Published online 30 January 2023

Nucleic Acids Research, 2023, Vol. 51, No. 3 ***1189–1207***
https://doi.org/10.1093/nar/gkad017

The enterohemorrhagic *Escherichia coli* insertion sequence-excision enhancer protein is a DNA polymerase with microhomology-mediated end-joining activity

Patricia A. Calvo[1], Víctor Mateo-Cáceres[2], Silvia Díaz-Arco[1], Modesto Redrejo-Rodríguez[2] and Miguel de Vega[1,*]

[1]Centro de Biología Molecular Severo Ochoa (Consejo Superior de Investigaciones Científicas-Universidad Autónoma de Madrid), Nicolás Cabrera 1, Madrid 28049, Spain and [2]Department of Biochemistry, School of Medicine, Universidad Autónoma de Madrid and Instituto de Investigaciones Biomédicas Alberto Sols (Universidad Autónoma de Madrid-Consejo Superior de Investigaciones Científicas), Madrid, Spain

Received October 31, 2022; Revised December 30, 2022; Editorial Decision January 03, 2023; Accepted January 05, 2023

ABSTRACT

Bacterial genomes contain an abundance of transposable insertion sequence (IS) elements that are essential for genome evolution and fitness. Among them, IS*629* is present in most strains of enterohemorrhagic *Escherichia coli* O157 and accounts for many polymorphisms associated with gene inactivation and/or genomic deletions. The excision of IS*629* from the genome is promoted by IS-excision enhancer (IEE) protein. Despite IEE has been identified in the most pathogenic serotypes of *E. coli*, its biochemical features that could explain its role in IS excision are not yet understood. We show that IEE is present in >30% of all available *E. coli* genome assemblies, and is highly conserved and very abundant within enterohemorrhagic, enteropathogenic and enterotoxigenic genomes. *In vitro* analysis of the recombinant protein from *E. coli* O157:H7 revealed the presence of a Mn^{2+}-dependent error-prone DNA polymerase activity in its N-terminal archaeo-eukaryotic primase (AEP) domain able to promote dislocations of the primer and template strands. Importantly, IEE could efficiently perform *in vitro* an end-joining reaction of 3'-single-strand DNA overhangs with ≥4 bp of homology requiring both the N-terminal AEP and C-terminal helicase domains. The proposed role for IEE in the novel IS excision mechanism is discussed.

INTRODUCTION

Insertion sequences (ISs) are the simplest transposable elements present in most bacterial genomes where they play relevant roles in genome evolution and fitness. The transposition and proliferation of ISs can result in diverse genomic changes, including deletions, inversions, duplications, insertional gene inactivation, pseudogene formation, and modification of gene expression, leading to a range of different phenotypes including increased antimicrobial resistance or upregulation of virulence genes in clinically-significant human pathogens (1–5).

More than 4000 types of IS elements have been identified and classified into 29 families in bacteria [(4,6) (https://tnpedia.fcav.unesp.br/index.php/Main_Page)]. Structurally, ISs are short (0.7–2.5 kb) DNA fragments containing one or sometimes two open reading frames encoding a transposase flanked by terminal inverted repeat (TIR) sequences. The transposition of IS elements requires the specific recognition of the TIR sequences by the cognate transposase, followed by strand cleavage at the IS boundaries and transfer of the IS into the target site *via* either a copy-and-paste mechanism that leaves a copy of the IS in the donor DNA, or a cut-and-paste mechanism that removes the IS from the donor genome (4). IS-mediated bacterial genome diversification has been extensively studied in the enterohemorrhagic *Escherichia coli* (EHEC) strain O157:H7, which produces potent cytotoxins such as Shiga toxins (hence, commonly referred to as Shiga toxin-producing *E. coli* or STEC). EHEC O157:H7 is a major cause of diarrhea, hemorrhagic colitis, and hemolytic-uremic syndrome, and it is frequently associated with severe diseases and outbreaks (7). Sequence analyses of O157 strains has enabled the identification of a large number of IS elements in their genomes (116 belonging to 25 families), which are largely responsible for the great diversification (8). Among them, the IS*3* family member IS*629* is the most abundant (23 copies) and is responsible for many polymorphisms associated with gene inactivation and/or genomic

*To whom correspondence should be addressed. Tel: +34 911964717; Fax: +34 911964420; Email: mdevega@cbm.csic.es

deletions (9). Detailed analysis of the transposition mechanism of IS*629* enabled the identification of a new protein, termed IS-excision enhancer (IEE), which promotes excision of IS*629* and other members of the IS*3*, IS*1* and IS*30* families in a transposase-dependent manner (10,11). The simultaneous presence of both IEE and IS*629* elements in almost all *E. coli* STEC strains from seropathotypes A and B, responsible for high pathogenicity and human disease, and the absence of at least one of them in the less pathogenic seropathotypes C and D, has led to the suggestion that their combined presence drives evolution towards increased virulence (12), which makes the study of this transposition system clinically important. Additionally, it has been recently reported how the combined action of both IEE and IS*3* can form new operons, highlighting how coevolution with mobile elements might shape the organization of prokaryotic genomes and gene regulation (13). Remarkably, beyond the aforementioned EHEC/STEC and enterotoxigenic (ETEC) *E. coli* strains, IEE homologs have also been identified in a broad range of bacterial species, suggesting their spread by horizontal gene transfer (10).

Whereas the activities and modes of action of the different types of known transposases have been widely studied [reviewed in (4)], those of IEE remain to be elucidated. Sequence analysis of IEE showed that the protein is composed of an N-terminal archaeo-eukaryotic primase (AEP) domain fused to a C-terminal DEAH-helicase domain [(10), see also Figure 1]. The AEP superfamily is distributed across all domains of life and is classified into 13 major families, most of them arranged into the AEP proper clade, the nucleo-cytoplasmic large DNA virus (NCLDV)-herpesvirus clade, and the PrimPol clade (14). AEPs are a varied group of enzymes with diverse catalytic features that have evolved to play several specialized roles during: (i) DNA replication, as the eukaryotic replicative primase (AEP proper clade) that synthesizes an RNA primer that is extended by Polα to provide a substrate for DNA synthesis by Polδ and Polε; (ii) DNA repair, as bacterial ligase D (LigD), a protein composed of an AEP domain belonging to the AEP proper clade fused to phosphoesterase and ligase domains that enable the enzyme to repair double-strand breaks (DSBs) through nonhomologous end-joining (NHEJ); (iii) damage tolerance, as human PrimPol that contains an AEP domain belonging to the NCLDV-herpes virus clade fused to a C-terminal zinc finger domain that can bypass DNA lesions and participate in the restart of stalled replication forks by either making use of its translesion synthesis (TLS) ability or by repriming downstream of damage [reviewed in (15,16)].

The fact that IEE and IS*629* are present in the most pathogenic serotypes of *E. coli*, together with both the absence of a known system for joining the DNA ends in the genome that result from IS excision, and the presence of an AEP domain that in other enzymes is involved in different DNA transactions (see above), prompted us to study the biochemical properties of IEE from *E. coli* O157:H7, with the aim of establishing the potential role of this enzyme in the novel transposition mechanism of the IS elements.

MATERIALS AND METHODS

IS-excision enhancer screening in GenBank database assemblies

To screen the 155 384 *E. coli* assemblies in the GenBank database (accessed on 24 February 2022), we devised a custom Python pipeline. The script downloads all the assemblies in *dehydrated* format (links only) using the NCBI Datasets tool (https://www.ncbi.nlm.nih.gov/datasets). Then, assembly *hydration* (actual download) and IEE search occur concurrently in cycles that process batches of 100 genomes. The search task entails the prediction of coding sequences with Prodigal (17) (Close ends '-c' option on, remaining parameters by default) and proteome search using BLASTP (18) with the IEE protein sequence (accession at NR database NP_309332.1) as the query (*E*-value < 1E−10, coverage > 70%). For the IEE-positive hits, the genome assembly, the IEE-containing contig and the identified IEE sequences were saved for downstream analyses (see Supplementary Table S1). The significant co-occurrence of IEE and IS*629* transposase (UniProt ID: O82918) was assessed using BLASTP against the IEE-harboring assemblies with an E-value cutoff of 1E−15.

IEE proteins were subsequently clustered at 95% identity with CD-Hit (19) (Supplementary Table S2 and Supplementary Text S1) and the resultant 20 representative sequences were aligned with Mafft (20) (–localpair –maxiterate 1000). The alignment was then displayed with the ggmsa package for R (21).

The IEE-harboring genome assemblies were characterized *in silico* for serotype, multilocus sequence type (MLST) and phylogenetic groups using ECTyper (22), MLST (https://github.com/tseemann/mlst), and EzClermont (23) tools, respectively. Data were aggregated and plotted using the R packages data.table and ggplot2, both available at CRAN (https://CRAN.R-project.org).

Reagents and oligonucleotides

Unlabeled ultrapure dNTPs and NTPs were purchased from GE Healthcare (Chicago,USA, catalogs: dNTPs #28-4065-52; UTP #27-2086-01, CTP #27–2066-01; ATP #27-2056-01; GTP #10784737), [γ^{32}P]ATP from Perkin Elmer Life Sciences (Waltham, MA, USA; catalog #NEG502A100UC), and oligonucleotides from Integrated DNA Technologies (Leuven, Belgium) (sequences are shown in the figures). When indicated, oligonucleotides were radiolabeled at the 5' end using [γ^{32}P]ATP (3000 Ci/mmol) and T4 polynucleotide kinase (New England Biolabs, Ipswich, MA, USA; catalog #M0201S). M13mp18 (+) strand was purchased from GE Healthcare (catalog # 27-1546-01). To obtain the hybrid double-stranded (ds)DNA molecules, the indicated oligonucleotides were annealed in the presence of 60 mM Tris–HCl (pH 7.5) and 0.2 M NaCl at 80°C for 5 min before slowly cooling to room temperature overnight. *E. coli* DNA Ligase A and NAD^+ were from New England Biolabs (catalogs #M0205S and #M0309, respectively). Q5 site-directed mutagenesis kit was purchased from New England Biolabs (catalog #E0554S). Proteinase K was purchased from Merck (Darmstadt, Germany, catalog #03115844001).

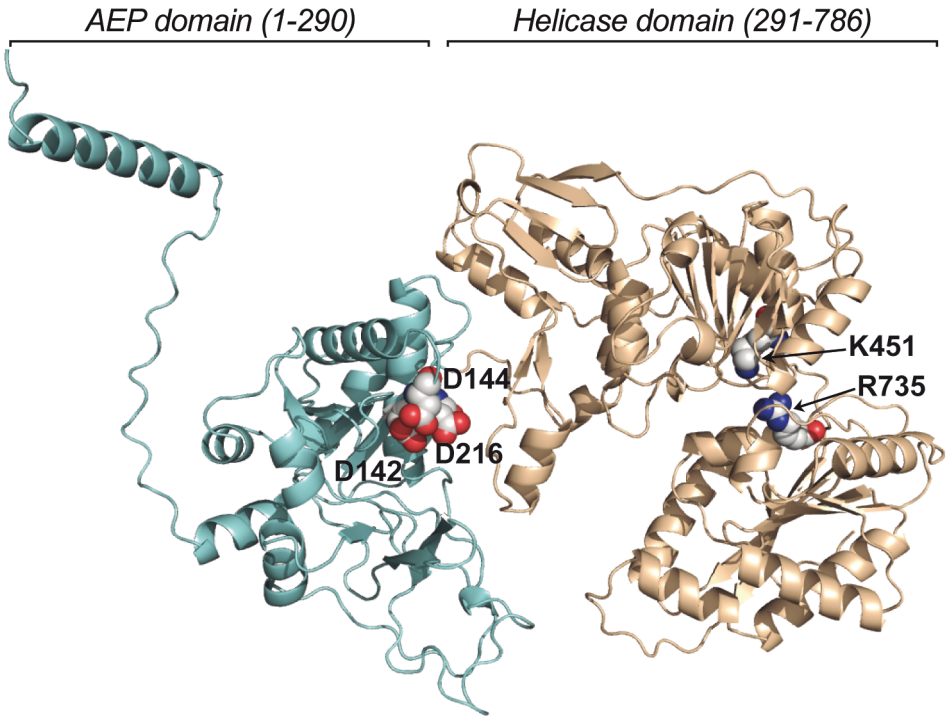

Figure 1. Structural model of IEE from *Escherichia coli* O157 reference EDL933 strain. The N-terminal AEP domain is colored in cyan and shows the predicted catalytic aspartates responsible for the polymerization reaction (represented as spheres). The C-terminal helicase domain is colored in brown. The conserved lysine that would bind the ATP molecule and the arginine finger are represented as spheres. The predicted structure of IEE from *E. coli* O157:H7 (26) was obtained through the Uniprot web site (https://www.uniprot.org/) from the Alphafold Database ID A0A0H3JF09 (30), and used under CC BY4.0. The original colors were changed as described above. Figure was generated using The Open-Source Pymol Molecular Graphics System, v. 2.5.0, Schrödinger, LLC (Open-Source PyMOL is Copyright (C) Schrodinger, LLC.)

Phi29 DNA polymerase and *B. subtilis* DNA polymerase X were overexpressed and purified as described (24,25).

