrophoresed in 1.2% agarose gel at 4 V/cm. (B) The PCR product in (A) was cloned in pGEMT-Easy vector. Clones were randomly picked and the inserts released by digesting with *Eco* RI. The digests were electrophoresed on a 1.2% agarose gel. The portion of the ethidium bromide-stained gel containing the inserts is shown.

nonpathogenic sibling species of *E. histolytica*. To check the status of Ehssp1 in this nonpathogenic species, PCR amplification was carried out using the genomic DNA of *E. dispar*. Primer sets flanking the ABR as well as those flanking the full-length of the ORF, amplified a single band, showing that the gene is, indeed, present in *E. dispar*. The primer set ABR-F/R amplified a band of 450 bp (Fig. 10A). The single band is, however, in contrast to the broad bands from *E. histolytica* with the same primers (Fig. 3C). To determine the degree of polymorphism of this gene in *E. dispar*, the PCR product was molecularly cloned and insert sizes of 15 randomly selected clones were determined (Fig. 10B). The cloned fragments varied in size from 0.41 to 0.49 kb. These could be classified into three size classes: 410 bp (4 clones), 435 to 450 bp (10 clones), and 490 bp (1 clone). The extent of polymorphism was further determined by nucleotide sequencing of seven clones belonging to the various size classes. Multiple alignment of the seven sequences showed that the size variation was mainly restricted to the ABR region (Fig. 11) with minor variations outside the ABR. Thus, size polymorphism in this gene family exists in *E. dispar* but is much less pronounced than in *E. histolytica*.

Ehssp1 is immunogenic in amebiasis patients. Full-length cDNA corresponding to the normally expressed copy of *Ehssp1* (contig 5076) was obtained by RT-PCR of total RNA using the F2/R2 primer set (shown in Fig. 3) The cDNA was cloned in the *Bam*HI/*Acc*65I sites of pET30a to get pETEhssp1, with a six-histidine tag at the N terminus. Upon induction with 1 mM IPTG, a 43-kDa band was observed (Fig. 12A, lane 1). The size of the band matched with the expected size deduced from the *Ehssp1* ORF (359 amino acids) and the His tag (38 amino

acids). The recombinant Ehssp1 protein was found in inclusion bodies. It could be identified by immunostaining with anti-His tag antibody. Western blot analysis showed that the Ehssp1 protein could bind antibodies present in pooled sera from invasive amebiasis patients but not those in normal individuals (Fig. 12B). The data suggest that Ehssp1 is immunogenic in amebiasis patients.

DISCUSSION

By virtue of the specialized niche that they occupy in their hosts, parasites are expected to have evolved novel molecules not encountered in free-living organisms. Such molecules are interesting from an evolutionary stand point and are likely to perform vital functions related to pathogenesis. Here we report a protein antigen, Ehssp1 in *E. histolytica*. This protein is of unique construction. It consists of relatively conserved N- and C-terminal regions flanking a central, size-polymorphic domain (ABR) of 5 to 104 amino acids. Apart from being size-polymorphic, the central domain is notable in being composed almost entirely of acidic and basic amino acids. The protein is encoded by a multigene family, of which only one (or a few) members are transcribed in exponentially growing *E. histolytica* trophozoites, while a large number of copies are stress-induced. At the nucleotide level, the ABR is composed predominantly of purines (83 to 95% A+G). Such a polypurine stretch coding for acidic and basic amino acids is also found in the ABRA protein of *P. falciparum* (14, 16). The function of this charged domain (which is not polymorphic) and of ABRA itself is not clear; although ABRA has a pro-

(A)

