3. Results-
3. 1. Cloning, expression and purification of the polypeptide encoded by EhLINE1 EN domain:

Although EhLINE1 is present in a few hundred copies per genome, it is not known whether any active copy of this element exists in *E. histolytica*. Genome sequence analysis showed that though most copies have multiple mutations, several entries in *E. histolytica* database had long ORFs. This raises the possibility that there may be at least a few functional copies of this retrotransposon in *E. histolytica* genome. In order to understand the process of retrotransposition of EhLINE1 in *E. histolytica*, we have cloned and expressed the endonuclease (EN) domain in *E. coli* to study its properties with respect to target site specificity of nicking *in vitro*. Since retrotransposition of these elements initiates with the bottom strand nick made by the endonuclease, we believe that an understanding of this process will help to decipher the overall mechanism of retrotransposition.

As most of the EhLINE1 copies are highly degenerate, we took advantage of the *E. histolytica* shot gun library clones (used in genome project) to reconstruct the EN domain. From nucleotide sequence analysis of EhLINE1, the conserved motifs of the EN domain thought to be important for activity were localized to nucleotide positions 4122 to 4479 (Fig. 11A). Two GSS entries encompassing this region, and lacking any stop codons were used to clone the EN domain. The clones corresponding to these entries (AZ669903 and AZ541065) were a kind gift from TIGR, USA. A 276 bp *EcoRI*-NdeI fragment from AZ669903 and a 600 bp *NdeI*-BamHI fragment from AZ541065 were ligated and cloned in pBS (pBluescript II KS+). For expression of EhLINE1-EN in *Escherichia coli* the *EcoRI*-NotI fragment (794 bp) from pBS was cloned into the *EcoRI*-NotI site of pET30b. The expressed protein contained a His-tag, and together with other vector sequences at the amino-terminus, it was 309 amino acids in size, with an expected molecular mass of 35.5 kDa. It was purified by Ni-agarose chromatography and its identity was confirmed using anti-his antibodies (Fig. 11). Since many endonucleases are found in extracts, it was possible that an *E. coli* activity copurified with EhLINE1 ENp. The PD..D motif in the EN domain is conserved among non-LTR retrotransposons containing REL endonuclease and also in certain restriction enzymes (Yang et al, 1999). The conserved amino acids in PD..D motif are essential for enzyme activity (Yang et al, 1999). Therefore, as a control for
Fig. 11. Purification of the recombinant EN protein. (A) The nucleotide positions of the RT and EN domains in EhLINE1 are given on the top line. The amino acid sequence of the EN domain is indicated and conserved motifs are shown by shading. The numbers in parentheses are the number of amino acids between adjacent motifs. (B) A 794 bp DNA fragment from nucleotide positions 4010 to 4804, containing the EN domain was cloned in the E. coli expression vector pET30b (pET-EN) as detailed in Materials and Methods. The expected molecular mass of the encoded polypeptide is 35.5 kDa. E. coli cells were induced with IPTG to express the cloned gene. Lysates were prepared from lanes 1, 5, pET-EN, uninduced; lanes 2, 6, pET30b, induced; lanes 3, 7, pET-EN, induced; lanes 4, 8, pET-ENM (mutated to PA..D), induced (see text for details). Samples were separated by 10% SDS-PAGE and the proteins were visualized by Coomassie blue staining. (C) Lysates from pET-EN-containing cells were separated by 10% SDS-PAGE followed by western blot analysis with monoclonal anti-his-tag antibody, as described in ‘Materials and Methods’. Chemiluminescent bands were visualized by autoradiography. Lane 1, Total cell-lysate from uninduced cells; lane 2, total cell-lysate from induced cells; lane 3, EN purified by Ni-agarose chromatography.
any contaminating activities from *E. coli* extracts, the PD..D motif in EN was changed to PA..D by site directed mutagenesis (section 2.24). The mutant protein (ENM) was expressed and purified in parallel with wild type (Fig. 11).

3.2. Lack of a strict target-site consensus sequence for insertion of EhLINE1 and EhSINE1:

To design a suitable substrate for testing the endonucleolytic activity of the EN protein we looked at the flanking sequences at the sites of insertion of EhLINE1 and the putative partner SINE, EhSINE1. Sequences were obtained from the GSS and contig databases and aligned using ClustalW. The 5'- and 3'- ends of the elements were defined as follows. Both elements started with 5'-AGA TC, after which the sequence of the EhLINE1 family diverged from the EhSINE1 family, but within each family they were highly conserved. The 3'-end of both shared a 74 bp sequence, which ended with ...CTTTTTATTT-3', with minor variations. The sequences upstream and downstream of EhLINE1 and EhSINE1 showed that these elements do not insert in a strictly site-specific manner. However, a T-rich stretch of nucleotides was almost universally present about 13-19 nucleotides upstream of the insertion site of both elements, and the sequence immediately downstream was generally rich in T and A. A representative sample of the multiple alignment highlighting these features is shown in Fig. 12.

