4914–4926 *Nucleic Acids Research, 2007, Vol. 35, No. 14* *Published online 10 July 2007*
doi:10.1093/nar/gkm516

Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease

Kostas Repanas[1], Nora Zingler[2], Liliana E. Layer[2], Gerald G. Schumann[2], Anastassis Perrakis[1] and Oliver Weichenrieder[1,*]

[1]Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands and [2]Paul-Ehrlich-Institut, Section PR2/Retroelements, 63225 Langen, Germany

Received April 12, 2007; Revised May 18, 2007; Accepted June 15, 2007

ABSTRACT

The human LINE-1 endonuclease (L1-EN) is the targeting endonuclease encoded by the human LINE-1 (L1) retrotransposon. L1-EN guides the genomic integration of new L1 and Alu elements that presently account for ~28% of the human genome. L1-EN bears considerable technological interest, because its target selectivity may ultimately be engineered to allow the site-specific integration of DNA into defined genomic locations. Based on the crystal structure, we generated L1-EN mutants to analyze and manipulate DNA target site recognition. Crystal structures and their dynamic and functional analysis show entire loop grafts to be feasible, resulting in altered specificity, while individual point mutations do not change the nicking pattern of L1-EN. Structural parameters of the DNA target seem more important for recognition than the nucleotide sequence, and nicking profiles on DNA oligonucleotides *in vitro* are less well defined than the respective integration site consensus *in vivo*. This suggests that additional factors other than the DNA nicking specificity of L1-EN contribute to the targeted integration of non-LTR retrotransposons.

INTRODUCTION

In the higher eukaryotes frequently more than 90% of the DNA does not code for functional proteins or RNA. Much of this DNA has originated from the action of mobile genetic elements, mostly retrotransposons that propagate in a copy-and-paste mechanism via an RNA intermediate. While these elements can be viewed as molecular parasites that are in an evolutionary race with their host genome, they can also be regarded as essential genomic components for slowly reproducing species to adapt to a changing environment. They generate allelic heterogeneity and create new possibilities for genetic recombination, increasing genomic fluidity (1–5).

Mobile genetic elements integrate into new genomic locations in two fundamentally different ways. DNA transposons and retrotransposons with long terminal repeats (LTR retrotransposons) use a transposase/integrase to insert a double-stranded DNA copy of the element at the target site. In this case, no DNA synthesis takes place at the site of integration. In contrast, non-LTR retrotransposons use a mechanism called target-primed reverse transcription (3). This process is initiated by a targeting endonuclease, which specifically binds to the site of genomic integration. It nicks one strand of the DNA and creates a free 3′ hydroxyl end, which is then used as a primer for reverse transcription of the retrotransposon RNA at the site of integration. Endonuclease and reverse transcriptase are two domains of a single retrotransposon-encoded protein. They are thought to rely on the assistance of 'host'-encoded proteins to complete the integration process (6–8).

Most non-LTR retrotransposons are APE-type non-LTR retrotransposons (9). Their targeting endonuclease belongs to a family of metal-dependent phosphohydrolases that includes nucleases like DNaseI (PDB-ID: 1dnk), APE1 (PDB-ID: 1dew), Exo III (PDB-ID: 1ako) and CdtB (PDB-ID: 1sr4) but also sugar phosphatases like I5PP (PDB-ID: 1i9z) and phospholipases like SmcL (PDB-ID: 1zwx) and Bc-SMase (PDB-ID: 2ddt). Members of this family share the same protein scaffold and the same catalytic residues, but a variation of the connecting surface loops has allowed them to develop quite diverse substrate specificities (10).

Under the pressure to survive in their respective host species non-LTR retrotransposons have evolved different strategies (9). Stringent elements like R1Bm from *Bombyx mori* (11) and Tx1L from *Xenopus laevis* (12) encode highly specific targeting endonucleases (13,14).

*To whom correspondence should be addressed. Tel: +4970716011358; Fax: +4970716011353; Email: oliver.weichenrieder@tuebingen.mpg.de
Present addressess:
Nora Zingler, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA
Oliver Weichenrieder, Max-Planck-Institute for Developmental Biology, Department of Biochemistry, 72076 Tuebingen, Germany

They integrate into unique genomic locations (a specific sequence within 28S rDNA for R1Bm or within the apparent DNA transposon Tx1D for Tx1L) where they do very little or no damage to the host. Promiscuous elements like the human LINE-1 (L1) element (15) may integrate into several hundred thousand genomic locations. They have a rather short integration-site consensus [5′-TTTT/AA-3′ for L1 (16–18)] that is nicked by the respective targeting endonuclease (19,20). The host limits the spread of such elements by transcriptional and post-transcriptional silencing mechanisms that reduce activity to tolerable levels (21–24).

Clearly, the respective endonucleases play a major role in target site selection (13,14,19,25). The intriguing question of how different targeting endonucleases recognize the DNA substrate and how easily new specificities can arise in the course of evolution remains open. There are indications that retrotransposons can evolve back and forth between a stringent and a promiscuous mode-of-action (26) and the ability to manipulate and design target specificity would be a crucial step in converting non-LTR retrotransposons into a genetic tool.

Previously, we described the crystal structure of the human L1 endonuclease (L1-EN) (27). Based on structure comparisons and sequence alignments we suggested that the prominent βB6–βB5 hairpin loop may insert into the DNA minor groove and may be particularly important for recognizing the DNA target. Here, we combine a mutational approach (specific point mutants and entire loop grafts) with structural and dynamic analyses. We determine minimal size and structural features of the DNA target and we show that size and flexibility of the βB6–βB5 hairpin loop are crucial for activity. Variation of the loop sequence results in an altered DNA nicking profile including novel sites. This indicates that the engineering of novel specificities may ultimately be feasible.