Expression and purification of recombinant IS-excision enhancer

Recombinant IEE: The open reading frame encoding IEE from *E. coli* O157:H7 reference strain EDL633 [GeneBank Accession number AAG55274.1; (26)] was synthesized by the Cusabio Technology LLC gene synthesis service, and was cloned between the NdeI and BamHI restriction sites in the pET28a(+) vector to express a recombinant protein fused to an N-terminal $(His)_6$-tag for purification on Ni^{+2}-affinity resin. *E. coli* Shuffle T7 cells, engineered by integrating the bacteriophage T7 RNA polymerase gene into *lacZ* to allow expression of genes under the regulation of the T7 promoter (27), were transformed with the resulting recombinant expression plasmid, named pET28-IEE. Transformed cells were grown overnight at 37°C in LB medium with kanamycin (50 μg/ml). Subsequently, cells were diluted into the same medium in the presence of 0.1 M sorbitol (Sigma) and incubated at 37°C until the OD_{600} reached 0.4. Then, IPTG (Sigma) was added to a concentration of 0.4 mM and incubation was continued for 4 h at 30°C. Cells were then collected by centrifugation for 10 min at 6143 × g at 4°C. Cells were freeze–thawed and ground with alumina at 4°C; the slurry was resuspended in Buffer A (50 mM Tris–HCl pH 7.5, 0.7 M NaCl, 7 mM β-mercaptoethanol, 25% glycerol, 1 mM PMSF) and centrifuged for 5 min at 650 × g at 4°C to remove alumina and intact cells. The supernatant was further centrifuged for 20 min at 23 430 × g at 4°C to separate insoluble proteins from the soluble extract, which was incubated for 30 min at 4°C in the presence of 25 U/ml Benzonase (Santa Cruz Biotechnology) and 2 mM $MgCl_2$ to digest the nucleic acids present in the sample. The soluble extract was then loaded onto a Ni^{2+}-NTA column (Qiagen) pre-equilibrated with Buffer A containing 5 mM imidazole. The affinity column was step-eluted with 50–300 mM imidazole in Buffer A (containing 0.3 M NaCl). The polypeptide composition of the column fractions was monitored by SDS-PAGE. The His-tagged IEE was recovered in the 100–300 mM imidazole eluate fractions, which were then dialyzed against Buffer A (0.1 M NaCl) for 4 h at 4°C and further loaded onto a phosphocellulose column pre-equilibrated with Buffer A (0.1 M NaCl). The affinity column was eluted stepwise with Buffer A (0.2–1 M NaCl). The polypeptide composition of the column fractions was monitored by SDS-PAGE. The His-tagged IEE was recovered in the 0.5–1 M NaCl eluate fractions, which were then loaded onto a second Ni^{2+}-NTA column pre-equilibrated with Buffer A (0.75 M NaCl and 30 mM imidazole). The IEE protein was eluted with Buffer A (0.75 M NaCl and 400 mM imidazole). The eluate fraction was dialyzed against a buffer containing 50 mM Tris–HCl, pH 7.5, 1 mM EDTA, 7 mM β-mercaptoethanol, 0.3 M NaCl and 50% glycerol and stored at −20°C.

Site-directed mutagenesis

The IEE catalytic-deficient mutants D142A/D144A and K451A were generated with the Q5 site-directed mutagenesis kit using plasmid pET28-IEE (see above) as a template for mutagenesis. Confirmation of the DNA sequence and the absence of additional mutations was performed by sequencing the entire gene. IEE mutant proteins were purified from soluble extracts of IPTG-induced *E. coli* Shuffle T7 cells by Ni-NTA and phosphocellulose chromatography as described for the wild-type enzyme.

Recombinant IEE polymerization domain (AEP-Dom): For the independent expression of the AEP domain (residues 1–288), pET28-IEE was used as template to change codon 289 (CGG) into the stop codon TAA using the Q5 site-directed mutagenesis kit. Confirmation of the DNA sequence and the absence of additional mutations was performed by sequencing the entire gene. *E. coli* BL21 (DE3) cells were transformed with the resulting plasmid and induction of protein expression and preparation of soluble bacterial lysates were performed as described for full-length IEE. The supernatant containing the AEP-Dom was loaded onto a Ni-NTA column pre-equilibrated with Buffer A (0.7 M NaCl, 10 mM imidazole, 0.05% Tween). The affinity column was step-eluted with 75, 100, 150, 200 and 300 imidazole; the polypeptide composition of the column fractions was monitored by SDS-PAGE. The His-tagged AEP-Dom was recovered in the 150–300 mM imidazole-eluate fractions and further dialyzed against a buffer containing 50 mM Tris–HCl, pH 7.5, 0.3 M NaCl, 1 mM EDTA, 7 mM β-mercaptoethanol, 0.05% Tween-20 and 50% glycerol and stored at −20°C.

DNA polymerization assays on defined DNA molecules

Polymerase activity was assayed by NTP- or dNTP-dependent extension of a 15-mer, labeled at its 5'-end with either T4 PNK and [γ^{32}P]ATP, or Cy5, as indicated, and hybridized to the specified template (see oligonucleotide sequences in figures). The incubation mixture (12.5 μl) contained 50 mM Tris–HCl pH 7.5, the indicated concentration of either $MnCl_2$ or $MgCl_2$, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/ml BSA, either 10 nM of the Cy5-labeled hybrid or 1 nM of the ^{32}P labeled substrate, and the indicated concentration of IEE and nucleotides. After incubation for the indicated times at 37°C the reactions were stopped by adding EDTA to 10 mM. Samples were analyzed by 7 M urea–20% PAGE and visualized using a Typhoon 9410 scanner (GE Healthcare).

DNA helicase assays

Assays were performed essentially as described (28). The 5'-^{32}P-labeled oligonucleotide (T)15-GTTTTCCCAGTCACGAC-(T)15 (29) was hybridized to M13mp18 (+) strand to create a partial duplex (middle 17 nucleotides) with 3' and 5' poly(dT)$_{15}$ tails. The reaction mixture (12.5 μl) contained 50 mM Tris–HCl pH 7.5, either 1 mM $MnCl_2$ or 5 mM $MgCl_2$, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/ml BSA, 1 nM of the indicated ^{32}P-labeled substrate, and the indicated concentration of the specified protein. The reaction was started by adding simultaneously 2 mM ATP and 100 nM of non-labeled oligonucleotide to prevent reannealing of the displaced oligonucleotide to the M13 genome molecule. After incubation for 30 min at 37°C, the reaction was stopped by adding 50 mM EDTA, 0.6% SDS, 10% glycerol and 2 mg/ml Proteinase K. After an additional incubation for 15 min at 37°C, the reaction products were resolved in a 10% non-denaturing polyacrylamide gel (0.1% SDS) and visualized using a Typhoon 9410 scanner.

ATPase assay

The incubation mixture contained, in 10 μl, 50 mM Tris–HCl pH 7.5, the indicated concentration of either $MnCl_2$ or $MgCl_2$, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/ml BSA, 1 mM ATP, 66 nM [γ-^{32}P]ATP (0.2 μCi), 2.5 mM $MgCl_2$, 100 nM of either wild-type IEE or mutant K451A and, when indicated, 50 ng/μl of either plasmid pUC18 (dsDNA) or single-stranded (ss)DNA from bacteriophage M13. After incubation for 20 min at 30°C, the reactions were stopped by adding 30 mM EDTA, and 1 μl of the resulting solution was spotted onto PEI-cellulose (Polygram Cel 300 PEI/UV254) for thin layer chromatography. Chromatograms were run in 0.15 M lithium formate, conditions where Pi migrates and ATP remains at the origin. The reaction products were visualized by autoradiography.

MMEJ assay

The incubation mixture contained, in a total of 12.5 μl, 50 mM Tris-HCl pH 7.5, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/ml BSA, 10 μM of the indicated dNTP, 1 mM $MnCl_2$, 25 nM of the indicated 5'-^{32}P-radiolabeled DNA and 40 nM of either wild-type IEE, AEP-Dom or mutant K451A. When indicated, *E. coli* LigA (1 U), 500 μM NAD^+ and 5 mM $MgCl_2$ were also added. After incubation for 15 min at 37°C reaction products were resolved by 7 M urea-20% PAGE and visualized using a Typhoon 9410 scanner. Quantitation of the ligated products was performed by calculating the ratio ligated product/(ligated product + 19mer elongation product) of two independent experiments.

Determination of the templating nucleotide that directs the insertion of the incoming nucleotide during the MMEJ reaction was performed essentially as described above, incubating the protein with 1 nM of the 5'-^{32}P-radiolabeled DNA and 20 nM of the indicated non-labeled hybrid. Reactions were started by adding 1 mM $MnCl_2$ and were incubated for 15 min at 37°C. Reactions were stopped by adding 10 mM EDTA and analyzed as described above.

Statistical analysis

All data were expressed as mean ± SEM from two or three independent experiments performed in a parallel manner. When indicated, comparisons between two groups were analyzed using two-tailed Student's *t*-tests. P values <0.05 were considered statistically significant.

RESULTS AND DISCUSSION

IS-excision enhancer is a highly conserved and distributed protein in *E. coli* strains composed of an AEP and a helicase domain

Sequence analysis (14) and AlphaFold full-length predicted structure analysis [UniProt ID A0A0H3JF09 (30) (31)] of the 786 amino-acid IEE from EHEC O157:H7 (Figure 1) indicated that it is composed of an N-terminal AEP domain belonging to the Z1568-like family (residues 1–290) fused to a C-terminal domain homologous to the superfamily II (SF2) of DNA/RNA helicases (residues 291–786).

Amino acid sequence alignment of the polymerization domain of several members of the AEP superfamily is presented in Figure 2A. As observed, within its primary structure IEE has sequences corresponding to the conserved motifs A (hDhD/E, where h is a hydrophobic residue) and C (hD/E) that contain the triad of carboxylate groups that would coordinate the divalent ions A (potentially coordinated by IEE residues D142, D144 and D216) and B (coordinated by residues D142 and D144) responsible for catalysis of the nucleotide incorporation (see also the structural modeling of the catalytic active site of the IEE AEP domain in Figure 2B). Additionally, IEE residue H178 would be homologous to the conserved histidine present in motif B (sxH/Q, 's' stands for a small residue), which interacts with the non-bridging oxygen atoms of the α and β phosphate groups of the incoming nucleotide in AEPs (32–36). Other conserved residues were aligned (see also alignment in Supplementary Figure S1) and modeled (Figure 2B), including S171, Q175 and G176, which would correspond to residues S176, R179 and G180 of the AEP domain (PolDom) of *Mycobacterium smegmatis* (Msm) PolD1 [PrimPolC, (37)] and to residues S172, K175 and G176 of the PolDom of *M. tuberculosis* (MtuLigD) that interact with the phosphate tail of the incoming UTP via hydrogen bonds (37,38). The crystallographic structures of the PolDom of *Pseudomonas aeruginosa* LigD (PaeLigD) with a Mn^{2+}-ATP bound at the active site (39), as well as the ternary complexes of the PolDom of MtuLigD and PrimPolC with an incoming UTP (37,38), showed that the ribose 2'-OH group of the incoming ribonucleotide is hydrogen bonded to the Nδ moiety of a conserved histidine (PaeLigD H651, MtLigD H111; PrimPolC H122), with this interaction responsible for the preferential insertion of ribonucleotides exhibited by these AEPs (38–40). As shown in Figure 2B and in Supplementary Figures S1 and S2, the above-mentioned conserved histidine is substituted by a tyrosine in IEE (Y128).

The C-terminal domain of IEE (residues 291–786) belongs to the superfamily II (SF2) of DNA/RNA helicases (10), containing the seven conserved motifs present in members of this superfamily (see Figure 2A). Accordingly, as in most SF2 helicases, the C-terminal domain of IEE would be constituted by two core RecA-like folds (Figures 1 and 2C) characterized by a central β-sheet flanked by α-helices (41), including residues from motif I such as the highly conserved lysine (IEE K451) essential for ATP binding (42–44), motif II (DExH/D) involved in binding and hydrolysis of the ATP (equivalent to the Walker A and B motifs of many ATPases), and motif VI, such as the arginine finger (IEE Arg735 in Figure 2A and C) involved in energy coupling (45).

As mentioned above, IEE has been reported to be very common in O157 EHEC/STEC and ETEC strains (Kusumoto *et al.* 2011, 2014). As the number of *E. coli* genome sequences continues to grow, we wanted to comprehensively update and analyze the frequency of IEE among diverse *E. coli* groups by screening for its presence in all *E. coli* genome assemblies available in GenBank (see Materials and Methods for details). We found that IEE is strikingly abundant, being detected in >33% of the *E. coli* genomes (51 821 IEEs in 51 265 positive strains out of 155 384 screened assemblies). The large majority of the identified proteins (98.5%) have a similar length and domain composition to the O157:H7 IEE. A small proportion of the sequences are <750 amino acids, with small deletions in the N-terminal AEP (1%) or C-terminal helicase (0.5%) domains, which, however, do not affect the catalytic motifs described above in the representative sequences (Supplementary Figure S3). Remarkably, notwithstanding the minority of truncated proteins, the overall conservation of the IEE amino acid sequence is very high, with 98.8% of the proteins having over 95% identity with O157:H7 IEE.

As anticipated, IEE was present mostly in *E. coli* O157:H7 (37%), but was also detected in other previously reported serotypes (12) including O26:H11 (13.8%), O103:H2 (11.6%), O111:H8 (8.1%) and O121:H19 (2.7%) (Figure 3A). We also found several new IEE-harboring serovars within enteropathogenic *E. coli* (EPEC) and STEC/EHEC pathotypes, including O123/O186 (1.6%), O71 (1.2%), and O146 (1%) serogroups, among others <1%. Regarding the multilocus sequence typing (MLST) of the genomes, IEE could be detected in genomes from up to 608 different sequence types (Supplementary Table S1), although the main groups correlated with the more common serotypes (Figure 3A), in agreement with the clonality of most serotypes. The vast majority of the IEE-containing genomes belonged to the B1 and E phylogroups (Figure 3B), whereas the most common human commensal strains from the phylogenetic group A (46) accounted for only 5.4% of the IEE-harboring strains. As expected, the high prevalence of the E group roughly corresponds to strains from the O157:H7 serovar (47). On the other hand, the abundance of B1 strains (51.5%) would correspond to STEC/EHEC strains from diverse serotypes or MLST clades (48). What was remarkable was the reduced number of IEE detected in other major groups of concern, such as extraintestinal pathogenic strains from the B2 and D phylogenetic groups (49).

As expected, analysis of the IEE-containing assemblies also yielded a significant co-occurrence (94.75%) with IS*629* or similar transposases, which underlines the functional association between IS*629* and possibly other related ISs with the IEE protein.

