

Characterization of the restriction enzyme-like endonuclease encoded by the *Entamoeba histolytica* non-long terminal repeat retrotransposon EhLINE1

Vijay Pal Yadav¹, Prabhat Kumar Mandal^{2,*}, Desirazu N. Rao³ and Sudha Bhattacharya¹

¹ School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, India

² School of Life Sciences, Jawaharlal Nehru University, New Delhi, India

³ Department of Biochemistry, Indian Institute of Science, Bangalore, India

Keywords

EhLINE; *Entamoeba histolytica*; nicking endonuclease; restriction endonuclease-like endonuclease (RE-like endonuclease); retrotransposon-encoded endonuclease

Correspondence

S. Bhattacharya, School of Environmental Sciences, Jawaharlal Nehru University, New Mehrauli Road, New Delhi 110067, India
Fax: +91 11 26172438
Tel: +91 11 26704308
E-mail: sb@mail.jnu.ac.in;
sbjnu110@gmail.com

*Present address

Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

(Received 11 September 2009, revised 26 September 2009, accepted 30 September 2009)

doi:10.1111/j.1742-4658.2009.07419.x

The genome of the human pathogen *Entamoeba histolytica*, a primitive protist, contains non-long terminal repeat retrotransposable elements called EhLINES. These encode reverse transcriptase and endonuclease required for retrotransposition. The endonuclease shows sequence similarity with bacterial restriction endonucleases. Here we report the salient enzymatic features of one such endonuclease. The kinetics of an EhLINE1-encoded endonuclease catalyzed reaction, determined under steady-state and single-turnover conditions, revealed a significant burst phase followed by a slower steady-state phase, indicating that release of product could be the slower step in this reaction. For circular supercoiled DNA the K_m was 2.6×10^{-8} M and the k_{cat} was 1.6×10^{-2} sec⁻¹. For linear *E. histolytica* DNA substrate the K_m and k_{cat} values were 1.3×10^{-8} M and 2.2×10^{-4} sec⁻¹ respectively. Single-turnover reaction kinetics suggested a nonco-operative mode of hydrolysis. The enzyme behaved as a monomer. While Mg²⁺ was required for activity, 60% activity was seen with Mn²⁺ and none with other divalent metal ions. Substitution of PDX₁₂₋₁₄D (a metal-binding motif) with PAX₁₂₋₁₄D caused local conformational change in the protein tertiary structure, which could contribute to reduced enzyme activity in the mutated protein. The protein underwent conformational change upon the addition of DNA, which is consistent with the known behavior of restriction endonucleases. The similarities with bacterial restriction endonucleases suggest that the EhLINE1-encoded endonuclease was possibly acquired from bacteria through horizontal gene transfer. The loss of strict sequence specificity for nicking may have been subsequently selected to facilitate spread of the retrotransposon to intergenic regions of the *E. histolytica* genome.

Introduction

Non-long terminal repeat (LTR) retrotransposable elements are widespread in eukaryotic genomes. They possess either one or two ORFs, which encode all the

functions needed for retrotransposition. These functions include reverse transcriptase and endonuclease activities, in addition to a nucleic acid-binding

Abbreviations

EhLINE1-EN, EhLINE1-encoded endonuclease; EhLINE1-ENM, EhLINE1-encoded endonuclease mutant; LTR, long terminal repeat; REL-ENDO, restriction endonuclease-like.

property needed to form a ribonucleoprotein particle. In non-LTR retrotransposons with two ORFs, the reverse transcriptase and endonuclease domains are part of the same ORF, with the nucleic acid-binding domain on a separate ORF. Phylogenetic analysis based on 440 amino acid residues of the reverse transcriptase domain, and on the nature and arrangement of other protein domains (endonuclease and nucleic acid-binding domains), has divided these retrotransposons into five distinct groups [1]. Of these, the R2 group is considered to be the most ancient. In this group the endonuclease domain is typically located at the C-terminus of the protein. The elements belonging to this group generally insert in a site-specific manner in repetitive regions of the genome. For example, members belonging to some clades of the R2 group insert in tandemly repeated spliced leader genes [2,3], while members of other clades insert in 28S rRNA genes [4]. In concurrence with their site-specific mode of insertion, the endonucleases encoded by these elements show considerable sequence similarity with restriction endonucleases and have been classified as restriction endonuclease-like (REL-ENDO) enzymes [5]. By contrast, the endonuclease encoded by the other four groups of non-LTR elements is an apurinic/apyrimidinic endonuclease [6].

The genome of the human pathogen *Entamoeba histolytica* contains hundreds of copies of three related families of non-LTR retrotransposable elements called EhLINEs/SINEs [7]. From comparative sequence alignment analysis, the EhLINEs are closest to the R4 clade in the R2 group of elements, and the endonuclease encoded by EhLINEs shares sequence similarity with REL-ENDO enzymes. EhLINEs/SINEs insert in intergenic regions of the *E. histolytica* genome. These regions (apart from being AT-rich) do not share obvious sequence similarity. The first step in the insertion of a non-LTR retrotransposon into its target site is a nick created by the element-encoded endonuclease at the target site [8]. The site specificity of the endonuclease is expected to be an important determinant in the selection of target sites for retrotransposon insertion. Thus, the origin and evolution of the endonucleases encoded by retrotransposons is of great interest because it has bearing on the origin of the elements themselves and on their subsequent spread through the genome. The presence of the REL-ENDO type of domain in eukaryotic retrotransposons is especially fascinating because this type of endonuclease is found only in bacteria. Therefore, it is of interest to understand the biochemical properties of the REL-ENDO endonuclease encoded by retrotransposons and to compare it with bacterial restriction endonucleases.

To understand the mechanism, and to predict the genomic sites where EhLINEs/SINEs may preferentially insert, we sought to functionally characterize the EhLINE1-encoded endonuclease (EhLINE1-EN). The endonucleases encoded by retrotransposons of the R2 group have not been well characterized to date. This is the first attempt to address the kinetic properties of an endonuclease belonging to this category. The EhLINE1-EN harbors the conserved catalytic sequence motif, PDX₁₂₋₁₄D, which is required for activity [9] and is closest to Type IIS restriction endonucleases. Although the enzyme is not strictly sequence specific, it nicks preferentially at certain hot-spots. The consensus hot-spot sequence is 5'-GCATT-3', with nicks occurring between A-T and T-T [10]. Here we report some of the salient enzymatic features of the EhLINE1-EN, and discuss the results with respect to its role in retrotransposition.

Results and Discussion

Kinetics of the endonuclease-catalyzed reaction with pBS supercoiled DNA substrate under steady-state conditions

As described earlier [9], recombinant EhLINE1-EN protein expressed and purified from *Escherichia coli* was used in the present study. The activity of this enzyme was optimized with regard to pH, temperature, Mg²⁺ and salt using pBS supercoiled DNA as the substrate. The optimum activity was in the pH range 6.0–8.0, at 37 °C, and at Mg²⁺ and NaCl concentrations of 10 and 100 mM respectively. Plasmid pBS DNA has previously been shown to be a substrate for EhLINE1-EN. Upon incubation with the endonuclease, the supercoiled plasmid was first converted into an open circle and then into a linear form, which was subsequently degraded upon further incubation [9].