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      5          15          25          35          45          55
D1    TGATGAAGAA ACCTAAGTA AAACACGGAA TGAATAATC AATAAAATCA TTTC AATTGA
D6    TGATGAAGAA ACCTAAGTA AAACACGGAA TGAATTAATC AATACAATCA TTTC AATTGA
D9    TGATGAAGAA ATACTAAGTA AAACACGGAA TGAATAATC AATAAAATCA TTTC AATTGA
D10   TGATGAAGAA ACCTAAGTA AAACATGGAA TGAATAAACC AATAAAATCA TTTC AAAAAGA
D11   TGATGAAGAA ACCTAAGTA AAACAAAGAA TAAACTAATA GACAAAATCA TTTC AACTGA
D13   TGATGAAGAA ATACTAAGTA AAACACGGAA TGAATTAATC AATACAATCA TTTC AATTGA
D14   TGATGAAGAA ATACTAAGTA AAACACGGAA TGAATTAATC AATACAATCA CTTC AATTGA
Ehssp1 TGATGAAGAA ACATTTGGTA AAACACGGAA TGAATTGATA GATAAAATCA TTTC AATTGA
Clustal Co ***** * * * * * ***** * * * * * * * * * * * * * * * * *

```

```

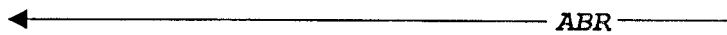
.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      65          75          85          95          105         115
D1    AGAAATGAAC ATGAAAGATA AACTATTCAG AACACCAAAA GAAAAAGATA AAATCACCCA
D6    AGAAATGATA ATGAAAGATA AAATATTTAG AAAACCAAAA GAAAAAGATA AAATCACCCA
D9    AGAAATGGTA ATGAAAGATA AACTATTCAG AACACCAGAA GAAAAAGATA AAATCACGGA
D10   AGAAATGATA ATGAAAGATA AACTATTCAG AAGACTAAAA GAAACAGATA AAACCACAGA
D11   AGAAATGATC AAGAAAGATA AAATATTCAG AAAACCAAAA GAAAAAGATA AAATCACAGA
D13   AGAAATGATA ATGAAAGATA AAATATTTAG AAAACCAAAA GAAAAAGATA AAATCACCCA
D14   AGAAATGATA ATGAAAGATA AAATATTTAG AAAACCAAAA GAAAAAGATA AAATCACCCA
Ehssp1 AGAAATGATA ATGAAAGATA AACTATTCAG AAGAACTGAA GAAACAGATA CAATCACATC
Clustal Co ***** * ***** ** * * * * * * * * * * * * * * * * * * * *

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      125         135         145         155         165         175
D1    AGAATTTAAT AAAAAGAAAG AAGAATGGAA AACATATTAT AGTGATTATT TTGAGAGGAA
D6    ATCATTTAAT AAAAATAAAG AAGAATGGAA CACATATTAC AGTGATTATT TTGAAAGGAA
D9    AGAATTTAAT AAAAAGAAAG AAGAATGGGA AACATATTAC AGTGATTATT TTGAGATGAA
D10   AGAATTTAAT AAAAAGAAAG AAGAATGGAA -ACACATTA- --TGATTATT TTGAGAAGAA
D11   ATCATTTAAT AAAAAGAAAG AAGAATGGAA AATACATTAT AGTGATTATT TTGAGAGGAA
D13   ATCATTTAAT AAAAATAAAG AAGAATGGAA CACATATTAC AGTGATTATT TTGAAAGGAA
D14   ATCATTTAAT AAAA-TAAAG AG--ATGGAA CACATATTAC AGTGATTATT TTGAAAGGAA
Ehssp1 AGAATTTAAT AAGAAGAAAG AAGAATGGAA AAAGCATTAC AGTGATTATT TAGAAAGAAA
Clustal Co * ***** ** * * * * * * * * * * * * * * * * * * * * *

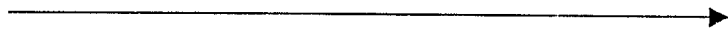
```



```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      185         195         205         215         225         235
D1    AAAGAAAAAA GAAGAAA--- ---A-----
D6    GAAGAAACAA GAAGAAGA--
D9    GAAGAAAAAA GAAGAAAAGA ^GAAACAAGA AGAAGAAAAG AAGAGAAAAG AGGAAGAAGA
D10   GAAGAGAAA   GAAAAAG---
D11   GAAGAAAAAA GAAGAAGA--
D13   GAAGAAACAA GAAGAAGA--
D14   GAAGAAACAA GAAGAAGA--
Ehssp1 GAAGAGAAA   GAAGAAGA--
Clustal Co **** * * * * *

```