3.3. The EN protein can cleave a nonspecific substrate, pBlueScript:

The lack of a strict target-site consensus for insertion of EhLINE1 prompted us to check if the EN protein could utilize a nonspecific substrate, pBS as found for the human L1 endonuclease (Feng et al, 1996). Supercoiled pBS DNA was efficiently nicked by the purified EN protein to yield open circular and linear DNAs (Fig. 13). Equal amount of mutated EN (ENM) had very low activity while no activity was found in extracts from uninduced, and pET30b-induced cells. In a time-course of endonucleolytic activity it was evident that supercoiled pBS DNA was first nicked to yield open circular DNA, which was then converted to linear DNA (Fig. 13). Since the amount of linear DNA produced at early time points was very low compared with
Fig. 12. Genomic sequences flanking the sites of insertion of EhLINE1 and EhSINE1. EhLINE1 flanking sequences were obtained from the *E. histolytica* genome database at TIGR and Sanger centre. EhSINE1 flanking sequences were from the GSS database. A representative sample is shown. Nucleotides occurring in at least 70% of entries are shaded. The T-rich stretch upstream of insertion sites of both elements is prominent.
Fig. 13. Endonucleolytic activity of the EN protein with pBS DNA as substrate. The protein was purified as described in Fig. 11. (A) Supercoiled pBS DNA (200 ng) was incubated with 80 ng protein at 37°C for 1 h. Lane 1, untreated pBS DNA; pBS DNA incubated with protein purified from E. coli containing – lane 2, pET-EN induced with IPTG; lane 3, pET-EN, uninduced; lane 4, pET 30b, induced. (B) pBS DNA (200 ng) was left untreated (lane C), or incubated with 80 ng of wild type EN (lane EN), or mutated EN (lane ENM) at 37°C for 30 min. DNA was electro-phoresed through 0.8% agarose at 5 V/cm for 4 h and visualized by ethidium bromide staining. (C) The time course of endonucleolytic cleavage of pBS with EN protein was determined by incubating pBS DNA (100 ng) for the indicated time periods with 40 ng of EN protein. The reaction was terminated by the addition of EDTA. DNA was electrophoresed through 0.8% agarose at 5 V/cm for 4 h and visualized by ethidium bromide staining. OC, open circular, L, linear, SC, supercoiled.
open circular DNA, the enzyme appears to make predominantly single strand nicks
and not double strand breaks. On further incubation the linear DNA gave a smear of
low molecular weight molecules. This (and experiments reported below) shows that
the enzyme can use both circular and linear DNAs as substrate. The slow appearance
of smears, and the presence of detectable levels of high molecular weight linear DNA
even after 240 min indicates that smears are probably due to closely spaced nicks on
the two DNA strands, although a low level of exonucleolytic activity cannot be ruled
out. Absence of exonuclease activity in our enzyme preparation is also borne out by
experiments described below in which end-labeled linear DNAs were used as
substrates.

3.4. EhLINE1 ENp nicks result in 5'-PO₄ and 3'-OH termini:

Nucleases can leave either 5'-PO₄, 3'-OH or 5'-OH, 3'-PO₄ termini. The EhLINE1 EN
domain resembles type IIS restriction endonucleases which leave 5'-PO₄, 3'-OH
termini. We tested whether the products of EhLINE1 EN were substrates for T4 DNA
ligase which requires 5'-PO₄ and 3'-OH termini. Nicked circles generated by EN were
incubated with T4 DNA ligase. Ligation of nicks was evident by the appearance of a
faster migrating band corresponding to closed circular DNA (Fig. 14) [Feng et al,
1996]. Thus EN generates 5'-PO₄ and 3'-OH termini.

3.5. Supercoiled pBS DNA has cleavage hotspots for EhLINE1 EN:

To look for any possible hotspots where the enzyme nicks preferentially, pBS DNA
was treated with EN protein for 10 min, following which the enzyme was denatured
and DNA was linearized with EcoRI. The fragments generated were detected by
Southern hybridization. The appearance of discrete bands along with a background
smear showed that the enzyme did have a preference for some sites (Fig. 15).