To determine the kinetics of an EhLINE1-EN-catalyzed reaction under steady-state conditions, reactions were carried out with the enzyme at a concentration of 2 nM and with pBS DNA at a concentration of 5–125 nM. Under these conditions, cleavage of all the plasmid molecules requires multiple enzyme turnover, each involving binding of enzyme to a DNA molecule, a catalytic reaction and subsequent dissociation from the product(s) before acting upon another molecule of DNA. The time-course experiment was performed with different concentrations of the substrate. Samples were withdrawn from the reaction mixture at the time-points indicated and analyzed using agarose-gel electrophoresis. Either the disappearance of supercoiled plasmid DNA or the appearance of open circle and

linear DNA was used as a measure of product formed. Similar results were obtained in both cases. As mentioned in the Materials and methods, all time-course results were the average of at least three independent determinations. The variation observed at each time-point was $< 4\%$ of the mean value (0.05–4.0). As shown in Fig. 1A, after incubation for 40 min (at the range of substrate concentrations mentioned) 80–90% of supercoiled DNA was still intact. Analysis of the time course of DNA hydrolysis by endonuclease revealed a significant burst phase followed by a slower steady-state phase (Fig. 1A). This pattern was observed at all concentrations of substrate. Such a reaction profile is consistent with product-burst

kinetics. The initial burst in product formation indicates that release of product from enzyme could be the slower step in this reaction. Clearly, the enzyme functions catalytically rather than stoichiometrically.

In order to determine the rate of reaction at each substrate concentration, the slope of the burst phase was considered to be the initial velocity (Fig. 1B). Although the variation in values of each data point was up to 4% in three replicates (as mentioned above), the slopes for each set showed very little variation (up to 0.3%). In subsequent experiments also (Figs 3B and 5A) where slopes were plotted, the variation observed was minimal. The rate of DNA cleavage was initially linear with increasing concentration of DNA substrate and saturated at around 50 nM (Fig. 1C). The substrate saturation followed a typical hyperbolic curve. Kinetic parameters (K_m and k_{cat}) were calculated from a Lineweaver–Burk plot (Fig. 1C). The K_m for pBS DNA was calculated to be $2.6 \pm 0.018 \times 10^{-8}$ M. The catalytic constant, k_{cat} , ($V_{max}/[E]$) was determined to be $1.6 \pm 0.01 \times 10^{-2} \text{ sec}^{-1}$. The K_m for pBS DNA was comparable with the low K_m values (0.5–17 nM) of restriction endonucleases determined with different DNA substrates under different conditions of buffer and temperature [11]. Furthermore, the turnover number of the enzyme was in the lower range of that reported for restriction endonucleases (1.6×10^{-2} – $16.6 \times 10^{-2} \text{ sec}^{-1}$) [11]. The low turnover number of a retrotransposon-encoded endonuclease may have a

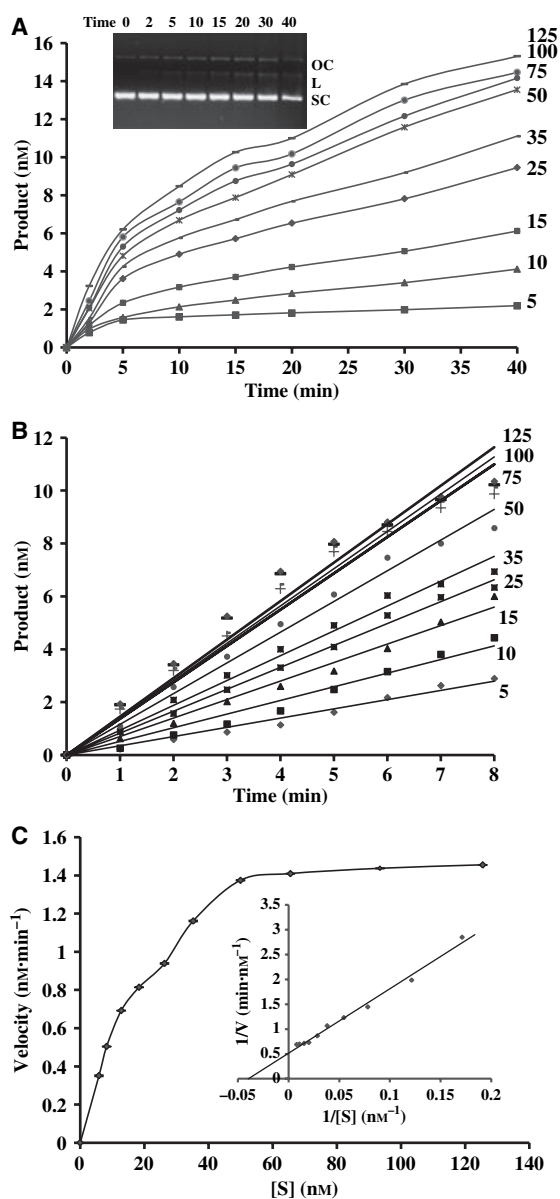


Fig. 1. Kinetics of supercoiled pBS DNA cleavage by endonuclease. (A) Steady-state kinetics of DNA cleavage by endonuclease. DNA cleavage assays were carried out with 2 nM EhLINE1-EN in a reaction mixture containing increasing concentrations (5–125 nM) of pBS DNA at 37 °C. Aliquots were withdrawn at different time-points (0–40 min) during the reaction and assayed by electrophoresis through 0.8% agarose (inset). The concentration of the supercoiled DNA form at each time-point was quantified as described in the Materials and methods. The disappearance of the supercoiled form of pBS DNA with time was plotted for the indicated concentrations of substrate (nM). L, linear; OC, open circle; SC, supercoiled. (B) Determination of initial velocities of the reaction. DNA cleavage assays were carried out as mentioned above. Aliquots were withdrawn every minute from 0 to 8 min. The disappearance of the supercoiled form of pBS DNA over time was plotted for the indicated concentrations of substrate (as in Fig 1A) and the slopes thus obtained were taken as the initial velocity at corresponding substrate concentrations. (C) DNA cleavage as a function of substrate concentration. Initial reaction velocities, obtained as described above, were plotted as a function of substrate concentration. A Lineweaver–Burk plot (inset) was used to calculate the kinetic parameters K_m and k_{cat} . The data are expressed as the average of three independent determinations, as mentioned in the Results and Discussion, and the standard deviation is indicated as error bars (\pm SD).

significant role in limiting the rate of retrotransposition events in the *E. histolytica* genome.

Kinetics with *E. histolytica*-specific 176-bp linear DNA substrate

We have shown (in an earlier publication) that a 176-bp oligonucleotide fragment, derived from a region of *E. histolytica* genome where EhSINE1 is known to insert, is nicked by this endonuclease at three major hot-spots [9]. One of these hot-spots (#3) corresponds with the EhSINE1 insertion site. We used this 176-bp fragment to study the kinetics of nicking at the hot-spot #3 (Fig. 2A).

Time-course measurements were performed under steady-state conditions with the 176-bp dsDNA fragment in which the bottom strand was end labeled. Reactions were carried out with the enzyme at a concentration of 2 nM and with DNA at a concentration of 2–125 nM. Samples were withdrawn from the reaction mixture at the indicated time-points and analyzed by denaturing PAGE, as mentioned in the Materials and methods. It was observed that more than 90% of

labeled DNA was intact under these conditions. The intensity of bands at site #3 (out of the three hot-spots) was quantified by densitometry in a phosphorimager and plotted against time. The data are the average of three independent determinations, with variation in data points being < 2.8% of the mean (0.34–2.8). Unlike supercoiled plasmid DNA, the rate of cleavage was found to be monophasic. In order to determine the rate of reaction at each substrate concentration, the slope of initial time-points (0–10 min) was considered as the initial velocity. The rate of DNA cleavage was initially linear (Fig. 2B) with an increasing concentration of DNA substrate and later saturated at around 60 nM. The substrate saturation followed a typical hyperbolic curve, as in the case of the pBS DNA substrate. Kinetic parameters (K_m and k_{cat}) were calculated from a Lineweaver–Burk plot (Fig. 2B). The K_m was calculated to be $1.3 \pm 0.032 \times 10^{-8}$ M. The catalytic constant, k_{cat} ($V_{max}/[E]$), was determined to be $2.2 \pm 0.055 \times 10^{-4}$ sec⁻¹. The kinetic parameters obtained for the 176-bp linear DNA substrate suggest that it is a less efficient substrate than pBS supercoiled DNA. This could