In conclusion, although initially described as common in *E. coli* O157:H7, our results show that IEE protein is highly conserved and shows a patchy distribution among EHEC, EPEC and ETEC/STEC genomes, suggesting not only promiscuous horizontal transfer across diverse *E. coli* serovars and phylogroups, but also a significant biological role in enhancing adaptability, as was recently proposed

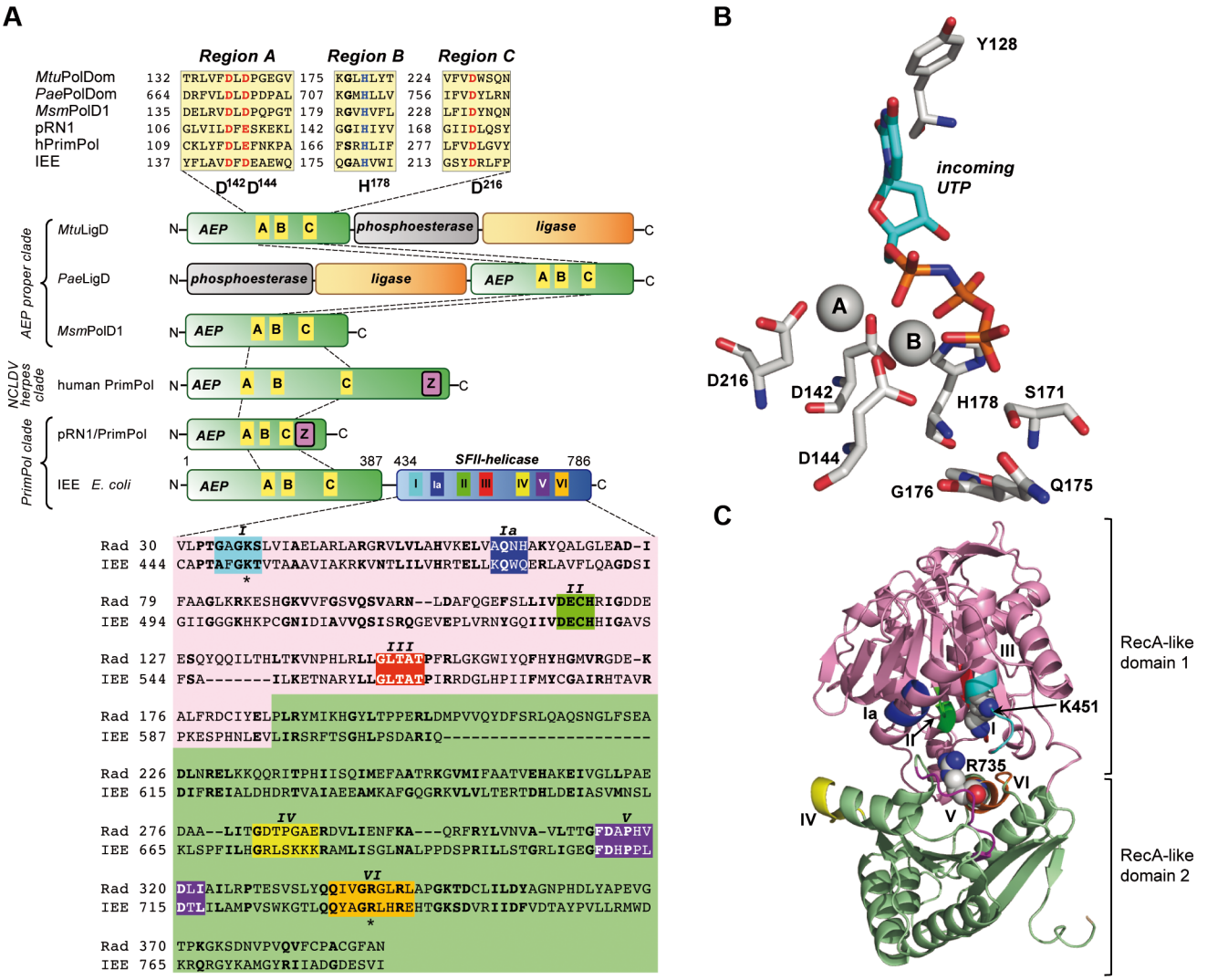

Figure 2. Catalytic active sites of IEE. (**A**) *Modular organization of several AEP domain-containing enzymes.* The conserved AEP domain (green bar) contains the three conserved regions A, B and C, which form the polymerization active site (in the alignment metal ligands are indicated in red letters and the conserved His that interacts with the incoming nucleotide is in blue). Members such as bacterial NHEJ Ligases D from *M. tuberculosis* and *P. aeruginosa* (*Mtu*- and *Pae*LigD) contain additional phosphoesterase (in grey) and ligase (in orange) domains that process 3'-ends and seal the resulting nicks after synapsis during the end-joining reaction. Some members such as the PrimPol of plasmid pRN1 from *Sulfolobus islandicus* and human PrimPol contain a Zn-finger domain (in magenta), which is essential to stabilize the initiating nucleotide during the DNA priming reaction (84). The N-terminal AEP domain of IEE is fused to a C-terminal helicase-like domain containing the seven conserved motifs present in superfamily II of these proteins. The figure also shows an alignment of the C-terminal helicase domain of IEE with the superfamily II helicase member RadD from *E. coli* (85). *Mtu*PolDom, Polymerization domain (AEP domain) of *Mtu*LigD; *Pae*PolDom, Polymerization domain of *Pae*LigD; *Msm*PolD, *M. smegmatis* PolD1 [PrimPolC (37)]; pRN1, plasmid pRN1 from *S. islandicus*; hPrimPol, human PrimPol. (**B**) *Model of the active site geometry of the AEP domain of IEE.* The model was obtained by superposition of the putative catalytic residues of the AEP domain of IEE with the crystallographic structure of the ternary complex of PrimPolC [PDB 6SA0 (37)]. (C) *Detailed view of the predicted structure of the C-terminal helicase domain of IEE.* The predicted ATP ligand residue Lys451 and the arginine finger (Arg735) are shown as spheres. The location in the three-dimensional structure of the conserved motifs is indicated following the same color code as in the alignment shown in (A). Figures (B) and (C) were generated using The Open-Source Pymol Molecular Graphics System, v. 2.5.0, Schrödinger, LLC (Open-Source PyMOL is Copyright (C) Schrodinger, LLC.)

(11,13). Thus, IEE could be part of a generalized mechanism that supports adaptability to changing environmental conditions.

IS-excision enhancer has DNA polymerization activity

To study its biochemical features, we cloned the entire open reading frame of IEE from the *E. coli* O157:H7 EDL633 reference strain into the pET28(a)+ expression vector, which carries an N-terminal $(His)_6$-tag. The recombinant protein was expressed in *E. coli* Shuffle (DE3) cells and purified (Materials and Methods). Based on amino acid sequence similarity analysis (see above), the N-terminal domain of IEE was proposed to have nucleotidyl transferase activity (14). Thus, to examine the presence of polymerization activity, we tested whether the enzyme could behave as a DNA-dependent DNA polymerase using a short DNA hybrid with a primer-template structure (see scheme in Fig-

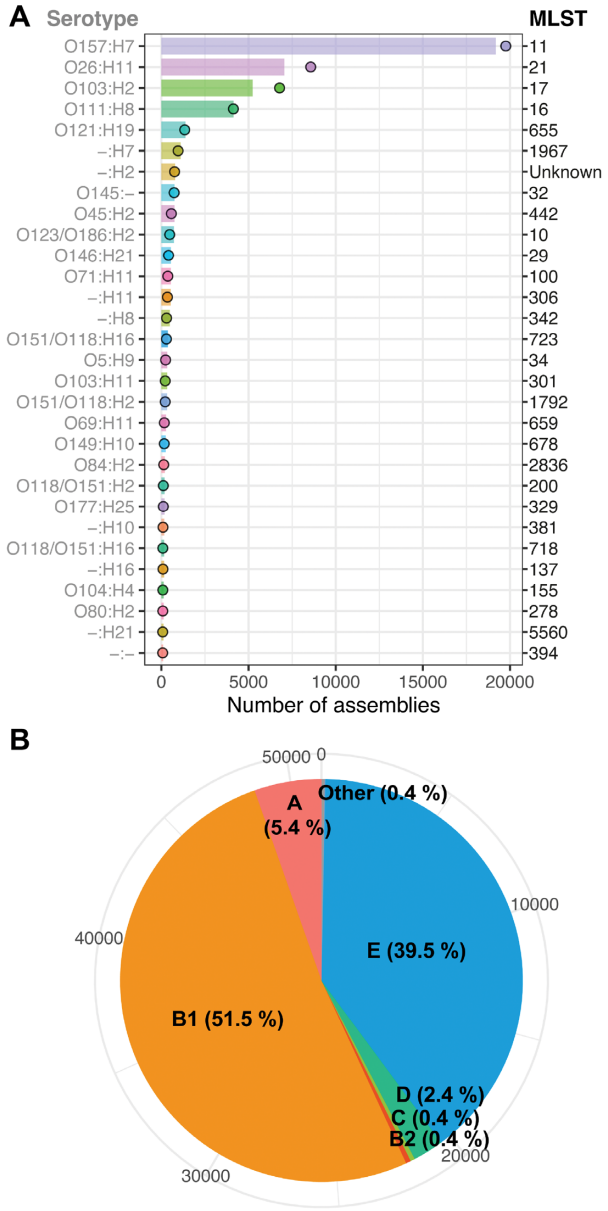

Figure 3. High abundance of IEE across diverse strains of *E. coli*. (**A**) Correlation between the most common predicted serotypes (left axis) and multilocus sequence typing (MLST) groups (right axis) among the IEE positive assemblies in GenBank. The serotypes are represented as colored bars and MLST groups as points. (**B**) Phylogenetic groups of the IEE-harboring assemblies. The percentage of each Clermont's classification phylogroup is indicated in the pie sections and the scale in the outer ring represents the actual number of assemblies. 'Other' stands for assemblies from minority groups, cryptic phylogroup, and labeled as non-*E. coli* genomes. See Materials and Methods for details.

ure 4A). Results showed that IEE could extend the DNA primer and produce elongation intermediates characteristic of distributive DNA synthesis, being 78- and 33-fold more efficient in inserting dNTPs over NTPs in the presence of 0.1 and 1 mM Mn^{2+}, respectively (Figure 4A). Analysis of mutations introduced at the predicted catalytic residues D142 and D144 (see Figure 2A and C) allowed us to infer that the DNA polymerase activity was intrinsic to IEE, and not due to a contaminant polymerase from the expression system. As mentioned above, the catalytic active site of extensively characterized AEP-containing proteins such as PaeLigD, MtuLigD and PrimPolC, includes a histidine residue (PaeLigD H651, MtuLigD H111 and PrimPolC H122) that contacts the ribose 2'-OH group of the incoming ribonucleotide [(37–39), see also the sequence alignment in Supplementary Figure S1]. Site-directed mutagenesis at the corresponding PaeLigD H651 allowed the authors to conclude that this residue is the main structural element responsible for the discrimination against dNTP insertion exhibited by these enzymes (40). Other AEP members that proficiently discriminate against ribonucleotides, such as *Pyrococcus furiosus* p41 (50) and human PrimPol (51), have a tyrosine residue in its place (Y72 and Y100, respectively) (52). A recent biochemical study conducted with the PrimPol mutant Y100H showed that this change stimulated the insertion of NMPs, which allowed authors to propose that the residue located at the equivalent position would dictate the nucleotide sugar use, histidine and tyrosine favoring the incorporation of NTP and dNTPs, respectively (52). Notably, both, multiple sequence alignments (Supplementary Figure S1) and structural modeling (Figure 2B and Supplementary Figure S2) revealed that IEE has a highly conserved tyrosine residue (Y128) in the corresponding position, likely accounting for its preferential usage of dNTPs and supporting the above-mentioned hypothesis.

As it can be observed in Figure 4A, the efficiency of the polymerization activity of IEE in the presence of 1 mM Mn^{2+} ions was 17- and 8-fold higher than with 1 and 10 mM Mg^{2+}, respectively. Similarly, and as shown in Figure 4B, the length of the polymerization products was also longer even at a relatively low concentration of Mn^{2+} (100 μM) than with physiological concentrations of Mg^{2+} [5–10 mM; (53)]. The presence of a saturating Mg^{2+} concentration (3 mM) seemed to have a synergistic effect on polymerization with suboptimal Mn^{2+} concentrations (10 and 20 μM) although far from the maximum activity, and with no significant differences regarding the absence or presence of Mg^{2+} at the tested Mn^{2+} concentrations higher than 20 μM (Figure 4C). Although the intracellular Mn^{2+} concentration in *E. coli* has been measured to be 10 μM, this can vary over at least two orders of magnitude [1–3 mM; (54)]. Thus, it is tempting to speculate that, *in vivo*, the AEP domain of IEE would utilize preferentially Mn^{2+} even when Mg^{2+} is present in molar excess. This phenotype of preferential usage of Mn^{2+} as a metal activator of polymerization activity has also been observed mainly in DNA polymerases involved in either DNA repair or bypass of DNA lesions, including: (i) the eukaryotic NHEJ DNA polymerases λ (55,56) and μ (57); (ii) members of the AEP superfamily such as the PolDom of bacterial LigD responsible for NHEJ reactions (36,58–61); (iii) human PrimPol involved in TLS (62); as well as in (iv) human TLS DNA polymerase ι (63). The ability of most of the aforementioned DNA polymerases to insert nucleotides opposite damaged bases, in the case of the TLS polymerases, or to promote misalignments to search for microhomologies during the end-joining reaction in NHEJ PolDoms, relies on the great flexibility of the active site that could be conferred by the more relaxed coordination of the Mn^{2+} ions when compared with Mg^{2+} (64). In

