SUMMARY
Entamoeba histolytica, a protozoan parasite is the causative agent of amoebiasis. This parasite is present globally but is more prevalent in the tropical and subtropical regions of the globe. Approximately, 500 million people are infected with *E. histolytica*, of which 40 million develop the disease. Thus, morbidity rates are high due to this disease while mortality rates are relatively low causing 40,000 deaths per annum.

The majority of other Entamoeba species do not cause any clinical symptoms except *Entamoeba polecki* and *Entamoeba gingivalis*. The life cycle of *E. histolytica* revolves around the two stages: the cyst, which is a host-infective stage, and the trophozoite, a vegetative host-tissue-invasive stage.

Molecular approaches to understanding the pathogenicity of this parasite have only begun as recently as 1986 when the ribosomal DNA was shown to reside in an extrachromosomal circle. Studies on this parasite have mainly focussed on the pathogenicity while studies on either the physiology or the biochemistry of this organism are lagging far behind.

VI.1 Isolation, expression and analysis of the cDNA encoding calcium binding protein from *E. histolytica*

The calcium binding protein was initially isolated as a fusion protein from a lambda gt11 cDNA expression library using a polyclonal antibody, aEhT.KCG. This antibody, aEhT.KCG, identifies amongst others, proteins present in the immune complex present in amoebiasis patients. Essentially, a single cDNA of 0.3 kb was isolated by immunoscreening. The insert size and the size of the β-galactosidase fusion protein suggested that the cDNA encoded a protein of 10 kDa.

The cDNA was further analyzed by nucleotide sequencing. The main features are summarized below.

1. The length of the sequence was 297 bp and agreed well with the size estimated earlier.
2. Open reading frame (ORF) analysis showed the presence of only one continuous frame encoding a polypeptide of 99 amino acids. This agreed well with the size of the fusion protein estimated earlier.
3. The cDNA lacked the 3'-end as the poly A tail was missing. The absence of a start codon at the 5'-end suggested that the cDNA was incomplete.
4. A search for possible homologues of the deduced amino acid sequence sug-
gested that the encoded polypeptide probably was a calcium binding protein as all the homologous proteins in the data bank were of the EF-hand superfamily.

5. Alignment with different calcium binding proteins of the EF-hand superfamily of proteins suggested that the incomplete cDNA contained three Ca$^{2+}$-binding domains.

VI.2 Isolation and characterization of the genomic clones encoding CaBP

Six genomic clones were isolated from a genomic library in lambda ZAP using the cDNA probe. The analysis of these clones showed two of them, carrying an insert size of 2.2 kb, contained the entire ORF. Restriction mapping analysis indicated that the EcoR I site present at the 3'-end of the cDNA was present as an internal site. The coding region was localized to the 3'-end of the genomic clone.

VI.3 Sequence analysis of the coding region

The nucleotide sequence of the complete coding region and stretches upstream and downstream of it was determined and analyzed by computer analysis.

The salient features are summarized below:

1. The genomic sequence matched with the cDNA sequence discounting the possibility of RNA editing.

2. There is only one Open Reading Frame (ORF) of 402 nucleotides present in the sequence, encoding a polypeptide of molecular mass 14,952 daltons.

3. Homology search with the deduced ORF showed that the protein has maximum homology of 30-35% with calmodulins and other calcium binding proteins from a variety of organisms. The homology was mostly around the putative Ca$^{2+}$-binding domains.

4. Comparison of the deduced amino acid sequence with other calcium binding proteins suggested the presence of four EF-hands. The Ca$^{2+}$-binding loops were highly conserved and consisted of the invariant Asp, Gly and Glu at the first, sixth and the twelfth positions. The other positions in the Ca$^{2+}$-binding loop were occupied by negatively charged amino acids consistent with the fact that the Ca$^{2+}$-binding ligands were present continuously and that the side chains of the protein bound Ca$^{2+}$.

Comparison with a large number of other calcium binding proteins of the EF-hand superfamily suggested that the sequences preceding and suc-
ceeding the Ca\textsuperscript{2+}-binding loop were quite different from that of calmodu-
lin. However, as this was quite true for a large number of other calcium
binding proteins belonging to this superfamily; it was deduced that the
protein belonged to this superfamily. This was further strengthened by the
fact that a few key residues are conserved in all these proteins and the
sequences preceding and succeeding the Ca\textsuperscript{2+}-binding loop could theoreti-
cally form alpha helices. These interpretations can only be confirmed by
direct structural analysis as our ability to designate a structure from
sequence is still not perfect.

5. Comparison of the four Ca\textsuperscript{2+}-binding domains suggested that these can be
classified into two groups. Ca\textsuperscript{2+}-binding domains I and III and domains II
and IV share homology of about 70\%, respectively. This agreed with the
hypothesis that the progenitor calcium binding protein had two Ca\textsuperscript{2+}-bind-
ing domains. Duplication of the early sequence may be responsible for
evolution of a large number of proteins with four Ca\textsuperscript{2+}-binding domains.