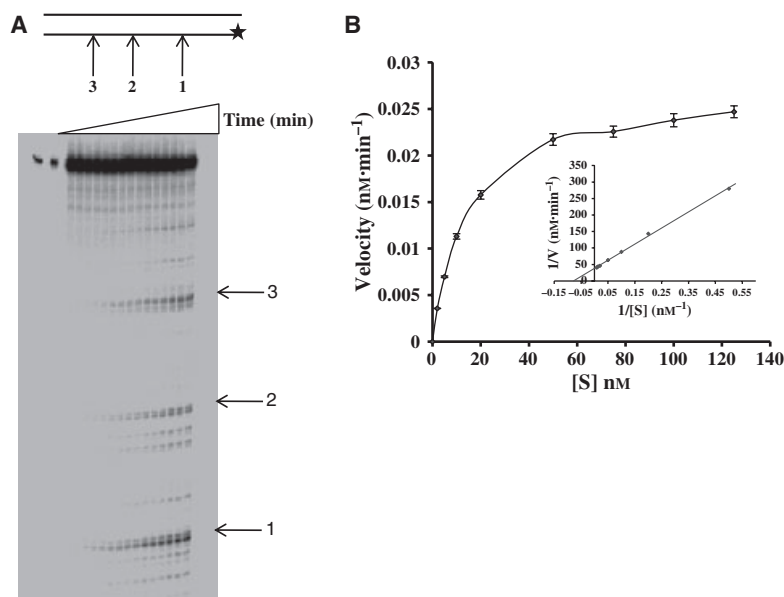


Fig. 2. Kinetics of 176-bp linear DNA cleavage by endonuclease. (A) Steady-state kinetics of DNA cleavage by endonuclease. DNA cleavage assays were carried out with 2 nM EhLINE1-EN in a reaction mixture containing increasing concentrations (2–125 nM) of 176-bp linear DNA at 37 °C. Aliquots were withdrawn at different time-points (0–60 min) during the reaction and were assayed by electrophoresis through a 6% polyacrylamide denaturing gel. The intensity of bands at site #3, at different time-points, was quantified as mentioned in the Materials and methods. The figure shows 176-bp end-labeled DNA (upper panel) and the time course of the reaction with 10 nM substrate (lower panel). The positions of three hot-spots are indicated by the numbers 1, 2 and 3 in both panels, and the end labeling of DNA is represented by a star. In the first two lanes of the autoradiograph (lower panel), different dilutions of end-labeled substrate DNA were loaded. (B) DNA cleavage as a function of substrate concentration. Initial velocities of the reaction were plotted as a function of substrate concentration. A Lineweaver–Burk plot (inset) was used to calculate the kinetic parameters K_m and k_{cat} . The data are expressed as the average of three independent determinations, as mentioned in the Results and Discussion, and the standard deviation is indicated as error bars (\pm SD).

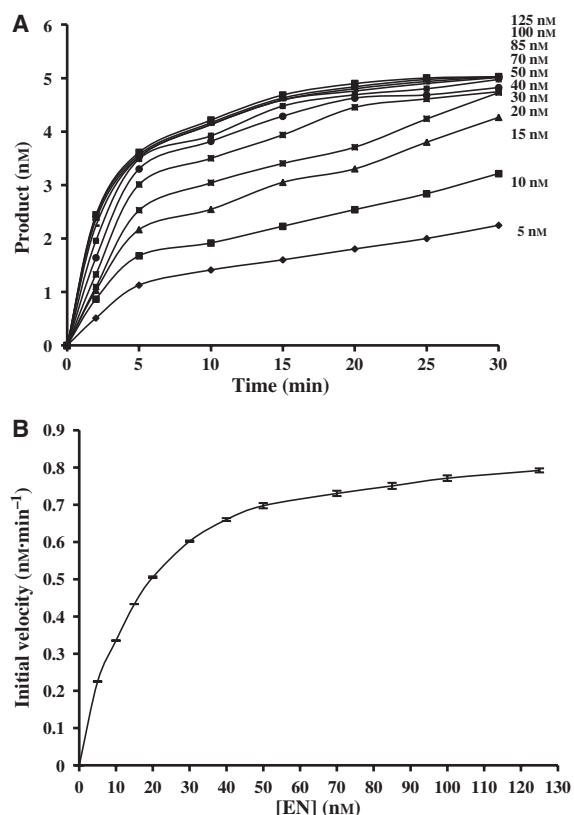


Fig. 3. Rate of pBS DNA cleavage at different concentrations of EhLINE1-EN under single turnover conditions. DNA cleavage assays were carried out with 5 nM pBS DNA in a reaction mixture containing increasing concentrations (5–125 nM) of EhLINE1-EN at 37 °C. Aliquots were withdrawn at different time-points (0–30 min) and assayed by electrophoresis, as mentioned in the Materials and methods. (A) The disappearance of the supercoiled form of pBS DNA with time was plotted for the indicated concentrations of EhLINE1-EN (nM). (B) Initial velocity versus EhLINE1-EN concentrations. The data are expressed as the average of three independent determinations, as mentioned in the Results and Discussion, and the standard deviation is indicated as error bars (\pm SD).

reflect the differences in the nature of the two substrates (linear versus supercoiled; 176 bp versus 2.9 kb). Alternatively, it could mean that although the endonuclease has high affinity for *E. histolytica* sequences, it nicks these very slowly to avoid excessive retrotransposition activity. At this point the data cannot distinguish between these possibilities. Subsequent kinetic analysis was performed with the supercoiled pBS substrate.

Kinetics of the endonuclease-catalyzed reaction under single-turnover conditions

Single-turnover enzyme reactions were carried out where the enzyme was in molar excess over the

substrate. Different concentrations of EhLINE1-EN (5–125 nM) were incubated with 5 nM pBS DNA substrate, and the reactions were monitored over time (Fig. 3A). The data represent the average of three independent determinations, with variation in the data points being $< 1\%$ of the mean (0.05–1.0). Each of the time courses showed an initial burst in product formation followed by a slower rate of product formation, as observed under steady-state experiments. However, the burst phases in single-turnover reactions were more rapid compared with steady-state reactions. About 80% of the pBS supercoiled DNA was converted into open circle and linear forms at the highest range of enzyme concentrations used (70–125 nM), within the first 10 min (Fig. 3A). In order to determine the rate of reaction at a given enzyme concentration, the slope of the burst phase was considered as the initial velocity. When initial velocities were plotted against increasing enzyme concentration, a linear relationship was obtained, which later plateaued at an enzyme concentration of around 40 nM (Fig. 3B), suggesting a noncooperative mode of hydrolysis of DNA by EhLINE1-EN up to a certain concentration. It is also likely that at higher concentrations of enzyme, limited substrate is available for the enzyme.

Order of binding of substrates

For the catalytic cycle of endonuclease, binding of DNA and Mg^{2+} could occur in a random order or a sequential order. To determine this, EhLINE1-EN was pre-incubated with pBS DNA or with $MgCl_2$ for 10 min on ice, and the reaction was initiated by adding $MgCl_2$ or pBS DNA, respectively. The Mg^{2+} concentration used in these experiments was lower than the K_d determined by fluorescence measurements. Three separate time-course experiments, with different mixing orders, were performed. The data represent the average value of three independent determinations, with variation in data points being $< 2.1\%$ of the mean (0.03–2.1). Under saturating substrate conditions, each time course showed an initial burst followed by a slower steady-state phase. As seen in Fig. 4, the pre-incubation of endonuclease with either DNA or $MgCl_2$, or starting the reaction with endonuclease, had no influence on the rate of product formation. In either set of experiments, pre-incubation of mixtures at 37 °C instead of at 4 °C, did not affect the rate of cleavage. From this it appears that the enzyme may have a random order of substrate binding during catalysis. However, detailed kinetic analysis needs to be performed to confirm the order of substrate binding.