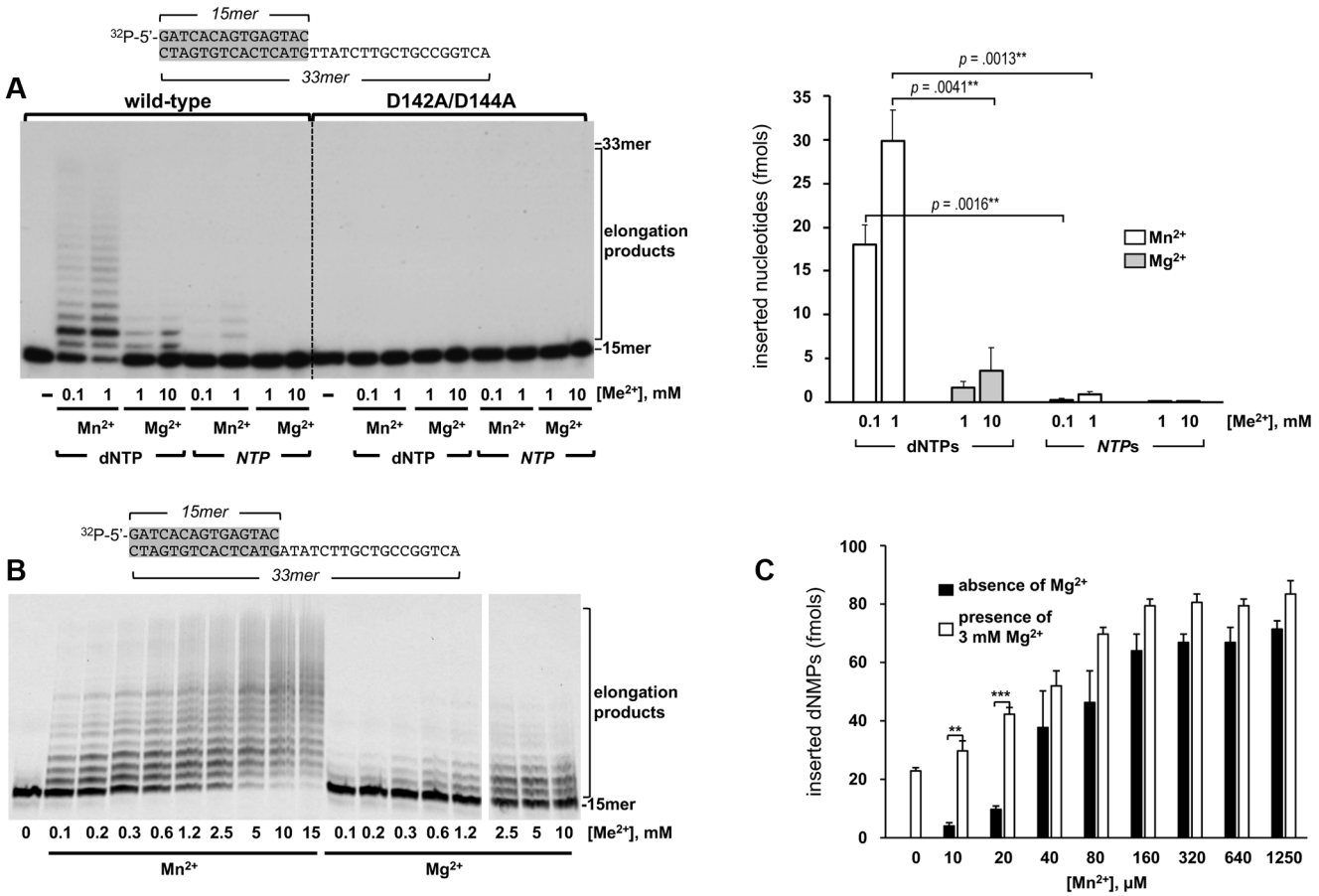

Figure 4. Polymerization activity of IEE on a template/primer substrate. (**A**) *IEE preferentially inserts dNTPs*. The assay was performed as described in Materials and Methods, with 10 nM of either wild-type or D142A/D144A mutant IEE, 1 nM of the ^{32}P-5-labeled primer/template substrate (schematized on top of the figure), 10 μM of either dNTPs or NTPs and the indicated concentration of either Mn^{2+} or Mg^{2+}. After incubation for 5 min at 37°C, the reactions were stopped by adding EDTA up to 10 mM. The position of the primer and elongation products is indicated. Bar chart shows the fmols of incorporated nucleotide under the indicated condition ($n = 3$ each; means ± SEM). Significance of results was determined with a two-tailed paired *t*-test. (**B**) IEE uses preferentially Mn^{2+} ions as metal activator of the polymerization reaction. The assay was carried out essentially as described in (A) in the presence of increasing concentrations of either Mn^{2+} or Mg^{2+}. Figure is a composite image made from different parts of the same experiment. (**C**) Polymerization in the simultaneous presence of Mg^{2+} and Mn^{2+} ions. The assay was carried out as described in (A) with increasing concentrations of Mn^{2+} (10, 20, 40, 80, 160, 320, 640 and 1250 μM), and either in the absence or presence of 3 mM Mg^{2+}. Bar chart shows the fmols of inserted nucleotide under the indicated condition ($n = 3$ each; means ± SEM). Significance of results was determined with a two-tailed paired *t*-test. $^{**}P < 0.01$; $^{***}P < 0.001$. Total nucleotide insertion was obtained by calculating the number of catalytic events giving rise to each polymerization product.

this sense, it is noteworthy that comparison of crystal structures of Pol ι ternary complexes with Mg^{2+} or Mn^{2+} at the active site revealed that Mn^{2+} achieves more optimal octahedral coordination geometry than Mg^{2+} (65). Therefore, the preferential use of Mn^{2+} by the AEP domain of IEE could be anticipating the ability of the enzyme to promote rearrangements of the DNA strands during the polymerization reaction (see below).

IS-excision enhancer can promote distortions of both template and primer strands

Unlike replicative DNA polymerases, the polymerization active site of AEPs is designed to promote and accommodate dislocations of the template and primer strands, as well as to extend 3'-terminal mismatched base pairs. In bacterial AEPs such as the PolDom of LigDs, the flexibility of the polymerization site allows the protein to search for microhomologies and bridge two protruding 3'-ends of a DSB during the nonhomologous end-joining reaction (36,60,66–69). The dislocation capacity of the primer and template strands is also instrumental for skipping DNA lesions by human PrimPol during TLS (62), facilitating the search for microhomologies beyond the unreadable damage.

Having shown that IEE is a superfamily AEP DNA polymerase using preferentially Mn^{2+} as metal activator, we next analyzed its ability to select among the four deoxynucleotides (base discrimination) to catalyze faithful DNA synthesis. Thus, we assayed the incorporation of each of the four dNTPs individually on the four primer/template DNA structures depicted in Figure 5A, covering the 16 possible template-substrate nucleotide pairs (4 matched + 12 mismatched). We found that IEE is a DNA-instructed DNA polymerase that although preferentially inserts the complementary nucleotide dictated by the first available templat-

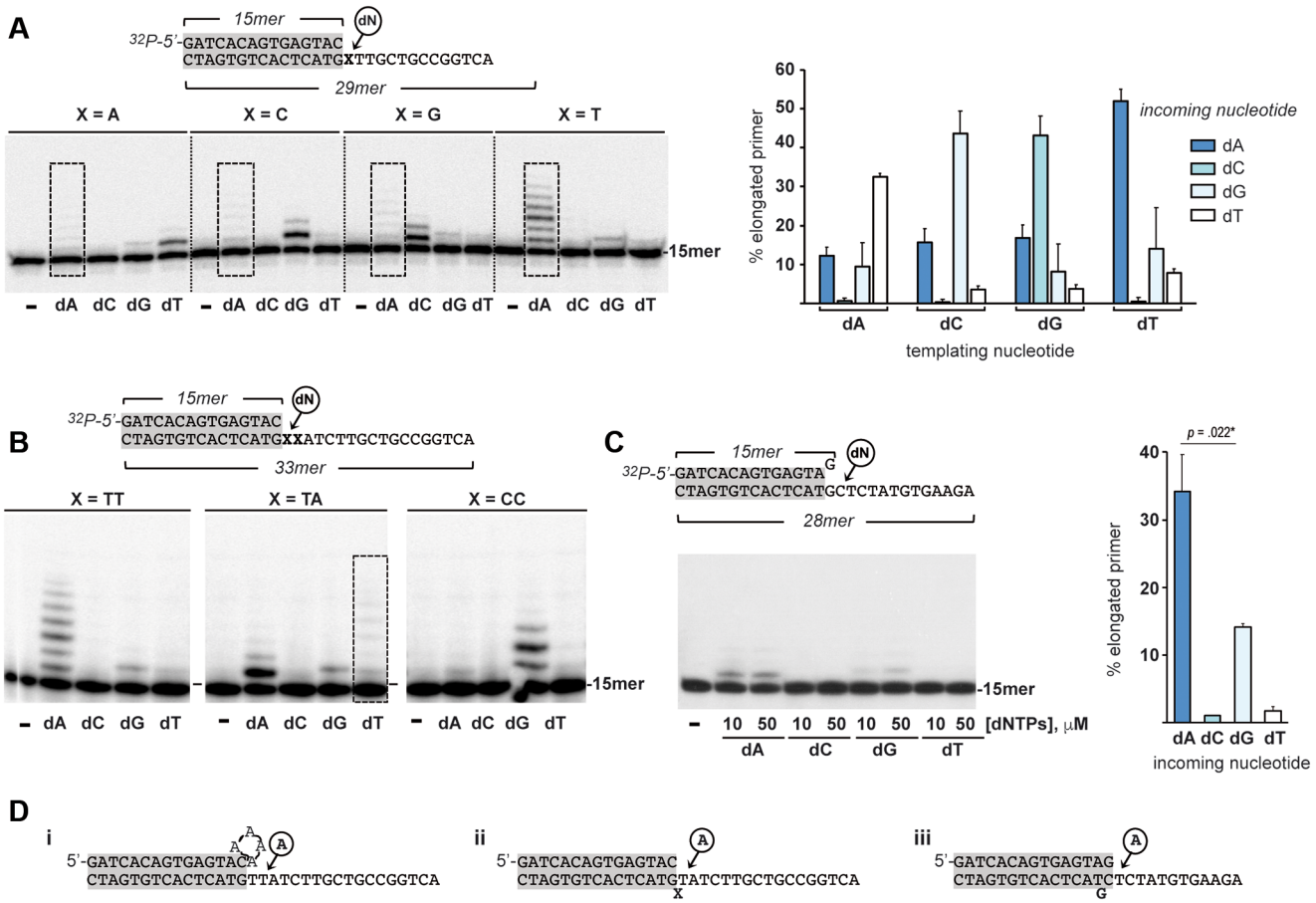

Figure 5. Primer and template dislocations mediated by IEE. (**A**) *IEE preferentially inserts the complementary nucleotide dictated by the first available templating base.* The assay was performed as described in Materials and Methods, with 1 nM of the 5'-labeled substrate schematized on the top of figure, 20 nM of IEE, 1 mM Mn^{2+} and 1 μM of the indicated nucleotide. After incubation for 10 min at 37°C, the reactions were stopped by adding EDTA up to 10 mM. Bar chart shows the percentage of elongated primer obtained in the presence of the indicated incoming and template nucleotide ($n = 3$ each; means ± SEM). (**B**) *Primer expansion mediated by slippage.* The assay was performed as described in (A) in the presence of 1 nM of the 5'-labeled substrates depicted on top of the figure. (**C**) *Slippage-mediated template dislocation.* The assay was carried out as described in (A), in the presence of 1 nM of the 5'-labeled substrate shown on top of the figure and the indicated concentration of the specified dNTP. Bar chart shows the percentage of elongated primer obtained in the presence of the indicated incoming nucleotide ($n = 3$ each; means ± SEM). (**D**) *Schematic representation of the primer and template dislocations mediated by IEE.*

ing base, shows a relatively high misinsertion capacity, elongating up to 16.8% of the primer molecules in the presence of an erroneous incoming nucleotide (see also bar chart in Figure 5A). Notably, under those experimental conditions, polymerization was not limited to the addition of a single nucleotide; IEE inserted mainly two nucleotides on those substrates where the first and second available templating bases were dissimilar. In the cases where the first and second templating bases were the same, the correct nucleotide produced an expansion mediated by reiterative slippage of the primer strand (primer dislocation), allowing IEE to add 5–7 nucleotides (see insertion of dA opposite T in Figure 5A and B, and insertion of dG opposite CC in Figure 5B). In addition to copying preferentially the first available nucleotide, IEE could also skip it, adding nucleotides opposite the next base (see insertion of dA indicated by rectangles in Figure 5A, and dT in the middle panel of Figure 5B). Thus, the IEE polymerization active site also tolerates a dislocation of the template base.

To examine the mismatch extension capacity of IEE, we performed primer extension assays starting from a dG:dG base-pair mismatch (see schematic representation in Figure 5C). The lack of 3'-5' exonucleolysis in IEE (see the lane in the absence of nucleotides in Figure 5C) prevents the removal of the 3'-mispaired nucleotide, with only two outcomes being possible: (i) direct mismatch extension inserting dG; and (ii) template dislocation (primer-realignment mediated) inserting dA. As shown in Figure 5C, whereas IEE was very inefficient in performing the direct mismatch extension (only 14.1% of elongated primer molecules with dG, see also bar chart in Figure 5C), in the presence of dA IEE realigned up to 34% of the 3'-terminal dG to form a correct base-pair with the templating dC, promoting a single base distortion in the template that allows further insertion of dA.

Altogether, these results show that IEE is an error-prone DNA polymerase, sharing with the TLS and NHEJ AEP enzymes the ability to induce/accept dislocations of the primer and template strands such as slippage-mediated primer dislocation (Figure 5D,i), dNTP selection-mediated template dislocation (ii), and template dislocation mediated by primer realignment (iii).

IS-excision enhancer promotes MMEJ *in vitro*

It is generally assumed that the excision of IS elements from bacterial genomes is a rare event since most bacteria lack the end-joining system needed to repair the donor DNA after IS excision. As we mentioned earlier, it has been reported that IEE promotes the excision of IS elements in enteropathogenic *E. coli* strains such as O157:H7 in a transposase-dependent manner (10). The presence of an N-terminal AEP domain in IEE with polymerization activities similar to those of bacterial end-joining PolDom with respect to the flexibility of their polymerization active sites to accommodate primer/template dislocations and to the reduced nucleotide insertion fidelity, prompted us to evaluate the capacity of IEE to promote the annealing and further elongation of DNA ends with limited complementarity. To do this, we made use of the dsDNA substrates depicted in Figure 6 (70), mimicking partially resected DSBs containing either no microhomology (dsDNA-0) or a 2, 4 and 6 nucleotide complementary region at the protruding 3' ends (dsDNA-2, -4 and -6, respectively). Accordingly, upon annealing, 4-nt gap hybrids are formed where the 3'-OH group of the protruding strand could be used as primer to copy the opposing overhang *in trans* (see scheme in Figure 6). The calculated melting temperatures of the complementary regions are 4°C, 16°C and 24°C, substantially lower than the reaction temperature (37°C), precluding the spontaneous annealing in the absence of the protein. As shown in Figure 6A, in the presence of the four dNTPs IEE elongated 28% and 46% of the primers when the annealing region was 4 and 6 bp long, respectively (see also bar chart in Figure 6A). The almost total absence of synthesis products with substrates dsDNA-0 and dsDNA-2 would rule out the possibility that the elongation products using dsDNA-4 and dsDNA-6 were due to a potential terminal transferase activity of IEE. That the base excision repair DNA polymerase X from *B. subtilis*, the replicative family B DNA polymerase from bacteriophage phi29, and the Klenow fragment from family A *E. coli* DNA polymerase I, were incapable of elongating the protruding 3' end of dsDNA-4 suggests that the observed elongation reaction with such a substrate is specific to IEE (see Supplementary Figure S4). As shown in Figure 6B (left panel), the addition of dTTP as the only nucleotide, or dTTP + dATP, resulted mainly in the synthesis of +1 and +(1–3) products, respectively, strongly suggesting that nucleotide incorporation by IEE is dictated by the sequence adjacent to the potential pairing region. To demonstrate unequivocally that nucleotide addition was dependent on the connection of the 3' ends and not on self-annealing (snap-back), the elongation reaction was analyzed by simultaneously adding two different protruding DNA 3'-ends (only one labeled) that differed in the nucleotide adjacent to the potential pairing region (labeled as X in the scheme in Figure 6B, right panel). As shown, IEE preferentially added to the 3'-end of the labeled strand the dNMP complementary to the X nucleotide of the non-labeled strand, indicating on the one hand that X was acting as template, and on the other hand that synapsis between the labeled and unlabeled dsDNA molecules occurred. In addition to the elongation products +1–4, the simultaneous presence of IEE and *E. coli* LigA gave rise to a small amount of a product with lower electrophoretic mobility that would correspond to the ligation of 18.9 ± 1.4% of the +4 product to the 5'-P end of the downstream strand in substrate dsDNA-6 (see Figure 6C).