```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      245         255         265         275         285         295
D1    ----- --A-----G AAATAAAGAA ACAAGAAGGA GAAAGATTAA GAAC-AATGA
D6    GAGAAGAAAA AAA---GAAG AAAGAAAGAA ACAAGAAGGA GAAAGATTAA GAAT-AATGA
D9    GAGAAGAAGA AAAGAAGAAG AAAGAAAGAA ACAAGAAGGA GAAAGATTAA GAAT-AATGA
D10   -----
D11   ---AAAAGA AAA--AGAGA GAAAAAAACA AGAAGAAGAA GGTAGATTAC AAAT-AATGA
D13   GAGAAGAAAA AAA---GAAG AAAGAAAGAA ACAAGAAGGA GAAAGATTAA GAATCAATGA
D14   GAGAAGAAAA AAAA--GAAG AAAGAAAGAA ACAAGAAGGA GAAAGATTAA GAAT-AATGA
Ehssp1 --G--GAAGA AAA-----AG AAAGAAAGAA ACAAGAAGAA GAAAGATTGC AAAT-AATGA
Clustal Co * * * * * * * * * * * * * * * * *

```

FIG. 11. (A) Alignment of nucleotide sequence of genomic PCR of *E. dispar* using ABR-F-ABR-R primers, along with the normally expressed copy of *E. histolytica* (Ehssp1). (B) The Clustal alignment of four clones showing the degree of variation in the ABR region. The ABR is shown with an arrow.

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      305      315      325      335      345      355
D1    AAGGAATG-- AATAC-AAT GAAGAAATTC AACAACTAGA AGAATGGACA AATAGAAAAG
D6    AAGGAATG-- AATAC-AATG GAAGAAATGA AACAACTAGA AGAATGGACA AATAGAAAAG
D9    AAGGAATG-- AATAC-AATG GAAGAAATTC AACAAATAGA AGAATGGACA GAAAGAAGAA
D10   AAAGAAGA-- AACAA----- GAAGAAATTC AACAAATAGA AGAGTGGACG TCAAGAAGAG
D11   GAGGAATATC AATACC AAAA CACCAAATTA AACAACTAGA AGAATGGACA TCAAGAAGAA
D13   AAGGAATG-- AATAC-AATG GAAGAAATGA AACAACTAGA AGAATGGACA AATAGAAAAG
D14   AAGGAATG-- AATAC-AATG GAAGAAATGA AACAACTAGA AGAATGGACA AATAGAAAAG
Ehsspl AAGGAATG-- AATAG-TATG GAAGAAATGA GACAAATAGA AGAATGGACA AATAAAAAAG
Clustal Co * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      365      375      385      395      405      415
D1    TATCCAATAT AATATTGGAT TCAGATGTTG ATAATTGGAA -TCAGAATAC ATCAGTCTTT
D6    TATCCAATAT ATTATTTGAT TCAGATATTG ATAATTGGAA -TATTAATAC ATCAGTACTT
D9    TATCCAATAT ACTATTTGAT TCAGAAATTC ATGATTGGAA -TAAGAATAC ATCAGTATTT
D10   TATCCAATAT ATTATTTGAT TCAGATATTG ATGATTGGGG GTGTTTATAC AACAAATATT
D11   TATCCAATAT ACTATTTGAT TCAGAGATTG ACGATTGGGG -TGTTTATAC ATCAGTATTT
D13   TATCCAATAT ATTATTTGAT TCAGATATTG ATAATTGGAA -TATTAATAC ATCAGTACTT
D14   TATCCAATAT ATTATTTGAT TCAGATATTG ATAATTGGAA -TATTAATAC ATCAGTACTT
Ehsspl TAGGAAACGT ACTATTTGAA TCAGATGTTG ATGATTGGAA -TAAAAATAC ATCAGTATTT
Clustal Co ** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      425      435      445      455      465      475
D1    AATAAAAGAA TAATAAATAA AGAGCATATA ATAATTATTA TTGAAGATAT AAATGGAAAT
D6    GATCAAAGAA TAATAAATAA AGAACATATA ATAATTATTA TTAAGATGA AGAAGGAAAT
D9    GATCAAAGAA TAATAAATAA AGAACATATA ATAATTATTA TTGAAGATAC AAATGGAAAT
D10   GAACAAAGAA TAATAAATAA AGAACATATA ATAATTATTA TTGAAGATAC AAAAGGAAAT
D11   GGAGAAAGAA TAATAAATAA AGAACATATA ATAATTATTA TTGAAGATTC AGATGGAAAT
D13   GATCAAAGAA TAATAAATAA AGAACATATA ATAATTATTA TTGAAGATAA AAAAGGAAAT
D14   GATCAAAGAA TAATAAATAA AGAACATATA ATAATTATTA TTGAAGATAA AAAAGGAAAT
Ehsspl GATCAAAAAA TAAAGAATAA AGAACATATA ATGATTGTTA TTGAAGATGA AGAAGGAAAT
Clustal Co *** ** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| .
      485      495
D1    AAATTTGGAG GATATGTTAA T
D6    AAATTTGGAG GATATGTTAA T
D9    AAATTTGGAG GATATGTTAA T
D10   AAATTTGGAG GATATGTTAA T
D11   AAATTTGGAG GATATGTTAA T
D13   AAATTTGGAG GATATGTTAA T
D14   AAATTTGGAG GATATGTTAA T
Ehsspl AAATTTGGAG GATATGTTAA T
Clustal Co ***** ***** *