3.6. Nicking activity of EN at the target sites occupied by EhSINE1:

To test the activity of EN on its natural substrate, we looked for sites of insertion of
EhLINE1/SINE1 in the E. histolytica genome, which may be unoccupied. Since E.
histolytica is thought to be polyploid (n=4) (Clark et al, 2000), it may be possible to
Fig. 14. Structure of nick created by EhLINE1 EN. Supercoiled pBS DNA [200 ng] (lane 2) was incubated with EN protein (80 ng) for 30 min (lane 3) and 60 min (lane 5). EN protein was heat inactivated. The products were purified and incubated with T4 DNA ligase for 2 h. (lanes 4 and 6). CC, closed relaxed circle DNA. DNA was electrophoresed through 1.0 % agarose at 3 V/cm for 5 h and visualized by ethidium bromide staining.
Fig. 15. Clevage hotspots of EN in supercoiled pBS plasmid. Supercoiled pBS DNA was incubated with EN for 10 min at 37°C. After inactivating the EN, products were digested with EcoRI. Lane1 - pBS DNA incubated with EN only; lane2 - pBS DNA treated first with EN and then with EcoRI. DNA was blotted onto nylon membrane after electrophoresis through 0.8 % agarose gel at 4 V/cm for 4 h and probed with the total pBS DNA probe. Autoradiographic exposure was for 4 h. The bands corresponding to hotspots are shown by arrowheads.
Results

find homologous regions of chromosomes, some of which harbor these elements while others are unoccupied. To look for unoccupied sites it is necessary to have contigs containing the full-length element, from where flanking sequences can be obtained. Since the assembly of the *E. histolytica* genome sequences is not complete, we undertook to do this exercise only for the 0.55 kb EhSINE1 at this point. The GSS database was adequate for this purpose since it contains, on an average, 2.0 kb clones with 0.8 kb of sequence available from either end. Thirty eight entries containing the full-length EhSINE1 could be retrieved from the database. Sequences flanking the ends of the element in each entry were stitched together and used to search the database for unoccupied sites. Only one unoccupied site was found in GSS entry AZ669709 for which the corresponding EhSINE1-containing GSS entry was AZ550231. The infrequent occurrence of unoccupied sites shows that once a site is occupied on one chromosome, the empty sites on its homologues are also filled up. Both EhSINE1 and EhLINE1 insertions are accompanied by short target site duplications (TSDs) as seen from sequence analysis. In the case of AZ550231, EhSINE1 insertion resulted in a 22 bp TSD, which is present only once in AZ669709 which lacks the insertion. A 174 bp fragment containing the unoccupied site was obtained by PCR amplification of *E. histolytica* genomic DNA, using primers designed from the sequence in AZ669709 (Fig. 16B). The fragment was sequenced to confirm its identity with the sequence of the GSS entry. The fragment was then labeled either in the bottom strand or in the top strand by PCR amplification using the corresponding end-labeled primer. It was incubated with the endonuclease and the resulting products were denatured and separated on a sequencing gel along with a sequencing reaction of the same template with each primer. The sites of nicking by the endonuclease on each strand were identified from the parallel sequencing run. The substrate in which the bottom strand was labeled showed three major nicking sites in which the products were doublets (marked 1-3, Fig. 16A), and two less preferred sites in which the products were primarily single bands (marked 4,5, Fig. 16A). Of these, bands 1 and 3 were the brightest in intensity. Band 3 corresponded with the exact site of insertion of EhSINE1 (see Fig. 16B for the sequence at which EhSINE1 inserts). The pattern of nicks on the top strand was different from the bottom strand. The enzyme acted on many sites but none was as strongly preferred as the sites in bands
Fig. 16A
Fig. 16. Nicking profile of EN protein on a 174 bp substrate containing an unoccupied insertion site of EhSINE1. This site was found in a GSS database entry (acc. no. AZ669709). The fragment containing the site was obtained by PCR amplification of *E. histolytica* genomic DNA. The sequence of the fragment used is shown in panel B and the primers used for PCR are marked at the two ends with horizontal arrows. Either the top strand or the bottom strand of the substrate was radiolabeled by using the appropriate primer end-labeled with $[\gamma-^{32}P]$ATP. The radiolabeled 174-bp fragment was gel-purified, and 100 ng was incubated with 40 ng of EN protein at 37°C for 1 h (lane 3 panel A). Controls used were protein purified from pET-EN uninduced cells (lane 2), and no added protein (lane 1). After enzyme digestion, the products were denatured by boiling in formamide and separated through 6% polyacrylamide gel at 60 W for 2 h. A sequencing reaction using the same primers was run in parallel (lanes G,A,T and C). Asterisk marks the point of insertion of EhSINE1. Numbered arrows mark the position of prominent nicks in the two strands. The 22 bp sequence that is duplicated upon EhSINE1 insertion is boldfaced in panel B.
Results