We also analyzed the ability of IEE to promote end-joining of ssDNA with terminal microhomology. To do this, IEE was incubated with ssDNA-4 in the presence of nucleotides. As shown in Figure 7A, IEE preferentially inserted dTMP onto the 3'-end of the strand (+1 product in the figure), giving rise to a +3 product in the presence of dTTP and dATP. As with dsDNA substrates, the simultaneous addition of labeled ssDNA-4 with non-labeled ssDNA substrates differing in the nucleotide adjacent to the pairing region (X in Figure 7B) provoked the preferential insertion of the dNMP complementary to the X nucleotide by IEE. Overall, the results show that IEE can mediate the bridging of two DNA ends.

The IS-excision enhancer C-terminal helicase-like domain allows the synapsis of 3'-ends

The C-terminal domain of IEE shows homology with SF2 helicases, containing the Walker A and Walker B motifs involved in NTP binding and hydrolysis. We next analyzed the ability of IEE to hydrolyze ATP by incubating the protein with 1 mM ATP (0.2 μCi [γ^{32}P]ATP) in the presence of either dsDNA or ssDNA, as other helicases require these substrates for ATP binding and further hydrolysis (71,72). As shown in Figure 8A, IEE released Pi, which was separated from non-hydrolyzed ATP by thin-layer chromatography, specifically in the presence of DNA (both ss- and dsDNA). The absence of ATPase activity in the catalytically dead mutant K451A mutant (Figure 8B) allowed us to conclude the presence of a DNA-dependent ATPase activity in the C-terminal domain of the protein. However, despite both the presence in the C-terminal domain of ATPase activity and homology with helicases, IEE failed to exhibit the ability to couple the ATPase to the unwinding of dsDNA substrates, either having 3' or 5' unpaired tails [see Figure 8C, RecD2 and PriA were used as control of 5'- and 3'-helicase activity, respectively (73–75)]. Analysis of the polymerization activity of the wild-type enzyme on a 5-nucleotide-gapped molecule revealed that IEE possesses a limited capacity to couple polymerization to the displacement of the downstream strand, since up to 44% of the 20mer products were further elongated (see Figure 9A). A similar result was obtained with mutant K451A, indicating that strand displacement is not concomitant with ATP hydrolysis. Similarly, the mutant enzyme could promote the synapsis of two 3'-ends (Figure 9B). To further analyze the role of the C-terminal helicase-like domain of IEE in the synthetic activity of the protein, we obtained a deletion mutant lacking residues 289–786 by inserting a stop at codon 289 (AEP-Dom, see Materials and Methods). As shown in Figure 10A, the AEP-Dom exhibited a polymerization efficiency similar to the wild-type protein on a primer/template molecule, but it stopped after adding 9–12 nucleotides. Indeed, the AEP-Dom could insert mainly three nucleotides when the length of the template was shortened (Figure 10B). To determine the relative orientation of the N-terminal AEP and C-terminal

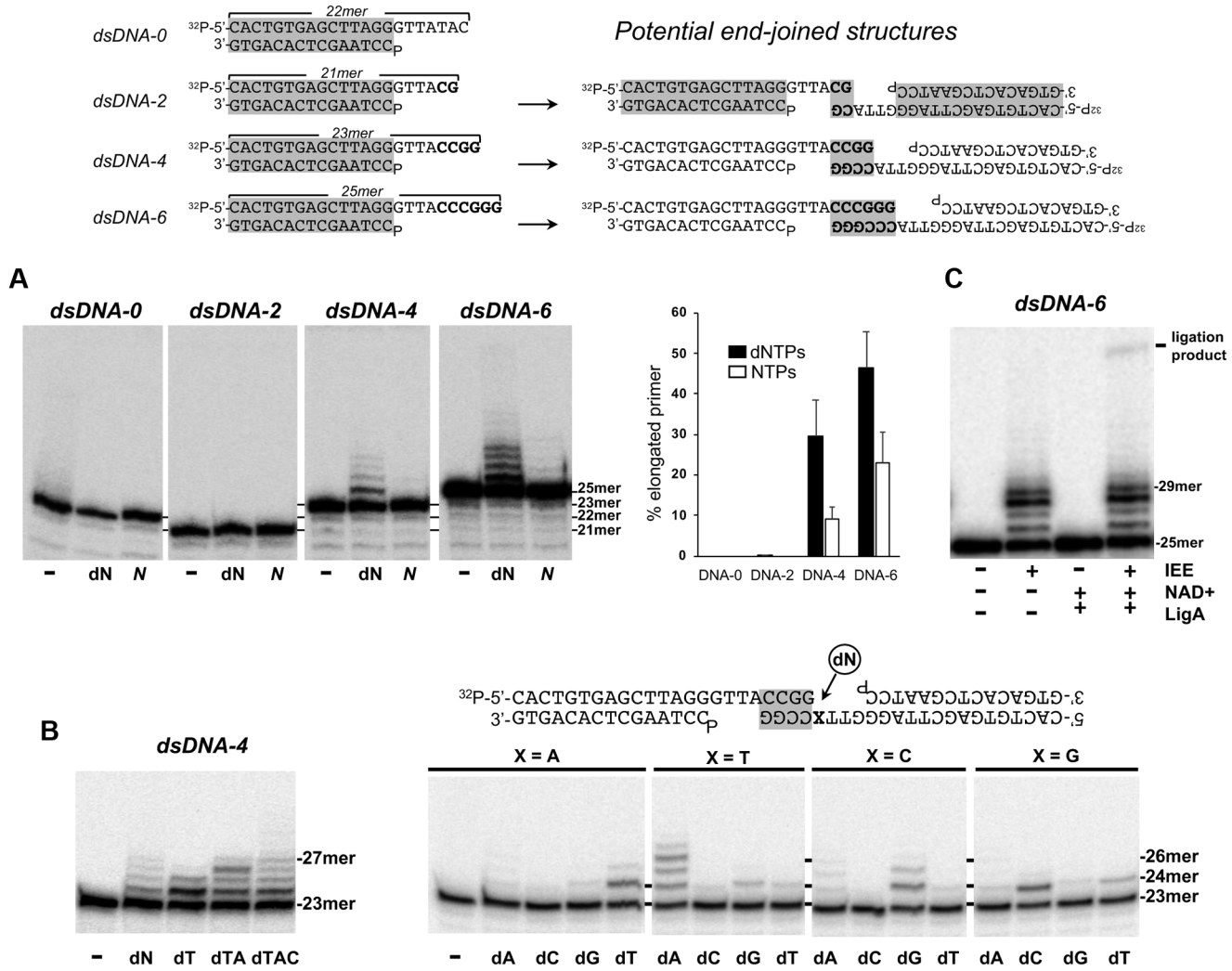

Figure 6. IEE promotes MMEJ *in vitro*. (**A**) *MMEJ reactions with dsDNA substrates.* The assay was performed as described in Materials and Methods, in the presence of 25 nM of the indicated 5'-labeled dsDNA schematized on top of the figure, 40 nM of IEE, and 10 μM of either dNTPs or NTPs, as indicated. After incubation for 15 min at 37°C, the reactions were stopped by adding EDTA up to 10 mM. Position of the substrates is indicated. Bar chart shows the percentage of elongated primers obtained in the presence of either dNTPs or NTPs ($n = 2$ each; means ± SEM). (**B**) *IEE elongates the protruding 3'-end using the opposing overhang as a template in trans*. Assays were performed as described in (A) in the presence of either 25 nM of 5'-labeled dsDNA-4 substrate (left panel) or 1 nM of 5'-labeled dsDNA-4 and 20 nM of the indicated non-labeled dsDNA molecule (right panel). Position of the substrate and the elongation products is indicated. (**C**) *Complete MMEJ reaction mediated by IEE and LigA*. The assay was performed as described in Materials and Methods, in the presence of 25 nM of dsDNA-6, 40 nM IEE, 10 μM dNTPs, 1 mM $MnCl_2$, 5 mM $MgCl_2$ and, when indicated, 500 μM NAD^+ and 1 unit of *E. coli* LigA (NEB). After incubation for 15 min at 37°C, samples were processed and analyzed as described in (A). Position of the substrate and reaction products is indicated.

helicase domains respect to the DNA, the structures of IEE (this work) and the ternary complex from PrimPolC (37) were overlapped by structural alignment of their catalytic active sites (see Figure 10C and Supplementary Figure S5). As shown, primer/template dsDNA portion would fit in the groove defined by the two core RecA-like folds of the helicase domain. Such an IEE-DNA complex, together with the shorter elongation products obtained with the AEP-Dom strongly suggest a role for the C-terminal domain in establishing contacts with the dsDNA portion to allow the complete replication of the templating strand by the wild-type protein when using primer/template substrates. Of note, the polymerization pattern observed with the AEP-Dom on the 5-nucleotide-gapped molecule showed a major pause once the gap was filled (20mer product) much stronger than with the wild-type IEE, the deletion mutant being 3-fold less efficient than the wild-type enzyme in displacing the downstream strand (Figure 10A, right panel, see also bar charts in Figure 10A). This result suggests that the partial displacement of the downstream strand coupled to DNA synthesis exhibited by IEE would depend on the proper stabilization of the IEE/DNA complex contributed by the helicase domain. Interestingly, AEP-Dom appears to have a higher efficiency than the wild-type enzyme in filling the gap, that could be reflecting a higher turnover of the enzyme due to the absence of the C-terminal helicase domain. Accordingly, IEE joins the increasing number of proteins possessing a helicase domain that, despite having the characteris-

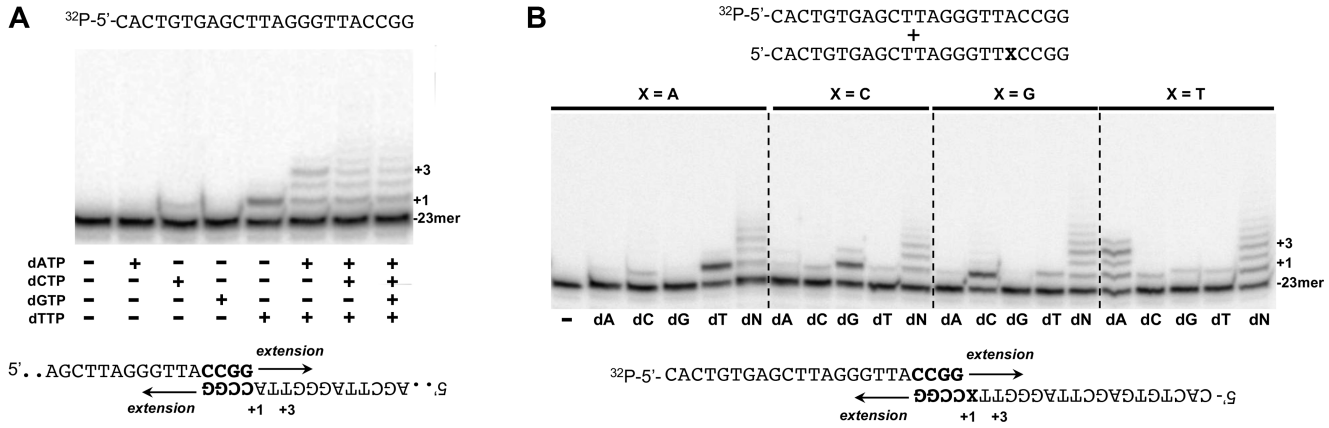

Figure 7. IEE can accomplish ssDNA synapsis. (**A**) *MMEJ reactions with a ssDNA substrate.* The assay was carried out as described in Materials and Methods, in the presence of 20 nM of the 5'-labeled molecule depicted on top of the figure, 40 nM IEE, and 10 μM of the indicated dNTP. After incubation for 15 min at 37°C the reactions were stopped by adding EDTA up to 10 mM. Position of the substrate and reaction products is indicated. (**B**) *IEE elongates the ssDNA 3'-end using the opposing ssDNA molecule as a template in trans*. Assays were performed as described in Materials and Methods, incubating 40 nM of IEE with 1 nM of the 5'-labeled ssDNA and 20 nM of the specified non-labeled ssDNA molecule (see scheme on top of the figure), and 10 μM of the indicated dNTP. After incubation for 15 min at 37°C the reactions were stopped by adding EDTA up to 10 mM. Position of the substrate and the elongation products is indicated.