```

(B)

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      5      15      25      35      45      55
D1    DEETLSKTRN EIINKIISIE EMNMKDKLFR TPKEKDKITQ EFNKKKEEWK TYYSDFERK
D6    DEETLSKTRN ELINTIISIE EMIMKDKIFR KPKEKDKITQ SFNKNKEEWN TYYSDFERK
D9    DEEILSKTRN EIINKIISIE EMVMKDKLFR TPEEKDKITE EFNKKKEEWE TYYSDFEMK
D11   DEETLSKTKN KLIDKIILTE EMIKKDKIFR KPKEKDKITE SFNKKKEEWK IHYSDFERK
Clustal Co *** ** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```



```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      65      75      85      95      105     115
D1    KKKEEK---- ----EIKK QEGE-RLRTM KGMNT-IEEI QQLEEWTRNRK
D6    KKQEEE---- ----RRKKEERK QEGE-RLRIM KGMNT-MEEM QLEEWTRNRK
D9    KKKEEKKQE EEKRRKEEEE RRRKEEERK QEGE-RLRIM KGMNT-MEEI QQIEEWTERR
D11   KRKEE---- ----KEKERK QEEEGRLQIM RGISIPKHQI QLEEWTSRR
Clustal Co * : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

FIG. 11—Continued

	125	135	145	155	165
D1	VSNIIFDSV	DNWNQNTSVF	NKRIINKEHI	IIIIEDINGN	KFGGYVN
D6	VSNILFSDI	DNWNINTSVL	DQRIINKEHI	IIIIKDEEGN	KFGGYVN
D9	ISNILFDSEI	DDWNKNTSVF	DQRIINKEHI	IIIIEDTNGN	KFGGYVN
D11	IFNILFDSEI	DDWGVYTSVF	GERIINKEHI	IIIIEDSDGN	KFGGYVN
Clustal Co	: **:***: :	*:*. ***: .:	***** *****:*	*** **	*****

FIG. 11—Continued

tease domain and exhibits chymotrypsin-like activity (22). Another antigen, FIRA, of *P. falciparum* (33) has some similarity with Ehssp1. It contains two, 81-amino acid stretches of charged amino acids which are separated by blocks of repeats. The protein is polymorphic due to variation in the number of repeats. The function of the charged domain in this antigen is also unknown. A mixed charge cluster containing the sequence EEDKKRRER has been identified proximal to the C terminus of most eukaryotic HSP70 proteins (15). Experimental data suggest the importance of this charge cluster in facilitating chaperone, transport, and secretion functions of this protein. Association of proteins that contain charged amino acid clusters with pathogenesis in *E. histolytica* was apparent in a recent study of genes that get differentially expressed in abscess-derived trophozoites. Three of the genes encoded proteins that contain regions rich in charged amino acids, such as lysine and glutamic acid (8). Other than shared repetitive stretches of the amino acid sequence PKKEEKKP, there was no similarity among these proteins. All the three genes were down regulated

in abscess-derived trophozoites but were unaffected by heat stress.

Apart from the centrally located charged cluster, the other notable feature of Ehssp1 is its polymorphism. In general, polymorphic antigens are very commonly found in protozoan parasites and are implicated in evasion of host immune response by the parasite, and in establishing pathogenesis. Examples of such antigens in *P. falciparum* are the var gene family (31), the stevor and rif genes (9), the MESA antigen (11), KAHRP protein (38), and the circumsporozoite protein (25). The VSP of *Giardia* (1) and VSG of *Trypanosoma brucei* (39) are also polymorphic multigene families of which only one copy is expressed at a time (4). The polymorphic, immunodominant molecule (PIM) of *Theileria parva* (37) is the predominant antigen recognized by sera from infected cattle. It consists of conserved 5' and 3' termini flanking a polymorphic region which is composed of various numbers of a tetrapeptide repeat. These antigenic molecules perform diverse functions and use a range of mechanisms to generate polymorphism,