one and three on the bottom strand. In addition, most of the bands generated from the top strand appeared to be singlets (Fig. 16A). The sequences surrounding the nicks on the two strands are listed in Table 1. From this it appears that the enzyme nicks at some preferred sequences and not randomly, as clear patterns could be discerned. The differences between the nicking patterns of the top and bottom strands could, in part, be due to the T-richness of the top strand compared with the corresponding A-richness of the bottom strand. The fact that TSDs created by insertion of these elements are variable in size (3-23 bp) also shows that the position of top strand nick may be influenced by factors other than the intrinsic specificity of endonuclease alone. The property of this endonuclease to nick at one of two consecutive phosphodiester bonds has also been observed with human L1 endonuclease (Feng et al, 1996) and R2Bm endonuclease (Xiong et al, 1988). We were not able to repeat this experiment with more unoccupied sites of EhSINE1 since only one such site could be identified in the database so far.

3.7. Nicking activity of EN at the genomic sequences flanking EhLINE1 and EhSINE1 inserted copies:

To further understand the nicking preferences of the endonuclease, we used as substrates fragments containing the boundary regions of sites where EhSINE1 and EhLINE1 had inserted. For EhSINE1 the GSS entry AZ550231 contained the entire element. Primers were designed to PCR amplify two fragments of 187 and 164 bp respectively from the 5'-end and from the 3'-end, containing a part of the element and adjoining genomic sequence (Fig. 17). *E. histolytica* genomic DNA was used for PCR amplification. To obtain radioactively-labeled fragments the bottom strand primer was end-labeled in each case. After incubating with the endonuclease the resulting products were separated on a sequencing gel as described for Fig. 16. Several bands of varying intensity were observed with both fragments. Of these the four most prominent bands are marked (Fig. 17). In the fragment from the 3'-end, there was a prominent nick at the 3'-boundary with the TSD (band 3, Fig. 17B), which corroborates the data with the unoccupied site in Fig. 16. The sequences at the nicking sites are listed in Table 1.
Fig. 17
Fig. 17. Nicking profile of EN protein on DNA fragments derived from the two ends of a EhSINE1 insertion. The sequence of the element and flanking regions was obtained from GSS entries AZ550231 and AZ542529. Sequence of the two fragments used as substrates for EN protein is shown in the bottom panel. Each fragment contained a part of the EhSINE1 element (shown in small case) and the adjoining genomic region from either the 5′-end or 3′-end of the element. The TSD is shown in bold, while sequences of primers used for PCR amplification of genomic DNA to obtain the fragments are marked with horizontal arrows. Both fragments were labeled in the bottom strand and used as substrate for nicking by EN protein. The reaction conditions and electrophoretic analysis were as described in legend of Fig. 16. Electrophoresis was done at 60 W for 1.5 h. Numbered arrows indicate positions of the most prominent nicks. In the fragment from 3′-end, the alignment of band no. 1 with the sequencing run was slightly skewed.
The same experiment was repeated for EhLINE1 by using contig 8160 from the *E. histolytica* genome database at Sanger centre. The contig contains the entire EhLINE1. Again, the boundary fragments were obtained by PCR amplification and the bottom strand was labeled. The nicking pattern is shown in Fig. 18 and the nicked sequences are listed in Table 1. Here, too, the enzyme nicked the bottom strand very close to the 3'-end of TSD in the fragment derived from the 3'-end (band 2, Fig. 18B). However, this band was not the most prominent, as found with the EhSINE1 insertion site (band 3, Figs. 16 and 17B). It is likely that mutations may have occurred after the original transposition event took place, making the sequence in contig 8160 a less favorable substrate. The fragment from the 5'-end did not give very prominent nicks.

Once better assembly of the *E. histolytica* genome sequence is available, it may be possible to find some unoccupied sites of EhLINE1 and use them as substrate. Although the EhLINE1-encoded endonuclease lacks site-specificity, it is clear from Table 1 that it nicks at some preferred sites. For example, the most commonly nicked sequence on the bottom strand was 5'-AT-3'. Another common feature on the bottom strand was the presence of a G, 3 to 4 nucleotides upstream of the nick in a large number of sites.