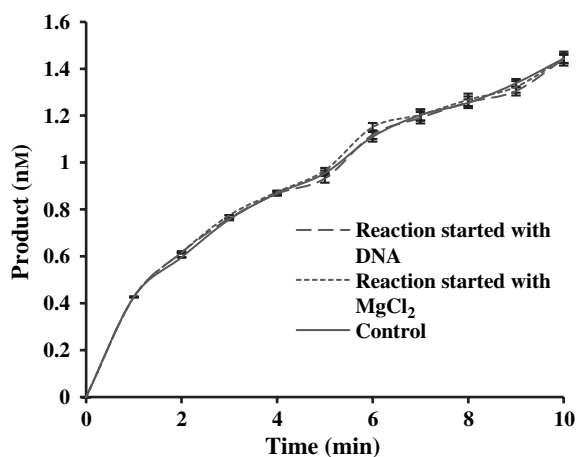


Fig. 4. Pre-incubation analysis of EhLINE1-EN. DNA cleavage reactions were carried out in reaction buffer (as mentioned in the Materials and methods) containing 2 nM EhLINE1-EN, 10 nM supercoiled pBS DNA and 5 mM MgCl₂. The cleavage reaction was started either by the addition of DNA to a solution containing enzyme pre-incubated with MgCl₂, or by the addition of MgCl₂ to a solution containing enzyme pre-incubated with DNA. In a control experiment, the cleavage reaction was started by the addition of enzyme to a solution containing DNA and MgCl₂. Aliquots were withdrawn at the time-points indicated and the time courses were plotted as mentioned above. The data are expressed as the average of three independent determinations for each set of experiment, as mentioned in the Results and Discussion, and the standard deviation is indicated as error bars (\pm SD).

Determination of oligomeric status

To establish the relationship between the initial velocity of the reaction and the enzyme concentration, the rate of pBS DNA cleavage at different concentrations of endonuclease was determined. Varying concentrations (0.5–10 nM) of endonuclease were added to reaction mixtures containing 50 nM pBS DNA and incubated for 10 min. The pBS DNA substrate used in these reactions was in molar excess compared with the highest range of enzyme concentration. The reaction rates were determined as described in the Materials and methods. An average of three independent determinations, with variation being $< 2.9\%$ of the mean (0.08–2.9), was made. When initial velocities were plotted against the corresponding enzyme concentrations, a linear plot was obtained (Fig. 5A). A linear relationship between initial velocities of the reaction and the corresponding enzyme concentration suggests that the reaction catalyzed by this enzyme follows first-order kinetics. It also indicates a noncooperative mode of catalysis of DNA by this endonuclease, and that the enzyme behaves as a monomer at a wide range of concentrations.

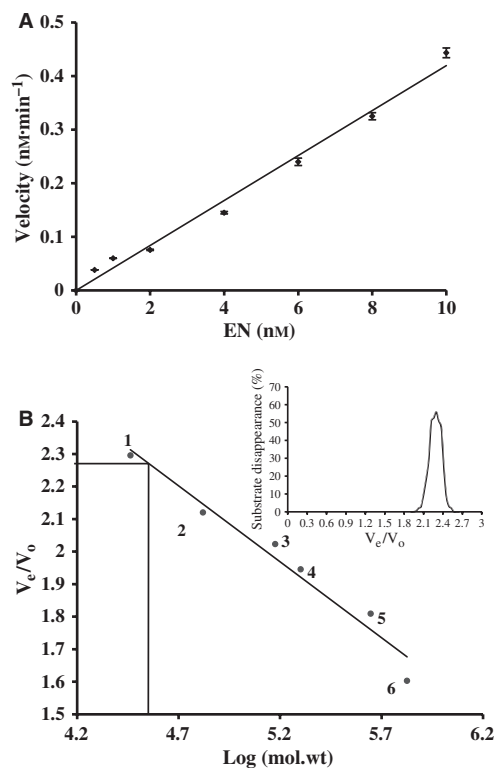


Fig. 5. (A) Oligomeric status of EhLINE1-EN. The rate of pBS DNA cleavage at different EhLINE1-EN concentrations is shown. DNA cleavage assays were carried out with 50 nM supercoiled pBS DNA in a reaction mixture containing increasing concentrations (0.5–10 nM) of EhLINE1-EN at 37 °C. The reactions were stopped after 10 min and assayed as mentioned in the Materials and methods. Initial reaction velocities were plotted as a function of enzyme concentration. The data are expressed as the average of three independent determinations, as mentioned in the Results and Discussion, and the standard deviation is indicated as error bars (\pm SD). (B) Determination of the molecular mass under native conditions by size-exclusion chromatography. A standard curve was derived from the elution profiles of the standard molecular mass markers, with V_e corresponding to the peak elution volume of the protein and V_o representing the void volume of the column, determined using blue dextran (2 000 000). The peak position of the EhLINE1-EN is indicated by a line. The molecular mass standards were: 1, carbonic anhydrase (29 kDa); 2, BSA (66 kDa); 3, alcohol dehydrogenase (150 kDa); 4, β -amylose (200 kDa); 5, apoferritin (443 kDa); 6, thyroglobulin (669 kDa). Inset: elution profile of EhLINE1-EN through a Superose-6 column. A 5 μ L sample of each eluted fraction was used in a cleavage assay with 5 nM supercoiled pBS DNA by incubation in reaction buffer at 37 °C for 30 min. Endonuclease activity was assayed as mentioned in the Materials and methods. Endonuclease activity was eluted in fractions 15.6–19.2 mL ($V_e/V_o = 2.023$ – 2.49), as shown.

Gel-permeation chromatography, using an analytical Superose-6 column, was performed to determine the molecular weight and oligomeric status of

EhLINE1-EN in solution. Individual fractions were assayed for endonuclease activity with pBS substrate (Fig. 5B). The peak fraction eluted at 17.4 mL, which corresponded to an approximate apparent molecular mass of 38 kDa (Fig. 5B). This was comparable to the calculated molecular mass of the monomeric EhLINE1-EN species (35.5 kDa), suggesting that the enzyme exists as a monomer in solution under native conditions. From our kinetic measurements described earlier, EhLINE1-EN appears to function as a monomer. The Type IIS restriction endonuclease *FokI* is also monomeric in solution [12,13]. Interestingly, it has been shown that *FokI* dimerizes on the DNA substrate and cleavage is carried out by the dimer [14,15]. Kinetic studies of a variety of methyltransferases [16] and restriction endonucleases [17,18] reveal that these enzymes oligomerize in the presence of their DNA substrate. From kinetic measurements described earlier, EhLINE1-EN appears to function as a monomer, even in the presence of DNA.