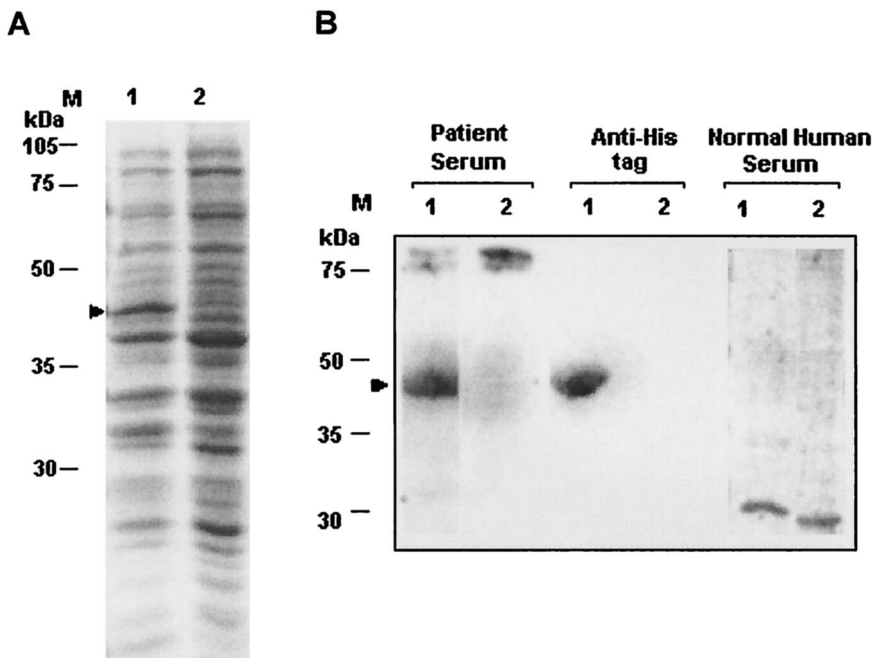


FIG. 12. Expression of Ehssp1-5076 protein and its antigenicity (A) Ehssp1 cDNA was cloned in pET30a vector and expressed in *E. coli*, BL21(DE3) to give pETEhssp1. One micromolar IPTG was used for induction and the total cell lysate of the induced (lane 1) and uninduced (lane 2) samples was electrophoresed in SDS-12% PAGE and stained with Coomassie blue. (B) Western blot analysis of induced (lane 1) and uninduced (lane 2) samples of pETEhssp1. Total cell lysate of the uninduced sample and the pellet fraction (containing the Ehssp1 protein in inclusion bodies) of the induced sample were separated in SDS-12% PAGE and transferred to a nitrocellulose membrane. The blot was probed with different antibodies as mentioned. The induced protein band is shown by arrow.

chief among them being variation in the number and sequence of short repeat motifs, recombination and gene conversion. However, most of these antigens are expressed on the cell-surface, while the structure of Ehssp1, as deduced from nucleotide sequence analysis, makes it unlikely for the protein to be located on the surface.

Perhaps the most intriguing observation regarding Ehssp1 is the change in the quantity and diversity of expression of this gene family in response to stress. Hsp70 is one of the most extensively studied stress-induced genes. Some members of this multigene family are expressed constitutively while others are stress-induced. In mouse, the 5' and 3' untranslated regions of two HSP70 genes (hsp70.1 and hsp70.3) differ from each other, suggesting their involvement in differential regulation of gene copies (40). Nucleotide sequence analysis of the putative 5' and 3' untranslated regions and sequences further upstream and downstream of the Ehssp1 ORF in the various members of the gene family did not show any striking differences between the copy that is transcribed in exponential cells and those that are turned on under stress (data not shown here). Further experiments are needed to understand the basis of this differential expression.

The transcription regulation of Ehssp1 may provide valuable insights into regulation of inducible gene expression in *E. histolytica*. The increase in expression of Ehssp1 in response to stress was not just due to increase in transcription of a single gene, as is the case for most stress-induced genes, but due to the transcription of different polymorphic copies of the gene, which are silent in the absence of stress. Thus, within the gene family different copies are regulated differentially. Expression of more than one copy of stress-induced genes has been shown for HSP70, HSP27, and HSP26 (34, 14, 18). However these multigene families express only two to four copies under stress conditions, unlike Ehssp1 where stress induces the expression of a very large repertoire of gene copies.