To gain further insight into the possible mode of retrotransposition of this element we analyzed the sequences flanking the inserted element, as available in the current database. We selected entries containing the full-length EhLINE1 and EhSINE1 (Fig. 19). The TSD could be readily identified in most cases and lay in the size range of 3-22 bp. Two important features that emerged from this analysis are as follows. First, of 13 entries containing full-length EhLINE1 and 17 containing full-length EhSINE1 shown in Fig. 19, 23 of 30 contained a C (G in bottom strand) within 0 to 5 nucleotides from the TSD 3'-end, while the chance of random occurrence of C in a 60-nucleotide stretch surrounding the TSDs in these entries was calculated to be 8%. Since a large number of nicks made by the endonuclease on the bottom strand lay close to a 5'-G residue (Table 1), this strengthens the possibility that the G in the bottom strand may be involved in target site recognition and nicking by the endonuclease (experiments described below), leading to insertion of the element. Second, the TSD upstream of the element was almost always preceded immediately
Fig. 18. Nicking profile of EN protein on DNA fragments derived from the two ends of an EhLINE1 insertion. The experiment was done as shown in Fig. 17. Genomic sequences flanking EhLINE1 were obtained from *E. histolytica* genome database at Sanger centre (Contig 8160). Sequences of the fragments used as substrate are shown in the bottom panel. PCR primers are marked at the ends of the fragments. The TSD is shown in bold and the EhLINE1 sequence is in small case. Numbered arrows indicate positions of prominent nicks.
<table>
<thead>
<tr>
<th>Unoccupied site</th>
<th>EhSINE1 insertion</th>
<th>EhLINE1 insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bottom strand</strong></td>
<td><strong>Top Strand</strong></td>
<td><strong>Bottom strand</strong></td>
</tr>
<tr>
<td>Bottom strand</td>
<td>Bottom strand</td>
<td><strong>3'-end</strong></td>
</tr>
<tr>
<td>1. 5'-GTATG</td>
<td>1. 5'-TTACT</td>
<td>1. 5'-GTATG</td>
</tr>
<tr>
<td>2. 5'-GCATT</td>
<td>2. 5'-TTAAT</td>
<td>2. 5'-GCATT</td>
</tr>
<tr>
<td>3. 5'-GTATT</td>
<td>3. 5'-TTATT</td>
<td>3. 5'-GTATT</td>
</tr>
<tr>
<td>4. 5'-TAATA</td>
<td>4. 5'-TTAAT</td>
<td>4. 5'-GTATT</td>
</tr>
<tr>
<td>5. 5'-GAGTA</td>
<td>5. 5'-CCTCT</td>
<td>12. 5'-ACATA</td>
</tr>
<tr>
<td>6. 5'-TTATT</td>
<td>13. 5'-CACTA</td>
<td></td>
</tr>
<tr>
<td>7. 5'-GGTTTG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 1. Sequences surrounding the nicking sites of the EN protein. The data is summarized from Figs. 16 to 18.
Fig. 19. Conserved sequence features of EhLINE1 and EhSINE1 insertion. EhLINE1 sequences were obtained from *E. histolytica* genome database at TIGR (sequences 1 to 6) and Sanger centre (sequences 7 to 13), while EhSINE1 sequences were from the GSS database. Boxes show the 3'-most sequence of the element. The TSDs at the two ends are in upper case. Regions of identity between them are underlined, while non-identical nucleotides are marked by dots. The sequences immediately upstream of the TSD, and at 3'-end of the element, which share close identity, are in lower case. Their regions of identity are underlined and non-identical nucleotides are marked by dots. The nearest C flanking the 3'-TSD at its 3'-end is boxed.
Results

by a T-rich stretch (which contributes to the block of Ts seen upstream of the elements in Fig. 12). This stretch showed a very good sequence match with the 3'-end of EhLINE1 and EhSINE1 (in lower case and underlined in Fig. 19). The role of this T-rich stretch is as yet unknown.

3.8. Substrate requirements of EhLINE1 EN:
To investigate the important features which EhLINE1 EN recognizes in its target site, we assayed the nicking activity for natural as well as artificial sequences. The data presented so far showed that EN while not being strictly sequence specific, nicked preferentially at some sequences (Table 1), including three major sites in a 174 bp DNA fragment containing the exact site of insertion of EhSINE1 (Fig. 16). In the following section, these sequence requirements were investigated in more detail.