Effect of metal ions on EhLINE1-EN

The endonuclease encoded by EhLINE1 is a restriction endonuclease-like endonuclease [9]. For the great majority of Type II restriction endonucleases (except *BfiI*), Mg^{2+} is an essential cofactor, which can be substituted with Mn^{2+} , Ca^{2+} , Fe^{2+} , Co^{2+} , Ni^{2+} , Zn^{2+} or Cd^{2+} , depending on the enzyme [19–21]. The effect of various metal ions on the cleavage activity of EhLINE1-EN was studied with supercoiled pBS DNA as the substrate in a standard buffer in the presence of 10 mM of the metal ions. For these experiments, all buffers and enzyme preparations were passed through Chelex-100 (Sigma). All metal ion stock solutions were prepared in Chelex-100-treated Milli-Q water. The endonuclease was inactive in the absence of divalent metal ions. Maximum activity was seen in the presence of Mg^{2+} . There was 60% activity in presence of Mn^{2+} compared with Mg^{2+} . Ca^{2+} , Cd^{2+} , Co^{2+} , Ni^{2+} , Zn^{2+} and Fe^{2+} did not support cleavage activity of this endonuclease (data not shown). The observed differential effect of divalent metal ions may be caused by differences in the binding of these ions to the endonuclease active site, or by the noninvolvement of some metal ions in catalysis, as reported for a number of restriction endonucleases [22–24]. In many endonucleases, either crystallographic or solution studies have implied at least two of the acidic residues of PDX₍₁₂₋₁₄₎D motif as being important in Mg^{2+} binding [22,25–27]. Therefore, it

will be interesting to determine the role of this motif in binding of divalent metal ions to the endonuclease.

Fluorescence spectroscopic analysis of EhLINE1-EN and EhLINE1-ENM

Previous work has shown that in several restriction endonucleases, substitution of either of the acidic residues in the PDX₁₂₋₁₄D motif with Ala resulted in no detectable enzyme activity [22], suggesting that both of these residues are important for metal ion binding. In a previous study, the PDX₁₂₋₁₄D motif was mutated to PAX₁₂₋₁₄D and a reduction of enzyme activity in this mutant (EhLINE1-ENM) was observed compared with the wild-type (EhLINE1-EN) enzyme [9]. The residual activity of the mutant enzyme was 10–30% of the activity of the wild-type enzyme. Appropriate controls were used to rule out any contaminating activity contributed by *E. coli*. Controls included an endonuclease-minus strain (ER2566, mutated in the *endA1* gene) used to express the EN and ENM proteins. Proteins purified from this *E. coli* strain gave identical results to those from *E. coli* BL21 (DE3). In another control, lysates from cells transformed with pET30b vector alone, and induced with isopropyl thio- β -D-galactoside (IPTG), were tested after passing through a Ni-nitrilotriacetic acid affinity column, and no endonuclease activity was found. Fluorescence spectroscopy was employed to evaluate the binding of metal ions to the enzyme. We took advantage of the presence of three tryptophan residues in the protein, and measured the change in tryptophan fluorescence intensity. Fluorescence quenching was observed with increasing concentrations of metal ions (Mg^{2+} , Ca^{2+} , Mn^{2+} ; quenching with Mg^{2+} is shown in Fig. 6A, B). However, monovalent cations, such as Na^+ , did not influence the fluorescence spectra of the protein (data not shown). The K_d values of three metal ions for EN and ENM proteins were calculated from a modified Stern–Volmer equation (Fig. 6C). The data presented in Table 1 show comparison of K_d values of EhLINE1-EN and EhLINE1-ENM for all three metal ions. Interestingly EN showed more affinity for Mn^{2+} than for Mg^{2+} . The order of affinity was $Mn^{2+} > Mg^{2+} > Ca^{2+}$, which is quite different from the order of activity ($Mg^{2+} > Mn^{2+} > Ca^{2+}$). For ENM protein the order of affinity was $Mg^{2+} > Ca^{2+} > Mn^{2+}$. The affinity of ENM protein for Mn^{2+} was 15-fold less than the affinity of EN for Mn^{2+} . In the case of Mg^{2+} and Ca^{2+} the affinity of ENM protein was only 2- and 1.5-fold less, respectively, compared with that of EN. These results indicate that Mg^{2+} and

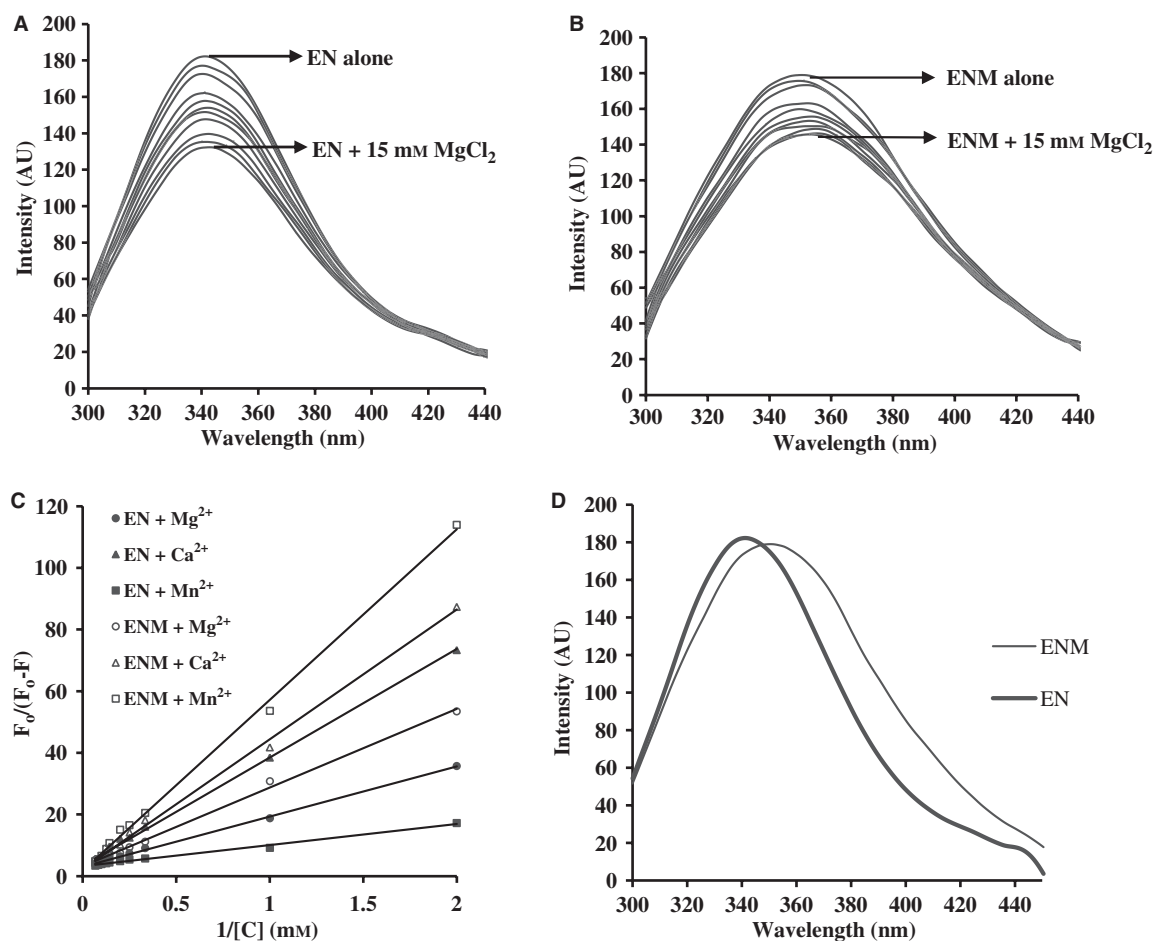


Fig. 6. Fluorescence spectral analysis of EhLINE1-EN and EhLINE1-ENM. Tryptophan fluorescence quenching analysis of (A) EN and (B) ENM. Proteins were incubated with increasing concentrations of MgCl₂ (final concentration 0.5–15 mM) and fluorescence measurements were made as described in the Materials and methods. (C) Modified Stern–Volmer plot. Tryptophan fluorescence quenching was carried out in the presence of Mg²⁺, Mn²⁺ or Ca²⁺. C, quencher concentration; F, fluorescence intensity in the presence of quencher; F₀, fluorescence intensity in the absence of quencher. (D) Fluorescence emission spectra of wild-type (EN) and mutant (ENM) proteins in the absence of quencher after excitation at 280 nm, as described in the Materials and methods.