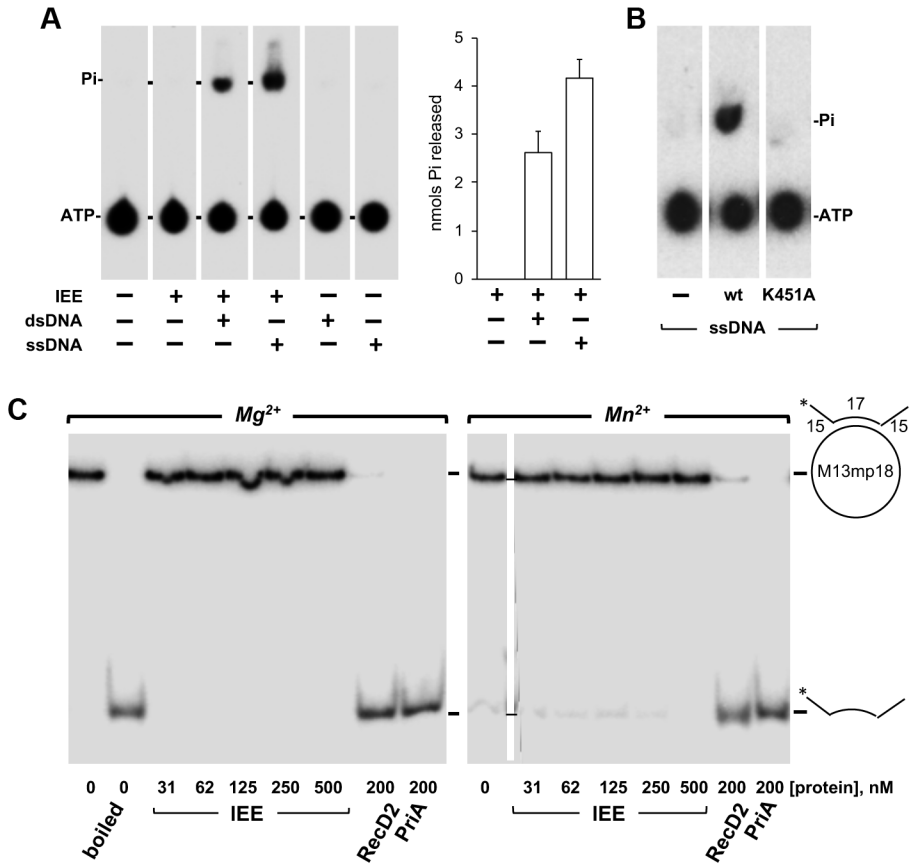

Figure 8. IEE has a DNA-dependent ATPase activity not coupled to unwinding of dsDNA substrates. (**A**) *IEE has a DNA-dependent ATPase activity*. The assay was performed as described in Materials and Methods by incubating 100 nM IEE with 1 mM ATP in the presence (+) or absence (−) of 50 ng/μl of either the double-stranded pUC18 plasmid or single-stranded DNA from phage M13. After incubation for 20 min at 30°C, reactions were stopped by adding 30 mM EDTA, and reaction products were analyzed by thin-layer chromatography and autoradiography. Bar chart shows the nmols of Pi released under the indicated condition ($n = 2$ each; means ± SEM). (**B**) *ATPase activity resides in the C-terminal helicase domain of IEE*. The assay was performed essentially as in (A), in the presence of 50 ng/μl of single-stranded DNA from phage M13 and 100 nM of either the wild-type or mutant K451A IEE (see main text for details). (**C**) *IEE has no helicase activity*. The assay was carried out as described in Materials and Methods by incubating the indicated concentration of either IEE, RecD2 or PriA with 1 nM of the indicated ^{32}P-labeled substrate and either 1 mM Mn^{2+} or 5 mM Mg^{2+}. After incubation for 30 min at 37°C, the reaction was stopped and the products were resolved in 10% non-denaturing polyacrylamide gel (0.1% SDS).

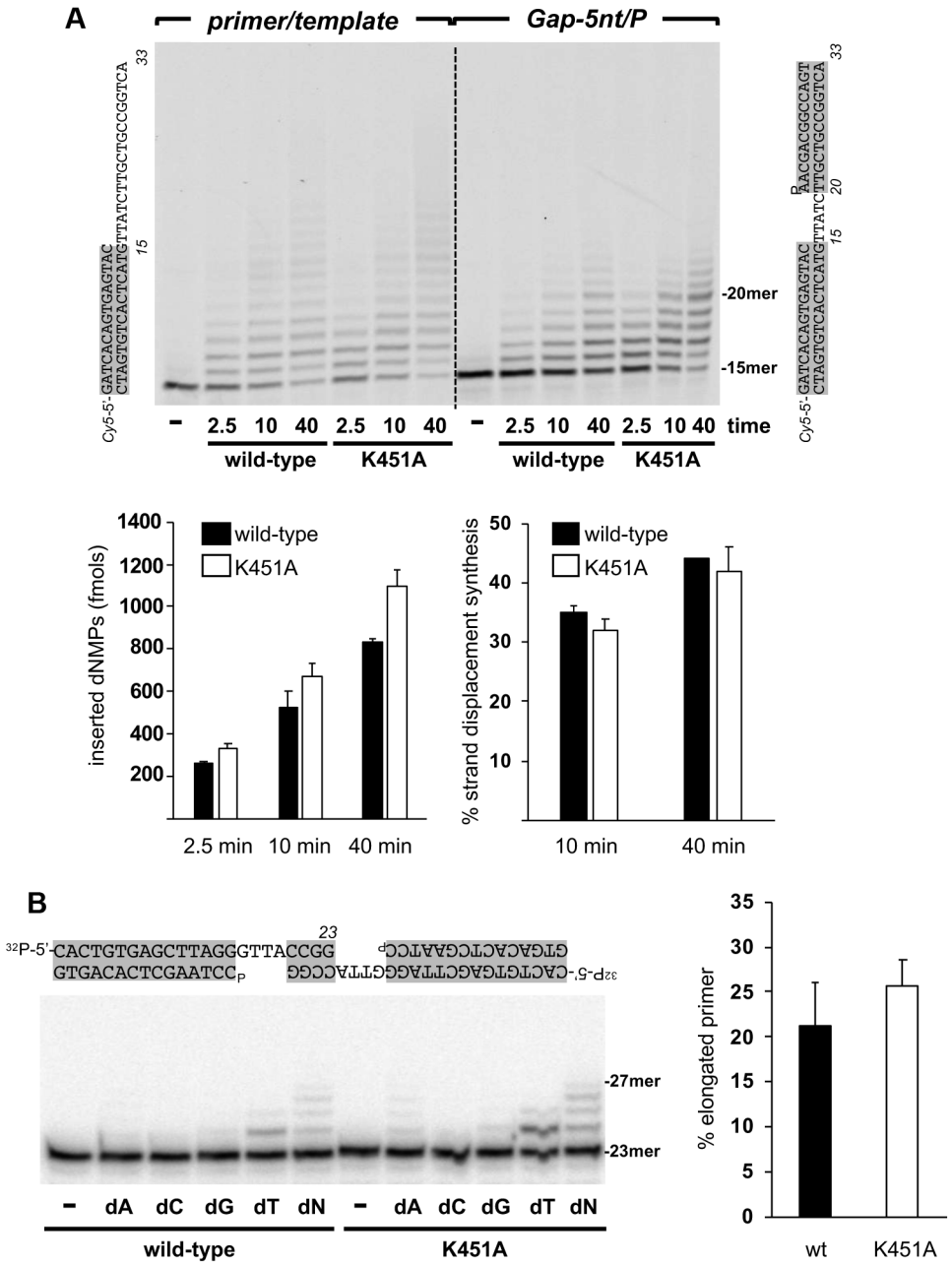

Figure 9. The limited strand displacement coupled to polymerization and the MMEJ reaction accomplished by IEE do not depend on ATP hydrolysis. (**A**) The assay was performed as described in Materials and Methods in the presence of 20 nM of either a primer/template (left panel) or a 5-nucleotide gapped (right panel) DNA, 50 μM dNTPs and 25 nM of either wild-type or mutant K451A mutant IEE. After incubation at 37°C for the indicated times, the reactions were stopped by adding EDTA up to 10 mM. The position of the primer and elongation products is indicated. Bar chart (left) shows the fmols of nucleotides incorporated at each reaction time with the primer/terminus substrate. The strand displacement capacity of both the wild-type and mutant K451A is estimated as the ratio between the products longer than 20 nucleotides and the products ≥ 20 nucleotides obtained with the gapped DNA substrate at the indicated times (bar chart on the right; $n = 2$ each; means ± SEM). (**B**) The reaction was performed as described in Figure 6B in the presence of 40 nM of either wild-type or mutant K451A IEE. Bar chart shows the percentage of elongated primers in the presence of the four dNTPs ($n = 3$ each; means ± SEM).

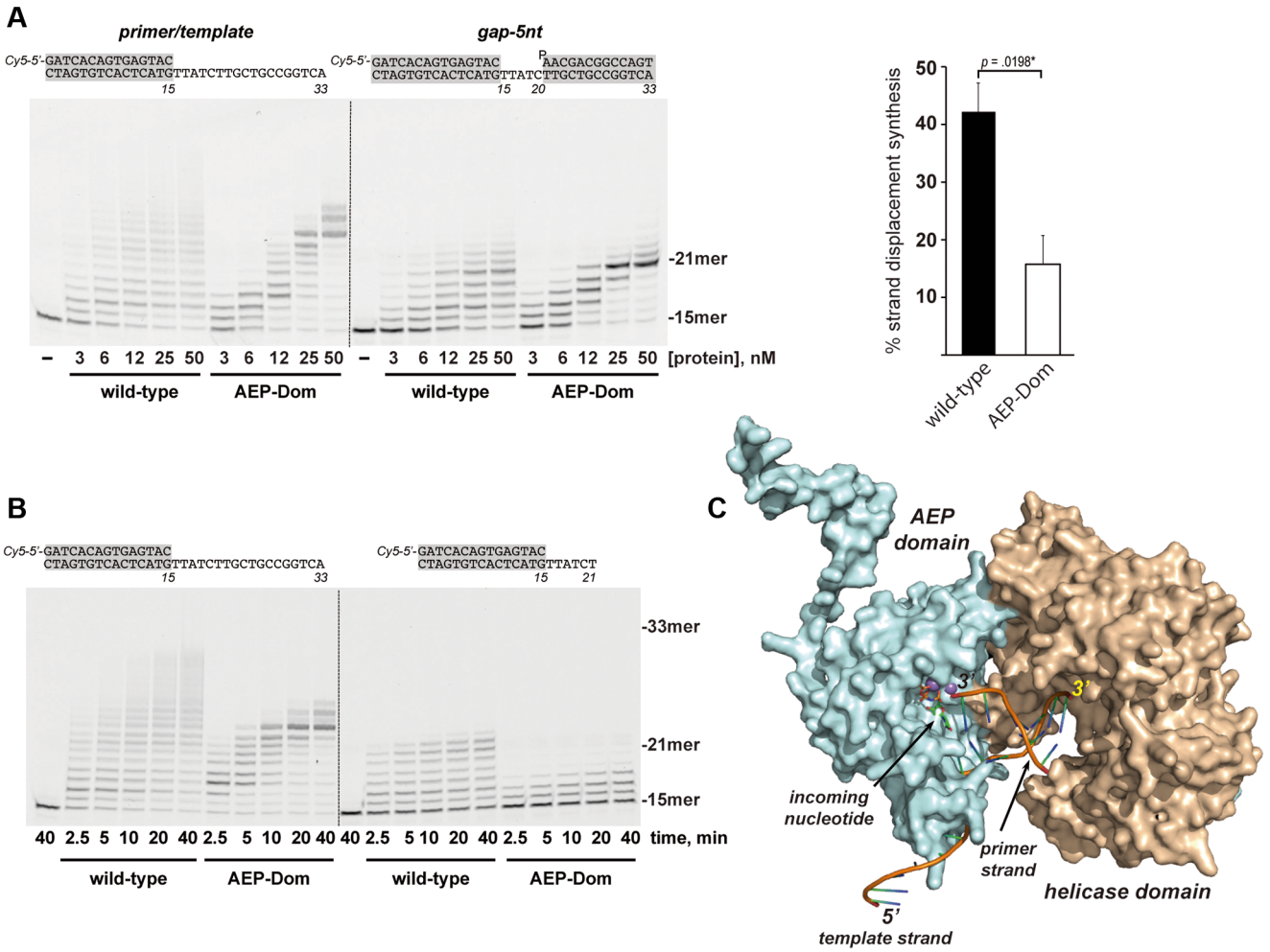

Figure 10. The helicase domain of IEE is required to couple polymerization to strand displacement and for the synapsis of two DNA ends. (**A**) *Deletion of the C-terminal helicase domain impairs the strand-displacement capacity of IEE.* The assay was carried out as described in Materials and Methods in the presence of 20 nM of the DNA substrate depicted on top of the figure and the indicated concentration of either the wild-type or the helicase deletion mutant (AEP-Dom) IEE. After incubation for 10 min at 37°C, the reactions were stopped by adding EDTA up to 10 mM. The position of the primer and elongation products is indicated. Bar chart on the right shows the quantitation of the strand displacement capacity of both the wild-type IEE and AEP-Dom estimated as the ratio between the products longer than 20 nucleotides and the products ≥20 nucleotides obtained in the presence of 25 nM of enzyme ($n = 3$ each; means ± SEM). Significance of results was determined with a two-tailed paired *t*-test. (**B**) *The helicase domain stabilizes the interaction of IEE with the DNA.* The assay was performed as above, in the presence of 20 nM of either the wild-type IEE or the AEP-Dom. The samples were incubated at 37°C for the indicated times and further processed as described in (A). (**C**) Modeling of the interaction of IEE with a primer/template DNA substrate. The figure was done by superposition of the catalytic active sites of IEE and the ternary *M. smegmatis* PrimPolC–DNA–NTP complex [(37); PDB 6SA0]. The incoming nucleotide, metal ions A and B and DNA from PrimPolC are shown as sticks, purple spheres and ribbons, respectively. IEE is shown as surface representation. Figure was generated using The Open-Source Pymol Molecular Graphics System, v. 2.5.0, Schrödinger, LLC (Open-Source PyMOL is Copyright (C) Schrodinger, LLC.)

tic sequence motifs as well as a DNA-dependent ATPase activity, lack an ATP-dependent unwinding activity, as described for proteins of the Swi/Snf family and the ATP-dependent restriction enzymes (76,77). It has been shown that SF2 proteins can have mechanical functions other than the unwinding of nucleic acids that would contribute to cellular metabolism (78,79), such as ATP-dependent remodeling of chromatin (80), or protein displacement (81). As IEE has been shown to promote the excision of the insertion elements in a transposase-dependent manner, it is tempting to speculate a role for its helicase domain also in the displacement of the transposase once it has incised the DNA.