3.8.1. Minimum substrate length for EN activity:
To determine the minimum substrate length, we selected the region surrounding site #3 in the 174 bp fragment (Fig. 16). Oligos were made to PCR amplify fragments of different lengths, all including the site #3 which was the target for nicking. The 174 bp fragment was used as template for PCR amplification. The PCR products were purified by running through a non-denaturing PAGE gel (as described in Materials and Methods), and used as substrates for EN. Deletion analysis showed that a 27 bp fragment (-11 to +16 with respect to the nick in site #3) was sufficient as a substrate for EN (Fig. 20).

3.8.2. Role of sequences immediately surrounding site #3 in nicking activity:
Next we made a series of substrates where mutations were introduced in site #3 of the 174 bp DNA to determine the sequences important for nicking. The site #2 was used as control for all substrates. Transition mutations were introduced using oligos with appropriately altered sequences to PCR amplify a 115 bp fragment from 174 bp region. The resulting DNA thus contained a normal site #2 and a mutated site #3. The activity of EN on the mutated site #3 was quantified using site #2 as an internal
Substrate analysis to determine the minimum substrate length for EN activity. The region surrounding site #3 of 174 bp fragment (see Fig. 16) was selected for deletion analysis. Substrates of different lengths were prepared by using the indicated oligos (underlined sequences) for PCR amplification. The sequences of the substrates used for EN assay are shown in the lower panel. Bottom strand primers were end-labeled for all substrates. Radiolabeled substrates were prepared as described in Materials and Methods. The substrates after treating with EN (+ lanes) and without EN (- lanes) were run in sequencing gel (top panel). For substrates 1 and 2, parallel sequencing reactions were run to know the sites of nicking. Arrowheads show the position of the nick corresponding to site #3. Substrates 1 and 2 were run in 8% polyacrylamide gel at 60 W for 3 h. Substrates 3 and 4 were run in 12% polyacrylamide gel at 60 W for 3 h. Substrate #2 having 8 bp upstream of the nicked site did not support the nicking activity. Substrate #4, 27 bp (-11 to +16) was the minimum substrate length which supported EN activity.
control. The results showed that changing the nucleotide upstream of the nick influenced the EN activity only marginally. However changing the nucleotides downstream of the nick had a much larger impact. Mutation in position +2 from T to C decreased the activity to 17% and in position +3 from A to G increased the activity to 178% as shown in Fig. 21.

3. 8. 3. Importance of the downstream Cs at the nicking site:
We had earlier shown that a C residue (in top-strand sequence) was very frequently present 3-4 nucleotides downstream of the nick (Table 1 & Fig 19). The frequent occurrence of a C residue at a conserved location may be significant since *E. histolytica* genome is highly AT rich. Site #3 contains three C residues downstream of the nick (5'-GAATACCTC-3'). Using the procedure described above, the C residues were changed singly, or in pairs and the activities of the substrates quantified with respect to normal site #2. The result showed that changing a single C residue alone did not affect EN activity (Fig. 22A). However changing the first two Cs to A or T reduced the activity very significantly, while changing them to G had less effect (Fig. 22B). The third C also contributed to the EN activity (Fig. 22C). Compared with the second C alone, the first C alone retained greater activity (Fig. 22C). The results showed that downstream C residues did play a significant role in substrate recognition. From this data, the preferred recognition sequence for EN was deduced to be 5'-AATGCC-3'

3'-TTACGG-5'.

3. 8. 4. Mutational analysis of site #2 in the 174 bp DNA:
To see whether the above deduced sequence (5'-AATGCC-3') was generally applicable, site #2 in the 174 bp region was also analyzed by mutation. The wild type sequence of this site was 5'-CAATGCA-3'. The site #1 was used as internal control. Transition mutations were introduced in oligos used to PCR amplify 85 bp fragment from the 174 bp region. In agreement with the results obtained for site #3, it was found that mutation in position +2 from T to C decreased the activity to 37%. Mutation in position +3 from G to A had less effect but reduced the activity to 70%.
Fig. 21. Mutation analysis of sequences immediately surrounding site #3. (A) Bottom Panel. Desired mutations were introduced in forward primer. Only the mutated nucleotides are indicated; the rest being the same in all. Horizontal arrows numbered 1-5 and C show the length and position of the forward primers containing appropriate mutations. C marks the control primer. Reverse primer was same for all substrates, and is also indicated by horizontal arrow. Forward primer and radiolabelled reverse primer were used to PCR amplify 115 bp substrate from 174 bp DNA fragment. Radiolabelled substrates were prepared as described in Materials and Methods. (Top Panel) The substrates after treating with EN (+ lanes) and without EN (- lanes) were run in 8% denaturing polyacrylamide gel at 60 W for 3 h. The bands corresponding to site #3 and site #2 are indicated by arrowhead. (B) The ratio of band intensity of site #3 / site #2 for each mutated substrate was calculated and compared with control substrate to calculate the percentage of activity.
Fig. 22. Mutation analysis of downstream C nucleotides in site #3. The downstream C nucleotides are shaded in gray. Horizontal arrows (number 1-7) show the length and position of primers used for mutation; C is control substrate. Reaction conditions and electrophoretic analysis were same as described in the legend of Fig. 21. The percentage activity compared to normal site #2 is shown. (A) Nicking profile of EN protein and percentage activity after changing single C residue. (B) Nicking profile of EN protein and percentage activity after changing the first two C residues together.
3rd C also contributes