Ca²⁺ bind EN and ENM proteins with similar affinities, although Ca²⁺ does not support the activity. A 15-fold decrease in affinity for Mn²⁺ in the case of the ENM protein implies that the first Asp residue of the PDX₁₂₋₁₄D motif may play a role in the binding of Mn²⁺. However, this residue may not be involved in binding to Mg²⁺ and Ca²⁺. Similar results have been reported for the restriction endonuclease *Pvu*II, where substitution of the first Asp residue of the PDX₁₂₋₁₄D motif did not affect the binding affinity of Mg²⁺, although it was required for proper positioning of this metal ion [22]. Interestingly, the emission spectra of the ENM protein showed a red shift of approximately 10 nm in the emission maxima (Fig. 6D). This shift indicates that in the ENM protein, tryptophan residues are almost completely exposed to the solvent and

the hydrophobic core of the protein is disturbed. Therefore, reduced activity of the ENM enzyme was not caused by the loss of metal binding, but may indeed be caused by a change in the tertiary structure of the protein.

CD spectroscopy

The fluorescence results described above indicate that the micro-environment of some of the aromatic residues of the protein could have been modified when the first Asp residue of PDX₍₁₂₋₁₄₎D motif was replaced with Ala. We therefore analyzed the differences in secondary and tertiary structure of the EN and ENM proteins by far- and near-UV CD spectroscopy, respectively. Near-UV CD spectra (250–300 nm; which are

Table 1. K_d values (mM) of EN and ENM for divalent metal ions.

	Mg ²⁺	Ca ²⁺	Mn ²⁺
EN	5.32	10.79	2.09
ENM	7.97	18.71	30.60

widely used to detect local conformational properties [28,29]) and far-UV CD spectra were recorded for EN and ENM proteins. In the far-UV region (200–250 nm) the CD spectrum of the ENM protein was nearly identical to that of the EN protein (data not shown), suggesting that the overall secondary structure of both proteins was quite similar. No significant changes in CD spectra were recorded upon the addition of substrate DNA to either EN or ENM proteins (data not shown). The fractions of secondary-structure components were estimated from far-UV CD spectra using Kohonen's self-organizing maps [30]. The estimated helical content was same for both proteins, and no detectable change in this content was noticed in the presence of DNA substrate (Table 2). However, there was a slight shift in CD spectra in the presence of Mg²⁺ ions in both cases (data not shown).

When CD spectra were recorded in the near-UV region, a remarkable difference was observed between the proteins (Fig. 7A). The molar ellipticity values of ENM at 250–280 nm were higher than those of EN, although the position of the observed peak and minima remained essentially unaltered. These results suggest that the single substitution of PDX₁₂₋₁₄D to PAX₁₂₋₁₄D causes a local conformational change in the protein tertiary structure, which could contribute to the reduction of enzyme activity in the ENM, substantiating our observation obtained by fluorescence spectroscopy.

Near-UV CD spectra were also recorded in the presence of Mg²⁺ and DNA. The spectra of EN and ENM proteins were slightly shifted in the presence of Mg²⁺ but the change in peak maxima was insignificant (data not shown). In the presence of the substrate DNA, a prominent shift in the spectrum was observed in the case of the EN protein (Fig. 7B). The shift observed in the case of the ENM protein was not significant (Fig. 7C). These observations suggest that the EN protein undergoes a conformational change upon

Table 2. Estimated fractions of secondary structure components of EN and ENM.

Secondary Structure	EN	ENM	EN+DNA	ENM+DNA
α -helix	0.346	0.346	0.346	0.346
β -sheet	0.226	0.226	0.226	0.226
Turn	0.121	0.121	0.121	0.121
Random coil	0.308	0.308	0.308	0.308

the addition of DNA, which may be critical for the enzyme activity. This result is consistent with the known behavior of restriction enzymes that exhibit major conformational changes upon binding to their DNA substrates [17,18].

In summary, the endonuclease encoded by EhLINE1 is similar to restriction endonucleases in many respects. For example, amino acid sequence comparison shows the presence of a PDX₁₂₋₁₄D motif similar to that found in the active site of Type IIS restriction endonucleases [31]. This endonuclease preferentially recognizes the asymmetric sequence 5'-GCATT-3' [10]. It displays a low K_m , suggesting high affinity for DNA, and a low turnover number that could be an evolutionary advantage to limit retrotransposition. The binding of the enzyme to DNA is accompanied by major conformational change. These similarities with bacterial restriction endonucleases suggest that the endonuclease encoded by EhLINE1, and other related non-LTR retrotransposons were possibly acquired from bacteria, through horizontal gene transfer. The loss of strict sequence specificity for nicking may have been subsequently selected to facilitate the spread of the retrotransposon to all intergenic regions of the *E. histolytica* genome. Analysis of more such endonucleases would reveal how these enzymes may have been engineered to suit the needs of retrotransposition. An understanding of the role of this endonuclease in target site selection would help in designing vectors for targeted gene insertion and genetic manipulation in *E. histolytica*, an organism in which it has so far not been possible to integrate genes by homologous recombination.

Materials and methods

Enzymes and substrate

Recombinant EhLINE1-EN and EhLINE1-ENM were purified from *E. coli* BL21 (DE3) clones, as described earlier [9]. *E. coli* strain ER2566 (*endA1*) was also used for protein expression and purification. Protein concentrations were estimated using the method of Bradford [32] with BSA as the standard. Supercoiled pBS plasmid was purified using a Qiagen plasmid purification kit. The 176-bp linear DNA was prepared by PCR amplification of *E. histolytica* genomic DNA, as described earlier [9]. The concentration of DNA was estimated by measuring absorption at 260 nm.

Steady-state and single turnover kinetics with pBS supercoiled DNA

Cleavage reactions were carried out in a buffer of 50 mM Tris/HCl (pH 7.5), 100 mM NaCl, 10 mM MgCl₂ and 1 mM