The time kinetics of induction of the multiple copies of Ehssp1 in comparison to a conventional HSP (HSP101) showed that the HSP101 expression reached saturation much faster (within 20 min of heat treatment) indicating that probably the pathways for induction of these two genes are different. Upon withdrawal of heat stress, the expression of all copies of Ehssp1 except the one expressed under normal conditions is shut off.

If Ehssp1 is somehow involved in the process of invasion of the intestinal mucosa by the virulent *E. histolytica*, then this protein may be dispensable in the nonpathogenic sibling species *E. dispar*, which is known to colonize the human gut but does not cause invasive disease. However, a homologue of Ehssp1 was found in *E. dispar*, although polymorphism of ABR was much less pronounced (8 to 38 amino acids). The functional significance of this is not clear at present. Stress response proteins are generally immunodominant antigens in many pathogenic infections. Both circulating antibodies and activated T cells specific for major heat shock proteins of diverse organisms (19) are found in patients. Ehssp1 also stimulates humoral immune response. Humoral response has been shown earlier for *E. histolytica* HSP70, in a group of patients with invasive amebiasis (23).

Ehssp1 is the first example of differentially regulated gene expression of a multicopy, polymorphic gene family in *E. his-*

tolytica. This could serve as a good model to study differential gene expression in this organism.

ACKNOWLEDGMENTS

We thank the Department of Science and Technology, Department of Biotechnology, and Council for Scientific and Industrial Research, Government of India, for financial support.

REFERENCES

1. Adam, R. D. 2001. Biology of *Giardia lamblia*. Clin. Microbiol. Rev. **14**:447-475.
2. Altschul, S. F., L. M. Thomas, A. S. Alejandro, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389-3402.
3. Bagchi, A., A. Bhattacharya, and S. Bhattacharya. 1999. Lack of a chromosomal copy of the circular rDNA plasmid of *Entamoeba histolytica*. Int. J. Parasitol. **29**:1775-1783.
4. Barry, J. D., and J. McCulloch. 2001. Antigenic variation in *trypanosomes*: enhanced phenotypic variation in a eukaryotic parasite. Adv. Parasitol. **49**:1-70.
5. Bhattacharya, A., S. Satish, A. Bagchi, and S. Bhattacharya. 2000. The genome of *Entamoeba histolytica*. Int. J. Parasitol. **30**:401-410.
6. Bhattacharya, S., A. Bhattacharya, and L. S. Diamond. 1988. Comparison of repeated DNA from strains of *Entamoeba histolytica* and other *Entamoeba*. Mol. Biochem. Parasitol. **27**:257-262.
7. Bruchhaus, L., M. Leippe, C. Lioutas, and E. Tannich. 1993. Unusual gene organization in the protozoan parasite *Entamoeba histolytica*. DNA Cell Biol. **12**:925-933.
8. Bruchhaus, L., T. Roeder, H. Lotter, M. Schwerdtfeger, and E. Tannich. 2002. Differential gene expression in *Entamoeba histolytica* isolated from amoebic liver abscess. Mol. Microbiol. **44**:1063-1072.
9. Cheng, Q., N. Cloonan, K. Fischer, J. Thompson, G. Waine, M. Lanzer, and A. Saul. 1998. stvor and rif are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. Mol. Biochem. Parasitol. **97**:161-176.
10. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156-159.
11. Coppel, R. L. 1992. Repeat structures in a *Plasmodium falciparum* protein (MESA) that binds human erythrocyte protein 4.1. Mol. Biochem. Parasitol. **50**:335-347.