1st C alone retains more activity than 2nd C alone

Fig. 22 (continued). Horizontal arrows (number 8-11) show the length and position of primers used for mutation. (C) Nicking profile of EN protein and percentage activity after changing the third C residue in combination with the other Cs.
Changing GC to AA reduced the activity to 8% (Fig. 23). From the above results the preferred recognition sequence for EN was deduced to be

\[ 5' -\text{AATGC} - 3' \]

\[ 3' -\text{TACG} - 5' \]

3.8.5. Role of nucleotides some distance away from the nicking site:

Nucleotides in the vicinity of the nicking site were checked for their role, if any, in substrate recognition. A 37 bp substrate containing 15 bp upstream and 22 bp downstream of the nick in site #3 of the 174 bp region was used as substrate. Transition mutations were introduced in every alternate nucleotide, keeping the central 9 bp (GAATACCTC) unchanged (Fig. 24). This increased the GC content of the substrate from 13.5% to 46%. Enzyme activity in the mutated substrate compared with the wild type showed that nucleotide sequence some distance away from the nicking site did not greatly influence EN activity.

3.8.6. Nicking activity with artificial substrates:

The above substrates were derived from a natural *E. histolytica* sequence in which EhSINE1 was known to insert. A completely artificial substrate was next tested for enzyme activity with EN. A simple AT-rich sequence was designed for this purpose (S32AT) (Fig. 25). It showed a major nicking site towards the middle of the fragment. Sequences in the vicinity of the nicking site were modified, especially to check if the presence of C residues downstream of nick would affect enzyme activity. The results obtained from this analysis are as follows.

1. The most prominent bottom-strand nick was in the sequence

\[ 5' -\text{AAT} - 3' \]

located towards the middle of the 32 bp fragment (S32AT).

2. Introduction of two C nucleotides immediately downstream of AAT increased EN activity substantially (substrate, S32C).

3. Nicking was reduced if AAT was replaced by GGC (S32G) and abolished when it was replaced with TTT (S32T).

4. A minimum of 15 nucleotides downstream of the nick seemed to be necessary for activity. Thus, if the downstream sequence was extended by three
Fig. 23. Mutation analysis of sequences immediately surrounding site #2 in. Substrates were prepare as described in Fig. 21. Bottom panel, Forward primers containing indicated mutations and radiolabelled reverse primer were used to PCR amplify 85 bp substrate from 174 bp DNA fragment. The sequence of PCR amplified product is shown at the bottom. The region studied by mutation is shaded in gray. The substrates after treating with EN (+ lanes) and without EN (- lanes) were run in 8% denatured polyacrylamide gel at 60 W for 3 h (Top Panel). The nicks in site #2 and site #1 are marked by arrowhead. The ratio of band intensity of site #2 / site #1 for each mutated substrate was calculated and compared with control substrate to calculate the percentage of activity. 1, 2, 3, 4 and 5 mark the mutated substrates; C denotes control substrate.
Fig. 24. Mutation analysis of nucleotides some distance away from nicking site. The sequences of mutated substrate (M) and control substrate (C) are identical in the 9-bp shaded part. In the remaining part the mutated substrate is more GC rich (see text for details). The primers used for synthesizing the substrates are marked by horizontal arrows and bottom strand primer was labeled for both the substrates (*). The substrates were prepared by annealing the overlapping complementary oligos followed by gap filling and PCR. The substrates, after treating with EN, were separated through 12% denaturing polyacrylamide gel at 50-60 W for 3 h. Lane 1, untreated control DNA; lane 2, untreated mutant DNA; lane 3, EN treated control DNA; and lane 4 EN treated mutant DNA. Arrowheads indicate the position of nicks.
Fig. 25. Nicking activity with artificial substrates. The sequences of substrates used are shown at the bottom and the primers used for PCR are marked at the two ends by horizontal arrows. Asterisk (*) marks the labeled primer. The substrate labelled "Control" was derived from site #3 in 174 bp DNA fragment. It is 30 bp, spanning -14 to +16 from nick in site #3. Substrate S32AT gave two prominent nicks towards the middle of the fragment. The other substrates were generated by changing (or adding) the nucleotides shown in bold in S32AT sequence. The prominent nicks in each substrate are shown by arrowhead. Substrates after treating with EN were separated through a 12% PAGE gel at 60 W for 3 h.
nucleotides (S35AT) two extra nicks were generated. However the extra nicks were abolished if the sequence was made T-rich (S35T).