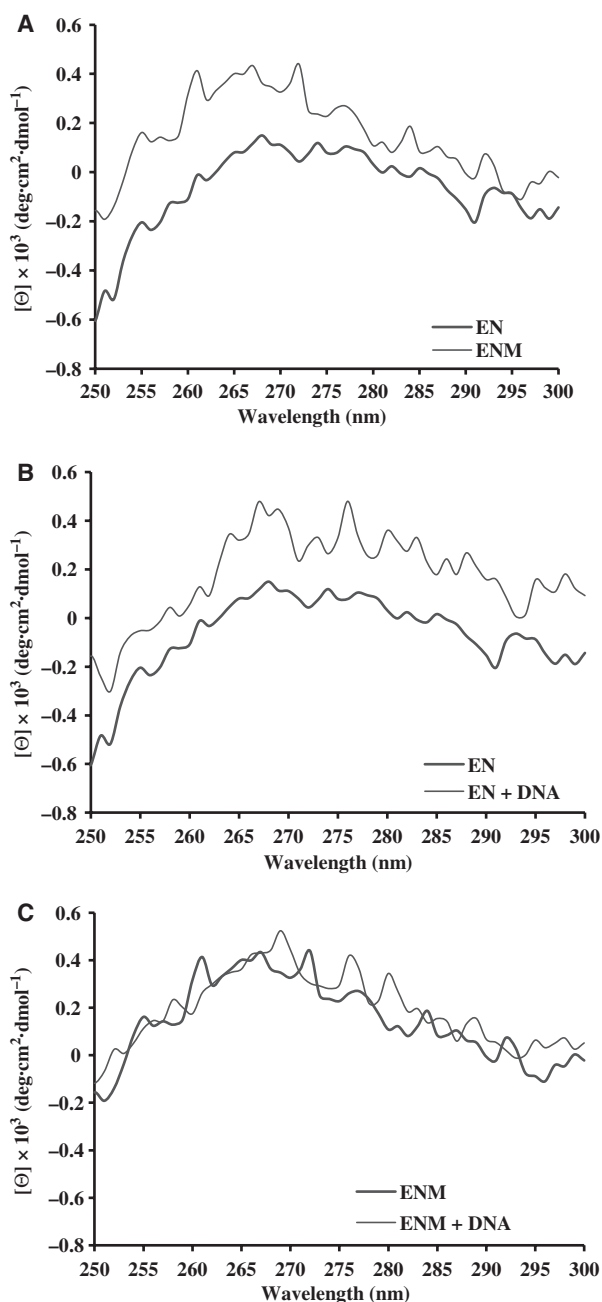


Fig. 7. Near-UV CD Spectra of EhLINE1-EN and EhLINE1-ENM. The proteins (5 μM) were dialyzed against 10 mM sodium phosphate buffer (pH 7.4) containing 100 mM NaCl. CD spectra were recorded from 250 to 300 nm as described in the Materials and methods. (A) Near-UV CD spectra of EN and ENM proteins. (B, C) Near-UV CD spectra of EN and ENM proteins, respectively, in the presence of 0.5 μM of the 34-bp DNA fragment.

dithiothreitol at 37 $^{\circ}\text{C}$. Enzyme and substrate concentrations were used as indicated in each experiment. The reactions were stopped by removing 20- μL aliquots of and

mixing them with 5 μL of stop mix [100 mM EDTA, 10 mM Tris/HCl (pH 8.0), 30% glycerol and 0.25% Bromophenol Blue]. Each sample was then electrophoresed through 0.8% agarose in Tris/borate (45 mM Tris/borate, 1 mM EDTA), containing 0.5 $\mu\text{g}\cdot\text{mL}^{-1}$ of ethidium bromide, at 3 $\text{V}\cdot\text{cm}^{-1}$. Under these conditions the covalently-closed-circular form of pBS migrated fastest, followed by linear and open circle forms. The intensities of bands corresponding to supercoiled, open circle and linear DNA were quantified using densitometry. The values obtained were converted into molar concentrations by using standards of known concentration of the same DNA. The reaction rates were calculated from either the decrease in the concentration of substrate (supercoiled DNA) or the corresponding increase in the amount of product (open circle and linear DNA).

To determine initial velocity dependence, cleavage assays were carried out as described above. In a series of similar reactions containing EhLINE1-EN (2 nM) and pBS DNA, the concentration of DNA was varied from 5 to 125 nM. Zero time-points were taken by adding stop mix to a sample of pBS DNA before adding enzyme. A double-reciprocal plot of the initial velocity versus DNA concentration allowed the determination of K_m and V_{max} values. The turnover number (k_{cat}) was calculated as the ratio of V_{max} to the enzyme concentration used. The equations used to obtain the kinetic constants V_{max} , K_m and k_{cat} were as described previously [33]. Unless otherwise indicated, all enzyme activity data were the average of at least three determinations. For steady-state kinetic experiments, substrate was in molar excess over enzyme and for single turnover experiments enzyme was in molar excess over substrate.

Steady-state kinetics with 176-bp linear DNA substrate

The bottom strand of the 176-bp DNA fragment was end-labeled as described previously [9]. It was incubated with EhLINE1-EN, in the buffer mentioned above, in a volume of 100 μL at 37 $^{\circ}\text{C}$. Aliquots of 10 μL were removed at different time-points and the reactions were stopped by the addition of 25 mM EDTA. For denaturing electrophoresis on 6% polyacrylamide gels containing 7 M urea, a 2- μL aliquot of the reaction product was mixed with 8 μL of formamide gel-loading dye (95% formamide, 20 mM EDTA, 0.05% Bromophenol Blue and 0.05% xylene cyanol FF), boiled for 5 min and chilled on ice before loading. Electrophoresis was carried out as described previously [9]. The gels were dried and autoradiographed. Band intensity at site #3 was quantified using densitometry. Values obtained were converted into molar concentrations by using standards of known concentrations of the same batch of 176-bp labeled DNA.

Pre-incubation studies

All buffers and enzyme preparations were passed through Chelex-100 (Sigma). MgCl₂ solution was prepared in Chelex-100-treated Milli-Q water. Pre-incubation experiments were carried out by incubating 2 nM EhLINE1-EN with either 10 nM pBS DNA or 5 mM MgCl₂ for 10 min on ice. The reaction was initiated by adding MgCl₂ or pBS DNA, respectively. At 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 min time-intervals, aliquots were withdrawn, mixed with stop solution and assayed as mentioned above. In a control experiment, MgCl₂ and pBS DNA were pre-incubated on ice for 10 min and the reaction was initiated by the addition of EhLINE1-EN.

Size-exclusion chromatography

Gel permeation was performed on an analytical Superose-6 HR 10/30 column (GE Amersham, Uppsala, Sweden) in 50 mM Tris/HCl (pH 7.5) containing 100 mM NaCl, 10 mM MgCl₂, 2 mM β-mercaptoethanol and 10% glycerol. To determine the molecular mass of native endonuclease protein, the column was calibrated with suitable molecular mass markers, ranging from 29 to 669 kDa, and different concentrations (2.8–7.2 μM) of EhLINE1-EN were loaded. The void volume (V_o) of the column was found (using blue dextran) to be 7.71 mL, and the bed volume was 24 mL. The elution volumes of the marker proteins and of EhLINE1-EN were determined. The molecular mass of EhLINE1-EN was calculated from the plot of the log of molecular mass versus V_e/V_o (where V_e corresponds to the peak elution volume of the protein).

Fluorescence measurements

Recombinant EhLINE1-EN and EhLINE1-ENM proteins were purified from *E. coli*, as described earlier [9], and dialyzed against buffer containing 10 mM Hepes (pH 7.4), 100 mM NaCl, 2 mM β-mercaptoethanol and 1 mM EDTA (which was passed through Chelex-100). Protein samples were transferred to 5-mm quartz cuvettes, placed in a Perkin-Elmer spectrofluorimeter, LS 55, and the fluorescence emission spectra of the samples were measured at a temperature of 25 °C. The emission spectra were recorded over a wavelength of 300–440 nm with an excitation wavelength of 280 nm. A slit width of 7.5 nm was used for excitation, and a slit width of 2.5 nm was used for emission. Divalent metal ions (Ca²⁺, Mg²⁺ and Mn²⁺) were added at final concentrations of 0.5–15 mM to reactions containing EN and ENM (1 μM each) and the spectra were recorded. Each recorded spectrum was an average of at least three scans. Appropriate corrections were made for dilution of the protein sample upon addition of the metal ion. The fluorescence intensities were plotted against the total metal ion concentration and

the data were analyzed according to Stern–Volmer and modified Stern–Volmer equations [34]. The Stern–Volmer relationship is represented by $F_0/F = 1 + K_{sv} [Q]$, where F_0 and F are fluorescence intensities in the absence and presence of cofactor, respectively, K_{sv} is the Stern–Volmer constant and $[Q]$ is the quencher (divalent metal ion) concentration. In the event where there is a heterogeneous population of fluorophores, the modified Stern–Volmer relationship is used, $F_0/(F_0 - F) = 1/\{[Q] \cdot f_a \cdot KQ\} + 1/f_a$, where f_a is the fractional number of fluorophores accessible to quencher and KQ is the quenching constant. The dissociation constants were calculated graphically using the modified Stern–Volmer plot (a plot of $F_0/(F_0 - F)$ versus $1/[Q]$, where $KQ = 1/K_d$).