12. Diamond, L. S., D. R. Harlow, and C. C. Cunnick. 1978. A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. Trans. R. Soc. Trop. Med. Hyg. **72**:431-432.
13. Heussler, V. T., P. Kuenzi, and S. Rottenberg. 2001. Inhibition of apoptosis by intracellular protozoan parasites. Int. J. Parasitol. **31**:1166-1176.
14. Hickey, E., S. E. Brandon, R. Potter, G. Stein, J. Stein, and L. A. Weber. 1986. Sequence and organization of genes encoding the human 27 kDa heat shock protein. Nucleic Acids Res. **14**:4127-4145.
15. Karlin, S., and L. Brocchieri. 1998. Heat shock protein 70 family: multiple sequence comparisons, function and evolution. J. Mol. Evol. **47**:565-577.
16. Kushwaha, A., P. P. Rao, V. S. Duttu, P. Malhotra, and V. S. Chauhan. 2000. Expression and characterisation of *Plasmodium falciparum* acidic basic repeat antigen expressed in *Escherichia coli*. Mol. Biochem. Parasitol. **106**:213-224.
17. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**:680-685.
18. Liang, P., R. Amons, J. S. Clegg, and T. H. MacRae. 1997. Molecular characterization of a small heat shock/alpha-crystallin protein in encysted *Artemia* embryos. J. Biol. Chem. **30**:19051-19058.
19. Lindquist, S., and E. A. Craig. 1988. The heat-shock proteins. Annu. Rev. Biochem. **63**:1-677.
20. Mai, Z., S. Ghosh, M. Frisardi, B. Rosenthal, R. Rogers, and J. Samuelson. 1999. Hsp60 is targeted to a cryptic mitochondria-derived organelle ("crypton") in the microaerophilic protozoan parasite *Entamoeba histolytica*. Mol. Cell. Biol. **19**:2198-2205.
21. Moerman, A. M., A. C. de Maria, S. L. Gomes, and C. Klein. 1998. Heat shock alters poly(A) tail length of *Dictyostelium discoideum* hsp32 RNA. DNA Cell Biol. **17**:635-641.
22. Nwagwu, M., J. D. Haynes, P. A. Orlandi, and J. D. Chulay. 1992. *Plasmodium falciparum*: chymotryptic-like proteolysis associated with a 101-kDa acidic-basic repeat antigen. Exp. Parasitol. **75**:399-414.
23. Ortner, S., B. Plaimauer, M. Binder, G. Wiedermann, O. Scheiner, and M. Duchene. 1992. Humoral immune response against a 70-kilodalton heat shock protein of *Entamoeba histolytica* in a group of patients with invasive amebiasis. Mol. Biochem. Parasitol. **54**:175-183.
24. Osteryoung, K. W., H. Sundberg, and E. Vierling. 1993. Poly(A) tail length of a heat shock protein RNA is increased by severe heat stress, but intron splicing is unaffected. Mol. Gen. Genet. **239**:323-333.

25. **Ozaki, L. S., P. Svec, R. S. Nussenzweig, V. Nussenzweig, and G. N. Godson.** 1983. Structure of the *Plasmodium knowlesi* gene coding for the circumsporozoite protein. *Cell* **34**:815–822.
26. **Polla, B. S.** 1991. Heat shock proteins in host-parasite interactions. *Immunol. Today* **12**:A38-A41.
27. **Purdy, J. E., L. T. Pho, B. J. Mann, and W. A. Petri, Jr.** 1996. Upstream regulatory elements controlling expression of *Entamoeba histolytica* lectin. *Mol. Biochem. Parasitol.* **78**:91–103.
28. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual*, 2nd ed., p. 9.31–9.57. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
29. **Sanger, F., D. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