A limitation of this experiment was that it was not possible to obtain a quantitative estimate of change in enzyme activity as a result of sequence changes in the substrate, since there was no internal control. However, the results permit us to state that the EhLINE1-encoded EN, while being flexible in its sequence requirement, has a strong preference for nicking the bottom-strand between A and T residues located upstream of CC (top strand sequence). This is in keeping with the results obtained with the naturally occurring target site analyzed in the 174 bp DNA fragment (Figs. 20-24).

3.9. Biochemical properties of EhLINE1 EN:

EhLINE EN activity was optimized with regard to, pH, temperature and salt, using supercoiled pBS DNA as substrate. EN was active over a broad range of pH. In the range tested (buffers below pH-6.0 were not tested) optimum activity was at pH 6.0-8.0 (Fig. 26A). At higher pH, the SC to OC conversion was efficient. However the OC to linear conversion was inefficient, leading to accumulation of OC DNA. EN activity varied with temperature, with the optimum activity being at 37°C (Fig. 26B).

Activity of EN was tested at different NaCl concentrations (Fig. 27). From the disappearance of SC form, it was clear that at very high salt (500mM) the enzyme was very weakly active. At 250 mM NaCl, the SC form was fully converted to OC. However, the OC to linear conversion was very poor. Between 50-100 mM NaCl, all three activities (SC to OC, OC to linear, linear to smear) were visible. The enzyme activity dropped at NaCl concentration below 25 mM, since there was substantial persistence of SC DNA. The same results were seen when a time course was done at different NaCl concentrations (Fig. 28). An interesting observation was that at low salt (5-25 mM NaCl) the kinetics of conversion of SC to OC showed an initial fast phase followed by a slow phase (Fig. 29). At 250 mM NaCl the kinetics of SC to OC conversion was the same as at 100 mM NaCl, although conversion of OC to linear was very slow at 250 mM. The divalent cation Mg$^{2+}$ was required for EN activity. However, residual activity was seen even without added Mg$^{2+}$, probably due to trace
Fig. 26. (A) Activity of EN at different pH. Super coiled pBS DNA (200 ng) and ENp (80 ng) were incubated in different buffers of same strength (50mM) ranging from pH 6.0-9.0 at 37°C for 1h. The concentrations of other components (100 mM NaCl, 10mM MgCl₂ and 1mM DTT) were same for all. The samples were run in 1% agarose gel at 3V/cm for 3h and visualized by ethidium bromide staining. pH 6.0, 7.0 and 9.0 buffers were made by using PIPES, HEPES and CHES respectively. Tris base was used to make buffers of pH 7.5 and 8.0. OC, open circular; L, linear; SC, supercoiled.

(B) Activity of EN at different temperatures. SC pBS DNA (100 ng) was incubated with EN (40 ng) at the indicated temperatures. The buffer composition was 50 mM Tris-Cl [pH 7.5], 100 mM NaCl, 10 mM MgCl₂ and 1 mM DTT. The samples were run and visualized as described in panel A.
Fig. 27. Effect of NaCl concentration on EN activity. The reaction conditions, and electrophoretic analysis were same as described in the legend to Fig. 26B, except the concentration of NaCl in reaction buffer was varied, as indicated.
Fig. 28. Time course of the endonucleolytic cleavage of pBS DNA with ENp at the indicated NaCl concentrations. The reaction conditions and electrophoretic analysis were same as described in Fig. 26B. Lane C-Time 0'(zero min).
Fig. 29. Kinetics of the nicking activity of ENp at different salt concentration. SC band intensity was quantified using “Scion PC Image Software” and plotted against time of incubation. Intensity of SC band in lane C (Fig. 28) was taken as 100% in each panel.
contamination in the buffer (Fig. 30). Optimum activity was seen between 10-20 mM MgCl₂.
Fig. 30. Effect of Mg$^{2+}$ concentration on EN activity. The reaction conditions and electrophoretic analysis are same as described in Fig. 26B, except the concentration of MgCl$_2$ in reaction buffer was varied as indicated.