CD spectral analysis

CD measurements were recorded on a Jasco J810 polarimeter, using a path length of 2 mm in the far-UV region (200–250 nm), at a protein concentration of 2 μM. In the near-UV region (250–300 nm), owing to the much weaker CD effects of aromatic amino acids, it was necessary to use a path length of 1 cm. The protein concentration used for the near-UV region was 5 μM. All experiments were carried out at 25 °C in 10 mM sodium phosphate buffer (pH 7.4) containing 100 mM NaCl. To record the CD spectra in the near-UV region in the presence of divalent metal ions, 10 mM Hepes (pH 7.4) was used. The protein solutions were incubated for 10 min in a final volume of 400 μL before the spectrum was recorded. To study the effect of ligands, proteins were incubated with 5 mM Mg²⁺ for 10 min before recording the spectra. In addition, spectra were recorded in the presence of a 34-bp DNA duplex (containing the insertion site of EhSINE1), which was incubated for 10 min before recording the spectra. The molar concentration of the DNA fragment was one-tenth of that of the protein concentration. Ellipticity is reported as molar ellipticity per mean residue (in-deg-cm²-dmol⁻¹). It was calculated from the equation: $[\Theta] = \Theta_{obs} \times 100M_r/(Cl_n)$, where Θ_{obs} is the ellipticity in millidegrees, M_r is the relative molecular mass of the protein and C , l , n are the concentration (in μg-mL⁻¹), path length (in cm) and molecular number of amino acids in the protein sequence ($n = 309$) respectively. Each experimental spectrum represents an average of 15 scanning accumulations.

Acknowledgements

We thank Ruchika Sharma for critically reading the manuscript and Madhusoodanan, UK, for technical assistance in CD spectroscopy. V.P.Y is a recipient of a Senior Research Fellowship of Council of Scientific and Industrial Research, Government of India. This work was supported by grants from the Department of

Science and Technology and Department of Biotechnology, India.

References

- Malik HS, Burke WD & Eickbush TH (1999) The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* **16**, 793–805.
- Aksoy S, Williams S, Chang S & Richards FF (1990) SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINES. *Nucleic Acids Res* **18**, 785–792.
- Malik HS & Eickbush TH (2000) NeSL-1 an ancient lineage of site-specific non-LTR retrotransposon from *Caenorhabditis elegans*. *Genetics* **154**, 193–203.
- Burke WD, Malik HS, Lathe WC & Eickbush TH (1998) Are retrotransposons long-term hitchhikers? *Nature* **392**, 141–142.
- Yang J, Malik HS & Eickbush TH (1999) Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc Natl Acad Sci USA* **96**, 7847–7852.
- Zingler N, Weichenrieder O & Schumann GG (2005) APE-type non-LTR retrotransposons: determinants involved in target site recognition. *Cytogenet Genome Res* **110**, 250–268.
- Bakre AA, Rawal K, Ramaswamy R, Bhattacharya A & Bhattacharya S (2005) The LINES and SINES of *Entamoeba histolytica*: comparative analysis and genomic distribution. *Exp Parasitol* **110**, 207–213.
- Luan DD, Korman MH, Jakubczak JL & Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605.
- Mandal PK, Bagchi A, Bhattacharya A & Bhattacharya S (2004) An *Entamoeba histolytica* LINE/SINE pair inserts at common target sites cleaved by the restriction enzyme-like LINE encoded endonuclease. *Eukaryot Cell* **3**, 170–179.
- Mandal PK, Rawal K, Ramaswamy R, Bhattacharya A & Bhattacharya S (2006) Identification of insertion hot spots for non-LTR retrotransposon: computational and biochemical application to *Entamoeba histolytica*. *Nucleic Acids Res* **34**, 5752–63.
- Pingoud A, Alves J & Geiger R (1993) Restriction Enzymes. In *Methods in Molecular Biology, Vol 16: Enzymes of Molecular Biology* (Burrell MM, ed), pp. 107–200. Humana Press Inc, Totowa, NJ.r5.
- Kaczorowski T, Skowron P & Podhajska AJ (1989) Purification and characterization of FokI restriction endonuclease. *Gene* **80**, 209–216.
- Skowron P, Kaczorowski T, Tucholski J & Podhajska AJ (1993) Atypical DNA-binding properties of class-IIS restriction endonucleases: evidence for recognition of the cognate sequence by a FokI monomer. *Gene* **125**, 1–10.
- Wah DA, Bitinaite J, Schildkraut I & Aggarwal AK (1998) Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci USA* **95**, 10564–10569.
- Bitinaite J, Wah DA, Aggarwal AK & Schildkraut I (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci USA* **95**, 10570–10575.
- Bheemanaik S, Reddy YV & Rao DN (2006) Structure, function and mechanism of exocyclic DNA methyltransferases. *Biochem J* **399**, 177–190.
- Halford SE, Bilcock DT, Stanford NP, Williams SA, Milsom SE, Gormley NA, Watson MA, Bath AJ, Embleton ML, Gowers DM *et al.* (1999) Restriction endonuclease reactions requiring two recognition sites. *Biochem Soc Trans* **27**, 696–699.
- Sistla S & Rao DN (2004) S-adenosyl-L-methionine-dependent restriction enzymes. *Crit Rev Biochem Mol Biol* **39**, 1–19.
- Pingoud A & Jeltsch A (1997) Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur J Biochem* **246**, 1–22.
- Sapranaukas R, Sasnauskas G, Lagunavicius A, Vilkaitis G, Lubys A & Siksnys V (2000) Noval subtype of type IIs restriction enzymes. *J Biol Chem* **275**, 30878–30885.
- Pingoud A & Jeltsch A (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res* **29**, 3705–3727.
- Nastri HG, Evans PD, Walker IH & Riggs PD (1997) Catalytic and DNA binding properties of PvuII restriction endonuclease mutants. *J Biol Chem* **272**, 25761–25767.
- Grabowski G, Maass G & Alves J (1996) Asp-59 is not important for the catalytic activity of the restriction endonuclease EcoRI. *FEBS Lett* **381**, 106–110.
- Horton NC, Newberry KJ & Perona JJ (1998) Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. *Proc Natl Acad Sci USA* **95**, 13489–13494.
- Grabowski G, Jeltsch A, Wolfes H, Maass G & Alves J (1995) Site-directed mutagenesis in the catalytic center of the restriction endonuclease EcoRI. *Gene* **157**, 113–118.
- Wolfes H, Alves J, Fliess A, Geiger R & Pingoud A (1986) Site directed mutagenesis experiments suggests that Glu 111, Glu 144 and Arg 145 are essential for endonucleolytic activity of EcoRI. *Nucleic Acids Res* **14**, 9063–9081.
- Dupureur CM & Conlan LH (2000) A catalytically deficient active site variant of PvuII endonuclease binds Mg(II) ions. *Biochemistry* **39**, 10921–10927.
- Kahn PC (1979) The interpretation of near-ultraviolet circular dichroism. *Methods Enzymol* **61**, 339–378.
- Provencher SW & Gloeckner J (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33–37.

- 30 Unneberg P, Merelo JJ, Chacon P & Moran F (2001) SOMCD: method for evaluating protein secondary structure from UV circular dichroism spectra. *Proteins* **42**, 460–470.
- 31 Eickbush TH (2002) R2 and related site-specific non-long terminal repeat retrotransposons. In *Mobile DNA II* (Craig NL, Craigie R, Gellert M & Lambowitz AM, ed), pp. 828–829. American Society for Microbiology, Washington, DC.
- 32 Bradford MM (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein dye binding. *Anal Biochem* **72**, 248–254.
- 33 Segel IH (1975) *Biochemical Calculations*, 2nd edn. pp. 208, John Wiley & Sons, Inc., New York, NY.
- 34 Lehrer SS (1971) Solute perturbation of protein fluorescence: the quenching of the tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. *Biochemistry* **17**, 3254–3263.