The AEP-Dom displayed a higher nucleotide insertion fidelity than the full-length IEE, copying preferentially the first templating nucleotide (see Figure 11A), and it did not produce the primer expansion observed with the wild-type protein when either the first and second or the three templating bases were the same. Although it cannot be ruled out that the C-terminal helicase domain is allowing AEP-Dom to efficiently accommodate the strand slippage intermediates, the relative orientation of both domains respect to the DNA substrate, as well as the role of the helicase domain in stabilizing the IEE/DNA complex (Figure 10) strongly suggest that the interaction between the helicase domain and the upstream dsDNA region of the substrate is required to

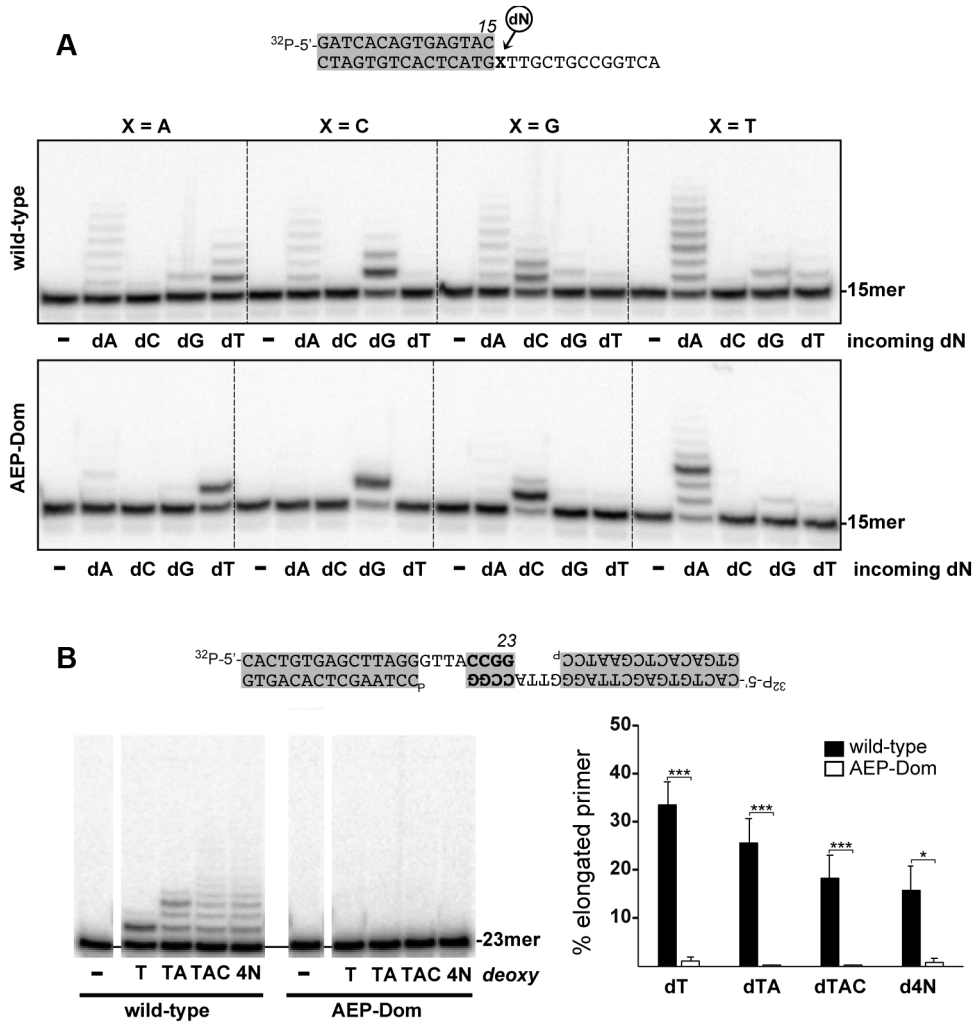

Figure 11. The helicase domain/DNA interaction allows dislocations of the primer and template strands and the synapsis of two DNA ends. (**A**) The assay was carried out in the presence of 20 nM of enzyme, 10 μM dNTPs and 1 nM of the indicated ^{32}P-labeled substrate. After incubation for 15 min at 37°C, the reactions were stopped by adding EDTA up to 10 mM. (**B**) The reaction was performed as described in Figure 6B in the presence of 40 nM of enzyme and 1 μM of the indicated dNTP. The figure is a composite image made from different parts of the same experiment. Bar chart on the right shows the percentage of elongated primers in the presence of the indicated combination of dNTPs ($n = 3$ each; means ± SEM). Significance of results was determined with a two-tailed paired *t*-test. *$P < 0.05$; ***$P < 0.001$.

allow IEE to hold down primer dislocations. Notably, the AEP-Dom was unable to promote the synapsis between two 3'-ends, since almost no elongation products were detected when the protein was incubated with the substrate DNA-4 (Figure 11B). Altogether, the results suggest a role for the C-terminal helicase domain in stabilizing the interaction between IEE and the DNA substrates.

Concluding remarks

Proliferation of IS elements generates beneficial mutations that promote both genome evolution and bacterial adaptation; however, the uncontrolled increase in the abundance of IS elements could be detrimental (82). Excision of IS elements, together with other genetic strategies such as inhibition of the transposase, epigenetic modifications, and negative regulation by specific repressors and RNA interference, are some of the mechanisms used by bacteria to control IS proliferation [reviewed in (83)].

Previous work revealed the existence of a novel transposase- and IEE-dependent IS excision system in bacteria (10). It is notable that most bacterial species lack an end-joining activity that could restore and safeguard the bacterial genome after the generation of DSBs by the transposase during the excision of the IS. Here, we determined the biochemical characteristics of the IEE protein, showing how the simultaneous presence of an AEP domain, typically present in proteins involved in DNA repair transactions or bypass of DNA lesions, and of a helicase domain, allows the enzyme to search for microhomologies and bridge two protruding 3'-ends of a DSB during an *in vitro* end joining reaction. These results lead us to postulate a functional coordination between IEE and the cognate transposase whereby, once the dsDNA breaks are generated by the transposase, IEE would be

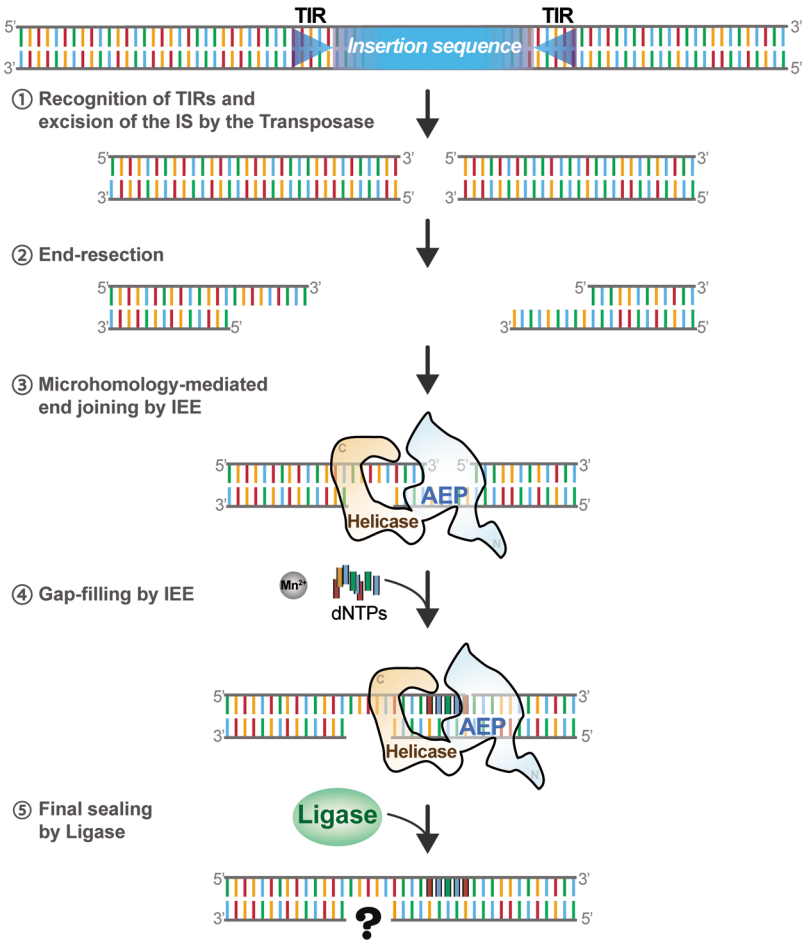

Figure 12. Proposed mechanism for the generation of the large deletions during IEE-dependent IS excision. The TIR sequences flanking the IS element are recognized and cleaved by the cognate transposase (1). The resulting DSB in the bacterial genome could be potentially resected by cellular nucleases to produce 3' single-stranded tails (2). IEE may be able to perform the microhomology annealing (3) and catalyze further extension of the paired 3'-end (4). Finally, the resulting nick could be sealed by a DNA ligase (5). The other 3'-end could be elongated either by a second IEE molecule or by another cellular DNA polymerase.

essential to search for microhomologies between both ends and allow their subsequent repair. IS elements and their transpositions are generally regarded as one of the major driving forces generating various genomic deletions. It has been shown that during IEE-dependent IS excision, deletions from a few nucleotides to >40 000 occur adjacent to the IS (10). In addition, a similar deletion has been recently reported during the *in vitro* formation of operons catalyzed by the IEE/transposase system (13). This observation led the authors to speculate that such deletions could be due to aberrant breaks generated by the transposase at sites remote from the IS. However, it cannot be ruled out that once the dsDNA breaks are generated, other unidentified cellular factors, such as exonucleases, could process the ends of these breaks until IEE finds a microhomology from which to carry out the end-joining reaction (Figure 12).

Finally, it is tempting to speculate that beyond its potential role in the end-joining reaction mentioned above, the biochemical features of IEE, such as its ability to promote dislocations of both the primer and template strand, could enable the enzyme to contribute to the increased variability and adaptability to different environmental conditions of IEE-encoding bacteria.

DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material.

The predicted structure of IEE from *E. coli* O157:H7 was obtained through the Uniprot web site (https://www.uniprot.org/) from the Alphafold Database ID A0A0H3JF09 (30), and used under the Creative Commons Attribution 4.0 International (CC BY 4.0) License.

All the analyzed *E. coli* genomic sequences were retrieved from GenBank Assemblies database (accessed on 24 February 2022), using NCBI Datasets tool. IEE-positive assemblies are detailed in Supplementary Table S1.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Dr Silvia Ayora (Centro Nacional de Biotecnología, Madrid) for proteins RecD2 and PriA.

FUNDING

MCIN/AEI/10.13039/501100011033 [PID2020-115978GB-I00 to M.V.]; MCIN/AEI/10.13039/501100011033/FEDER A way to make Europe [PID2021-123403NB-I00 to M.R.R.]; Comunidad de Madrid (V PRICIT call Research Grants for Young Researchers from Universidad Autónoma de Madrid) [SI3/PJI/2021-00271 to M.R.R.]; institutional grant from Fundación Ramón Areces to the Centro de Biología Molecular Severo Ochoa. Funding for open access charge: Consejo Superior de Investigaciones Científicas.

Conflict of interest statement. None declared.

REFERENCES

1. Consuegra,J., Gaffé,J., Lenski,R.E., Hindré,T., Barrick,J.E., Tenaillon,O. and Schneider,D. (2021) Insertion-sequence-mediated mutations both promote and constrain evolvability during a long-term experiment with bacteria. *Nat. Commun.*, **12**, 980.
2. Hawkey,J., Hamidian,M., Wick,R.R., Edwards,D.J., Billman-Jacobe,H., Hall,R.M. and Holt,K.E. (2015) ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. *BMC Genomics*, **16**, 667.
3. Jellen-Ritter,A.S. and Kern,W.V. (2001) Enhanced expression of the multidrug efflux pumps AcrAB and AcrEF associated with insertion element transposition in *Escherichia coli* mutants selected with a fluoroquinolone. *Antimicrob. Agents Chemother.*, **45**, 1467–1472.
4. Siguier,P., Gourbeyre,E., Varani,A., Ton-Hoang,B. and Chandler,M. (2015) Everyman's guide to bacterial insertion sequences. *Microbiol. Spectr.*, **3**, MDNA3-A0030–2014.
5. Vandecraen,J., Chandler,M., Aertsen,A. and Van Houdt,R. (2017) The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.*, **43**, 709–730.
6. Ross,K., Varani,A.M., Snesrud,E., Huang,H., Alvarenga,D.O., Zhang,J., Wu,C., McGann,P. and Chandler,M. (2021) TnCentral: a prokaryotic transposable element database and web portal for transposon analysis. *Mbio*, **12**, e02060-21.
7. Karch,H., Tarr,P.I. and Bielaszewska,M. (2005) Enterohaemorrhagic *Escherichia coli* in human medicine. *Int. J. Med. Microbiol.*, **295**, 405–418.
8. Makino,K., Ishii,K., Yasunaga,T., Hattori,M., Yokoyama,K., Yutsudo,C.H., Kubota,Y., Yamaichi,Y., Iida,T., Yamamoto,K. *et al.* (1998) Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. *DNA Res.*, **5**, 1–9.
9. Ooka,T., Ogura,Y., Asadulghani,M., Ohnishi,M., Nakayama,K., Terajima,J., Watanabe,H. and Hayashi,T. (2009) Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. *Genome Res.*, **19**, 1809–1816.
10. Kusumoto,M., Ooka,T., Nishiya,Y., Ogura,Y., Saito,T., Sekine,Y., Iwata,T., Akiba,M. and Hayashi,T. (2011) Insertion sequence-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions. *Nat. Commun.*, **2**, 152.
11. Nakamura,K., Seto,K., Isobe,J., Taniguchi,I., Gotoh,Y. and Hayashi,T. (2022) Insertion sequence (IS)-excision enhancer (IEE)-Mediated IS excision from the lacZ gene restores the lactose utilization defect of shiga toxin-producing *Escherichia coli* O121:H19 strains and is responsible for their delayed lactose utilization phenotype. *Appl. Environ. Microbiol.*, **88**, e00760-22.
12. Toro,M., Rump,L.V., Cao,G., Meng,J., Brown,E.W. and Gonzalez-Escalona,N. (2015) Simultaneous presence of insertion sequence excision enhancer and insertion sequence IS629 correlates with increased diversity and virulence in shiga toxin-producing *Escherichia coli*. *J. Clin. Microbiol.*, **53**, 3466–3473.
13. Kanai,Y., Tsuru,S. and Furusawa,C. (2022) Experimental demonstration of operon formation catalyzed by insertion sequence. *Nucleic Acids Res.*, **50**, 1673–1686.
14. Iyer,L.M., Koonin,E.V., Leipe,D.D. and Aravind,L. (2005) Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res.*, **33**, 3875–3896.
15. Blanco,L., Calvo,P.A., Díaz-Talavera,A., Carvalho,G., Calero,N., Martínez-Carrón,A., Velázquez-Ruiz,C., Villadangos,S., Guerra,S. and Martínez-Jiménez,M.I. (2019) Chapter nine - mechanism of DNA primer synthesis by human primpol. In: Zhao,L. and Kaguni,L.S. (eds). *The Enzymes, DNA Repair*. Academic Press, Vol. **45**, pp. 289–310.
16. Guilliam,T.A., Keen,B.A., Brissett,N.C. and Doherty,A.J. (2015) Primase-polymerases are a functionally diverse superfamily of replication and repair enzymes. *Nucleic Acids Res.*, **43**, 6651–6664.
17. Hyatt,D., Chen,G.-L., LoCascio,P.F., Land,M.L., Larimer,F.W. and Hauser,L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.*, **11**, 119.
18. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinf.*, **10**, 421.
19. Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
20. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
21. Zhou,L., Feng,T., Xu,S., Gao,F., Lam,T.T., Wang,Q., Wu,T., Huang,H., Zhan,L., Li,L. *et al.* (2022) ggmsa: a visual exploration tool for multiple sequence alignment and associated data. *Brief. Bioinform.*, **23**, bbac222.
22. Bessonov,K., Laing,C., Robertson,J., Yong,I., Ziebell,K., Gannon,V.P.J., Nichani,A., Arya,G., Nash,J.H.E. and Christianson,S. (2021) ECTyper: in silico *Escherichia coli* serotype and species prediction from raw and assembled whole-genome sequence data. *Microb. Genomics*, **7**, 000728.
23. Waters,N.R., Abram,F., Brennan,F., Holmes,A. and Pritchard,L. (2020) Easy phylotyping of *Escherichia coli* via the EzClermont web app and command-line tool. *Access Microbiol.*, **2**, acmi000143.
24. Baños,B., Lázaro,J.M., Villar,L., Salas,M. and de Vega,M. (2008) Characterization of a *Bacillus subtilis* 64-kDa DNA polymerase X potentially involved in DNA repair. *J. Mol. Biol.*, **384**, 1019–1028.
25. Lázaro,J.M., Blanco,L. and Salas,M. (1995) Purification of bacteriophage ϕ29 DNA polymerase. *Methods Enzym.*, **262**, 42–49.
26. Perna,N.T., Plunkett,G., Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A. *et al.* (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature*, **409**, 529–533.
27. Lobstein,J., Emrich,C.A., Jeans,C., Faulkner,M., Riggs,P. and Berkmen,M. (2012) SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Factories*, **11**, 56.
28. Cui,S., Klima,R., Ochem,A., Arosio,D., Falaschi,A. and Vindigni,A. (2003) Characterization of the DNA-unwinding activity of human RECQ1, a helicase specifically stimulated by human replication protein A. *J. Biol. Chem.*, **278**, 1424–1432.
29. Tuteja,N., Rahman,K., Tuteja,R. and Falaschi,A. (1991) DNA helicase IV from Hela cells. *Nucleic Acids Res.*, **19**, 3613–3618.
30. The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.
31. Baek,M., DiMaio,F., Anishchenko,I., Dauparas,J., Ovchinnikov,S., Lee,G.R., Wang,J., Cong,Q., Kinch,L.N., Schaeffer,R.D. *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, **373**, 871–876.
32. Augustin,M.A., Huber,R. and Kaiser,J.T. (2001) Crystal structure of a DNA-dependent RNA polymerase (DNA primase). *Nat. Struct. Biol.*, **8**, 57–61.
33. Kazlauskas,D., Sezonov,G., Charpin,N., Venclovas,C., Forterre,P. and Krupovic,M. (2018) Novel families of archaeo-eukaryotic primases associated with mobile genetic elements of bacteria and archaea. *J. Mol. Biol.*, **430**, 737–750.