6. The comparison of 5' upstream sequences of a number of E. histolytica genes
showed consensus sequence for transcription initiation (ATTCA) present as
a single base mismatch (AATCA), 11 bp upstream of the translation initia-
tion site. A perfect match for the proposed Goldberg-Hogness box (YATT-
TAAA) is also present 24-26 bp upstream of the transcription initiation
site.

7. Codon bias analysis showed that the designated ORF is similar to many other
E. histolytica genes known so far.

VI.4 Ca\textsuperscript{2+}-binding and mobility assays

The Ca\textsuperscript{2+}-binding property of the putative calcium binding protein was
determined by expressing both the cDNA and the genomic clone in E. coli. The
cDNA was expressed as a fusion protein containing a maltose binding protein
from the expression vector, pMAL-C. The expressed protein bound Ca\textsuperscript{45}, assayed
after western blotting in the presence of high concentrations of Mg\textsuperscript{2+} and K\textsuperscript{+}.
The entire ORF was also expressed in E. coli using the expression system, pET-
3c. The recombinant protein could bind Ca\textsuperscript{2+} and undergo conformational change
(mobility shift) in SDS-PAGE. These experiments confirmed that the protein
belonged to the EF-hand superfamily of calcium binding proteins.
VI.5 The *E. histolytica* calcium binding protein (EhCaBP) is expressed in trophozoites

Northern blot analysis demonstrated the presence of a single species of poly A+ RNA of 0.52 kb suggesting that the gene is expressed in trophozoites. The expression of the protein was monitored by polyclonal antibody, against the recombinant EhCaBP (rEhCaBP) raised in rabbits. Both Western blot analysis and ELISA confirmed the presence of the protein molecule in trophozoites of *E. histolytica*. The size of the amoebic protein is identical to the recombinant protein. Moreover, competition experiments using the *E. histolytica* lysates suggested that the two proteins may be antigenically identical.

VI.6 Expression of the recombinant protein in *E. coli*

Polymerase chain reaction (PCR) using primers designed to amplify the ORF from the genomic clone was used to amplify and clone the coding region in the T7-based expression vector, pET-3c. The presence of a single amplified band of 0.4 kb suggested that the ORF had been amplified specifically by the primers used.

The orientation of the cloned ORF in the vector, pET3-c, was confirmed by restriction enzyme digestions. The expressed recombinant protein migrated with a molecular mass of 15 kDa under denaturing conditions. Interestingly the recombinant protein was soluble and not in inclusion bodies as observed for many other proteins.

VI.7 Biochemical characterization of recombinant *E. histolytica* calcium binding protein

The recombinant protein was found to be stable at 100°C and in the presence of trichloroacetic acid. It could bind an anion exchanger in absence of Ca2+ and could be eluted specifically with Ca2+. It also got precipitated with 60% ammonium sulphate just like calmodulin. It has been pointed out before that as in case of calmodulin, this protein also has four putative Ca2+-binding domains and it migrates differently on binding Ca2+.

However, the recombinant protein was unable to bind phenyl-sepharose in a Ca2+-dependent manner, an important property of calmodulins. This can be explained on the basis of hydrophobicity pattern as the amoebic protein has different pattern compared to calmodulins. These experiments indicate that the calcium binding protein of *E. histolytica* is closely related but not similar to the calmodulins. Incidentally, during the course of this study, the
presence of a calmodulin was reported in *E. histolytica*.

### VI.8 Immunochemical analysis

The polyclonal antibody against the expressed recombinant protein was used for immunochemical analysis. There may not be any size polymorphism of the calcium binding protein in different pathogenic strains of *E. histolytica*. The amoebic calcium binding protein was present in all strains of *E. histolytica* including the nonpathogenic xenic strain, SAW1734 R Cl AR. However, this protein was either not present or was antigenically very dissimilar in other species of *Entamoeba*.

Studies on the immunogenicity of this molecule in patients suffering from amoebic liver abscess, suggested that this protein was not immunogenic.

### VI.9 Molecular characterization

Degenerate primers were used to amplify the coding region of the calcium binding protein from different strains and species of *Entamoeba* and analyzed by southern hybridization. As expected all strains of *E. histolytica* including the nonpathogenic strain, SAW1734 R Cl AR showed the presence of a similar sized band (0.4 kb). No amplified product was detected from *Entamoeba moshkovskii* strain Laredo. The absence of the gene in other *Entamoeba* may explain the data obtained from immunochemical studies. The amoebic calcium binding protein could change conformation on binding Ca$^{2+}$ as demonstrated by the mobility shift assay. This points to a possible role of this protein in *E. histolytica* in activating target enzymes after binding Ca$^{2+}$. Incidentally, calmodulin and a large number of other calcium binding proteins also share this property. Further work on this protein would focus on the target enzymes and help to elucidate the pathways involved in signal transduction.