30. **Sharma, R., A. Bagchi, A. Bhattacharya, and S. Bhattacharya.** 2001. Characterization of a retrotransposon-like element from *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* **116**:45–53.
31. **Smith, J. D., B. Gamain, D. I. Baruch, and S. Kyes.** 2001. Decoding the language of var genes and *Plasmodium falciparum* sequestration. *Trends Parasitol.* **17**:538–545.
32. **Sonna, L. A., J. Fujita, S. L. Gaffin, and C. M. Lilly.** 2002. Invited review: Effects of heat and cold stress on mammalian gene expression. *J. Appl. Physiol.* **92**:1725–1742.
33. **Stahl, H. D., P. E. Crewther, R. F. Anders, G. V. Brown, R. L. Coppel, A. E. Bianco, G. F. Mitchell, and D. J. Kemp.** 1985. Interspersed blocks of repetitive and charged amino acids in a dominant immunogen of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* **82**:543–547.
34. **Sung, D. Y., E. Vierling, and C. L. Guy.** 2001. Comprehensive expression profile analysis of the Arabidopsis Hsp70 gene family. *Plant Physiol.* **126**:789–800.
35. **Tovar, J., A. Fischer, and C. G. Clark.** 1999. The mitosome, a novel organelle related to mitochondria in the amitochondriate parasite *Entamoeba histolytica*. *Mol. Microbiol.* **32**:1013–1021.
36. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from SDS and acid/urea gels to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **76**:4350–4354.
37. **Toye, P., E. Gobright, J. Nyanjui, V. Nene, and R. Bishop.** 1995. Structure and sequence variation of the genes encoding the polymorphic, immunodominant molecule (PIM), an antigen of *Theileria parva* recognized by inhibitory monoclonal antibodies. *Mol. Biochem. Parasitol.* **73**:165–177.
38. **Triglia, T., H. D. Stahl, P. E. Crewther, A. Silva, R. F. Anders, and D. J. Kemp.** 1988. Structure of a *Plasmodium falciparum* gene that encodes a glutamic acid-rich protein (GARP). *Mol. Biochem. Parasitol.* **31**:199–201.
39. **Vanhamme, L., E. Pays, R. McCulloch, and J. D. Barry.** 2001. An update on antigenic variation in African trypanosomes. *Trends Parasitol.* **17**:338–343.
40. **Walter, L., F. Rauh, and E. Gunther.** 1994. Comparative analysis of the three major histocompatibility complex-linked heat shock protein 70 (Hsp70) genes of the rat. *Immunogenetics* **40**:325–330.
41. **Weber, J. L., J. A. Lyon, R. H. Wolff, T. Hall, G. H. Lowell, and J. F. Chulay.** 1988. Primary structure of a *Plasmodium falciparum* malaria antigen located at the merozoite surface and within the parasitophorous vacuole. *J. Biol. Chem.* **263**:11421–11425.
42. **Willhoeft, U., H. Buss, and E. Tannich.** 2002. The abundant polyadenylated transcript 2 DNA sequence of the pathogenic protozoan parasite *Entamoeba histolytica* represents a nonautonomous non-long-terminal-repeat retrotransposon-like element which is absent in the closely related nonpathogenic species *Entamoeba dispar*. *Infect. Immun.* **70**:6798–6804.
43. **Yadava, N., M. R. Chandok, J. Prasad, S. Bhattacharya, S. K. Sopory, and A. Bhattacharya.** 1997. Characterization of EhCaBP, a calcium binding protein of *E. histolytica* and its binding proteins. *Mol. Biochem. Parasitol.* **84**:69–82.
44. **Zugel, U., and S. H. Kaufmann.** 1999. Role of heat shock proteins in protection from and pathogenesis of infectious diseases. *Clin. Microbiol. Rev.* **12**:19–39.

Editor: W. A. Petri, Jr.