34. Lao-Sirieix,S.-H., Nookala,R.K., Roversi,P., Bell,S.D. and Pellegrini,L. (2005) Structure of the heterodimeric core primase. *Nat. Struct. Mol. Biol.*, **12**, 1137–1144.
35. Lipps,G. (2011) Structure and function of the primase domain of the replication protein from the archaeal plasmid pRN1. *Biochem. Soc. Trans.*, **39**, 104–106.
36. Pitcher,R.S., Brissett,N.C., Picher,A.J., Andrade,P., Juárez,R., Thompson,D., Fox,G.C., Blanco,L. and Doherty,A.J. (2007) Structure and function of a mycobacterial NHEJ DNA repair polymerase. *J. Mol. Biol.*, **366**, 391–405.
37. Brissett,N.C., Zabrady,K., Płociński,P., Bianchi,J., Korycka-Machała,M., Brzostek,A., Dziadek,J. and Doherty,A.J. (2020) Molecular basis for DNA repair synthesis on short gaps by mycobacterial primase-polymerase C. *Nat. Commun.*, **11**, 4196.
38. Brissett,N.C., Martin,M.J., Pitcher,R.S., Bianchi,J., Juarez,R., Green,A.J., Fox,G.C., Blanco,L. and Doherty,A.J. (2011) Structure of a preternary complex involving a prokaryotic NHEJ DNA polymerase. *Mol. Cell*, **41**, 221–231.
39. Zhu,H., Nandakumar,J., Aniukwu,J., Wang,L.K., Glickman,M.S., Lima,C.D. and Shuman,S. (2006) Atomic structure and nonhomologous end-joining function of the polymerase component of bacterial DNA ligase D. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 1711–1716.
40. Sánchez-Salvador,A. and de Vega,M. (2020) Structural determinants responsible for the preferential insertion of ribonucleotides by bacterial NHEJ poldom. *Biomolecules*, **10**, 203.
41. Story,R.M., Weber,I.T. and Steitz,T.A. (1992) The structure of the *E. coli* recA protein monomer and polymer. *Nature*, **355**, 318–325.
42. Byrd,A.K. and Raney,K.D. (2012) Superfamily 2 helicases. *Front. Biosci. Landmark Ed.*, **17**, 2070–2088.
43. Hausmann,S., Geiser,J., Vadas,O., Ducret,V., Perron,K. and Valentini,M. (2020) Auxiliary domains of the HrpB bacterial DExH-box helicase shape its RNA preferences. *RNA Biol.*, **17**, 637–650.
44. Ozdemir,A.Y., Rusanov,T., Kent,T., Siddique,L.A. and Pomerantz,R.T. (2018) Polymerase θ-helicase efficiently unwinds DNA and RNA-DNA hybrids. *J. Biol. Chem.*, **293**, 5259–5269.
45. Singleton,M.R., Dillingham,M.S. and Wigley,D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, **76**, 23–50.
46. Tenaillon,O., Skurnik,D., Picard,B. and Denamur,E. (2010) The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.*, **8**, 207–217.
47. Clermont,O., Condamine,B., Dion,S., Gordon,D.M. and Denamur,E. (2021) The E phylogroup of *Escherichia coli* is highly diverse and mimics the whole *E. coli* species population structure. *Environ. Microbiol.*, **23**, 7139–7151.
48. Denamur,E., Clermont,O., Bonacorsi,S. and Gordon,D. (2021) The population genetics of pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.*, **19**, 37–54.
49. Picard,B., Garcia,J.S., Gouriou,S., Duriez,P., Brahimi,N., Bingen,E., Elion,J. and Denamur,E. (1999) The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect. Immun.*, **67**, 546–553.
50. Liu,L., Komori,K., Ishino,S., Bocquier,A.A., Cann,I.K.O., Kohda,D. and Ishino,Y. (2001) The archaeal DNA primase: biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus*. *J. Biol. Chem.*, **276**, 45484–45490.
51. García-Gómez,S., Reyes,A., Martínez-Jiménez,M.I., Chocrón,E.S., Mourón,S., Terrados,G., Powell,C., Salido,E., Méndez,J., Holt,I.J. *et al.* (2013) PrimPol, an archaic primase/polymerase operating in human cells. *Mol. Cell*, **52**, 541–553.
52. Díaz-Talavera,A., Calvo,P.A., González-Acosta,D., Díaz,M., Sastre-Moreno,G., Blanco-Franco,L., Guerra,S., Martínez-Jiménez,M.I., Méndez,J. and Blanco,L. (2019) A cancer-associated point mutation disables the steric gate of human primpol. *Sci. Rep.*, **9**, 1121.
53. Lusk,J.E., Williams,R.J.P. and Kennedy,E.P. (1968) Magnesium and the growth of *Escherichia coli*. *J. Biol. Chem.*, **243**, 2618–2624.
54. Kehres,D.G. and Maguire,M.E. (2003) Emerging themes in manganese transport, biochemistry and pathogenesis in bacteria. *FEMS Microbiol. Rev.*, **27**, 263–290.
55. Blanca,G., Shevelev,I., Ramadan,K., Villani,G., Spadari,S., Hübscher,U. and Maga,G. (2003) Human DNA polymerase lambda diverged in evolution from DNA polymerase beta toward specific Mn(++) dependence: a kinetic and thermodynamic study. *Biochemistry*, **42**, 7467–7476.
56. García-Díaz,M., Bebenek,K., Krahn,J.M., Pedersen,L.C. and Kunkel,T.A. (2007) Role of the catalytic metal during polymerization by DNA polymerase lambda. *DNA Repair (Amst.)*, **6**, 1333–1340.
57. Domínguez,O., Ruiz,J.F., Lain de Lera,T., García-Díaz,M., González,M.A., Kirchhoff,T., Martínez,A.C., Bernad,A. and Blanco,L. (2000) DNA polymerase mu (Polμ), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J.*, **19**, 1731–1742.
58. Brissett,N.C., Martín,M.J., Bartlett,E.J., Bianchi,J., Blanco,L. and Doherty,A.J. (2013) Molecular basis for DNA double-strand break annealing and primer extension by an NHEJ DNA polymerase. *Cell Rep.*, **5**, 1108–1120.
59. Gong,C., Martins,A., Bongiorno,P., Glickman,M. and Shuman,S. (2004) Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *J. Biol. Chem.*, **279**, 20594–20606.
60. de Vega,M. (2013) The minimal *Bacillus subtilis* nonhomologous end joining repair machinery. *PLoS One*, **8**, e64232.
61. Zhu,H. and Shuman,S. (2007) Characterization of *Agrobacterium tumefaciens* DNA ligases c and D. *Nucleic Acids Res.*, **35**, 3631–3645.
62. Martínez-Jiménez,M.I., García-Gómez,S., Bebenek,K., Sastre-Moreno,G., Calvo,P.A., Díaz-Talavera,A., Kunkel,T.A. and Blanco,L. (2015) Alternative solutions and new scenarios for translesion DNA synthesis by human primpol. *DNA Repair (Amst.)*, **29**, 127–138.
63. Frank,E.G. and Woodgate,R. (2007) Increased catalytic activity and altered fidelity of human DNA polymerase ι in the presence of manganese. *J. Biol. Chem.*, **282**, 24689–24696.
64. Yang,W., Lee,J.Y. and Nowotny,M. (2006) Making and breaking nucleic acids: two-Mg^{2+}-ion catalysis and substrate specificity. *Mol. Cell*, **22**, 5–13.
65. Choi,J.-Y., Patra,A., Yeom,M., Lee,Y.-S., Zhang,Q., Egli,M. and Guengerich,F.P. (2016) Kinetic and structural impact of metal ions and genetic variations on human DNA polymerase ι. *J. Biol. Chem.*, **291**, 21063–21073.
66. Aniukwu,J., Glickman,M.S. and Shuman,S. (2008) The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev.*, **22**, 512–527.
67. Gong,C., Bongiorno,P., Martins,A., Stephanou,N.C., Zhu,H., Shuman,S. and Glickman,M.S. (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat. Struct. Mol. Biol.*, **12**, 304–312.
68. Yakovleva,L. and Shuman,S. (2006) Nucleotide misincorporation, 3′-Mismatch extension, and responses to abasic sites and DNA adducts by the polymerase component of bacterial DNA ligase D. *J. Biol. Chem.*, **281**, 25026–25040.
69. Zhu,H. and Shuman,S. (2005) A Primer-dependent polymerase function of *Pseudomonas aeruginosa* ATP-dependent DNA ligase (LigD). *J. Biol. Chem.*, **280**, 418–427.
70. Kent,T., Chandramouly,G., McDevitt,S.M., Ozdemir,A.Y. and Pomerantz,R.T. (2015) Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nat. Struct. Mol. Biol.*, **22**, 230–237.
71. Jung,H., Lee,J.A., Choi,S., Lee,H. and Ahn,B. (2014) Characterization of the *Caenorhabditis elegans* HIM-6/BLM helicase: unwinding recombination intermediates. *PLoS One*, **9**, e102402.
72. Kusano,K., Berres,M.E. and Engels,W.R. (1999) Evolution of the RECQ family of helicases: a *Drosophila* homolog, Dmblm, is similar to the human bloom syndrome gene. *Genetics*, **151**, 1027–1039.
73. Masai,H., Deneke,J., Furui,Y., Tanaka,T. and Arai,K.-I. (1999) *Escherichia coli* and *Bacillus subtilis* PriA proteins essential for recombination-dependent DNA replication: involvement of ATPase/helicase activity of PriA for inducible stable DNA replication. *Biochimie*, **81**, 847–857.
74. Walsh,B.W., Bolz,S.A., Wessel,S.R., Schroeder,J.W., Keck,J.L. and Simmons,L.A. (2014) RecD2 helicase limits replication fork stress in *Bacillus subtilis*. *J. Bacteriol.*, **196**, 1359–1368.
75. Wang,J. and Julin,D.A. (2004) DNA helicase activity of the RecD protein from *Deinococcus radiodurans*. *J. Biol. Chem.*, **279**, 52024–52032.

76. Dürr,H., Flaus,A., Owen-Hughes,T. and Hopfner,K.-P. (2006) Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. *Nucleic Acids Res.*, **34**, 4160–4167.
77. Jankowsky,E. (2011) RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.*, **36**, 19–29.
78. Jankowsky,E. and Fairman,M.E. (2007) RNA helicases — one fold for many functions. *Curr. Opin. Struct. Biol.*, **17**, 316–324.
79. Pyle,A.M. (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.*, **37**, 317–336.
80. Hargreaves,D.C. and Crabtree,G.R. (2011) ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.*, **21**, 396–420.
81. Jankowsky,E. and Bowers,H. (2006) Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res.*, **34**, 4181–4188.
82. Lorenz,S.C., Gonzalez-Escalona,N., Kotewicz,M.L., Fischer,M. and Kase,J.A. (2017) Genome sequencing and comparative genomics of enterohemorrhagic *Escherichia coli* O145:H25 and O145:H28 reveal distinct evolutionary paths and marked variations in traits associated with virulence & colonization. *BMC Microbiol.*, **17**, 183.
83. Skipper,K.A., Andersen,P.R., Sharma,N. and Mikkelsen,J.G. (2013) DNA transposon-based gene vehicles - scenes from an evolutionary drive. *J. Biomed. Sci.*, **20**, 92.
84. Martínez-Jiménez,M.I., Calvo,P.A., García-Gómez,S., Guerra-González,S. and Blanco,L. (2018) The Zn-finger domain of human primpol is required to stabilize the initiating nucleotide during DNA priming. *Nucleic Acids Res.*, **46**, 4138–4151.
85. Chen,S.H., Byrne,R.T., Wood,E.A. and Cox,M.M. (2015) *Escherichia coli* radD (yejH) gene: a novel function involved in radiation resistance and double-strand break repair. *Mol. Microbiol.*, **95**, 754–768.