MATERIALS AND METHODS

Preparation and purification of L1-EN variants

Point mutants and loop variants of L1-EN were generated in the context of the retrotransposition reporter plasmid pCEP4/L1.3/ColE1/*mneoI*$_{400}$ (16) as described in the Supplementary Data. For the expression of mutated L1-EN domains we PCR-amplified DNA corresponding to residues 1–239 of wild-type L1-EN (27) from the respective retrotransposition reporter plasmids and inserted the products into the NcoI/XhoI cloning sites of expression plasmid pETM11 (28). Proteins with N-terminal poly-histidine tags were overexpressed in *Escherichia coli* Rosetta II cells (Novagen) and purified over Ni-chelating chromatography and heparin affinity columns. Protein was quantified spectroscopically or on denaturing SDS polyacrylamide gels. For Tx1L-EN protein (residues 1–239) DNA was amplified from plasmid pE1EN (13) using primers Tx1L-EN-N1 and Tx1L-EN-C239. For crystallization the respective proteins were expressed and purified without tag as described in (27). Purified protein (>1 mg/ml) was stored frozen at −80°C at NaCl concentrations above 300 mM.

Retrotransposition reporter assay

Retrotransposition frequencies of wild-type and mutant L1 constructs were determined by applying the rapid, quantitative transient L1 retrotransposition assay described previously (29). HeLa cells (2×10^5) were plated in each well of a six-well dish and grown to 50–80% confluency in DMEM. The following day, triplicate dishes were transfected using 6 μl Fugene-6 transfection reagent (Roche) and 2 μg of a Qiagen plasmid DNA preparation per well. At 24-h post-transfection, the transfection mixture was removed and replaced by DMEM. At 72-h post-transfection, the medium was replaced with DMEM containing 400 μg/ml G418. After 10–14 days, G418R colonies were stained with Giemsa solution and counted. The recovery of integrated L1 elements for sequencing is described in (16).

Plasmid nicking

Supercoiled pBluescript plasmid DNA was prepared from *E. coli* DH5α cells. Closed circle plasmid DNA was obtained by simultaneous digestion and re-ligation of supercoiled DNA (15 μg/ml) with 5 U/ml HindIII and 900 U/ml T4 DNA ligase resulting in only trace amounts of dimeric product. DNA was quantified after linearization on agarose gels containing ethidium bromide. Nicking reactions (10 or 60 μl) were done in single tubes or 96-well trays in 20 mM Na-HEPES (pH = 7.5), 100 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/ml bovine serum albumin (BSA) and 4 mM dithiothreitol. Final concentrations were 2 nM DNA (3.6 μg/ml) and 2–128 nM protein, which had been previously diluted in protein buffer (20 mM Na-HEPES (pH = 7.5), 300 mM NaCl, 10 mM $MgCl_2$, 0.3 mg/ml BSA and 10 mM dithiothreitol). After 30 min at 37°C, reactions were stopped by the addition of DNA loading buffer containing EDTA (17 mM final). Reaction products were separated on 1.0 or 1.4% agarose gels (0.5 × TBE) containing 0.5 μg/ml ethidium bromide. Nicking activity was quantified by densitometry, determining the fraction of supercoiled plasmid DNA converted to the open circle form.

Oligonucleotide nicking

Gel-purified synthetic oligonucleotides were labeled at the 5′ end with radioactive phosphate (^{32}P) using [γ-^{32}P]ATP and T4 polynucleotide kinase and were re-purified on a gel. Equimolar amounts (450 nM) of unlabeled complementary and substrate strands were mixed with a trace amount of labeled substrate. The mixture was annealed in 5 mM Na-HEPES (pH = 7.5) by heating to 90°C and slow-cooled to room temperature. After testing various pH and salt conditions nicking reactions (50 μl) were done in 50 mM Na-HEPES (pH = 6.5), 150 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/ml BSA and 1 mM dithiothreitol. Final concentrations were 180 nM DNA (0.5–7.5 μg/ml) and 20–2000 nM protein, which had been previously diluted in protein buffer [5 mM Na-HEPES (pH = 7.5), 300 mM NaCl, 10 mM $MgCl_2$, 0.5 mg/ml BSA and 5 mM

dithiothreitol]. After 30 min at 37°C, reactions were stopped by the addition of 175 µl of 380 mM Na-acetate (pH = 7.5), followed by phenol extraction and ethanol precipitation. Reaction products were separated on 10% denaturing polyacrylamide gels and quantified in a phosphoimager. The intensity of each band was converted into the relative abundance of each nicking site (Supplementary Table 1) and used to generate sequence logos (http://ep.ebi.ac.uk/EP/SEQLOGO).

Crystallization

Untagged LTx [20 mM Na-HEPES (pH = 7.0), 200 mM NaCl] was concentrated to 15 mg/ml. Sitting drops (200 nl protein plus 200 nl reservoir solution) were set up at room temperature using a Mosquito robot. Single crystals appeared over night from a reservoir (75 µl) containing 160 mM $MgCl_2$, 370 mM $(NH_4)_2SO_4$ and 33.8% PEG 6000. Untagged LR1 [20 mM Na-HEPES (pH = 7.0), 200 mM NaCl] was concentrated to 10 mg/ml. Hanging drops (2 µl protein plus 2 µl reservoir solution) were set up manually at 4°C. Crystals appeared after several days over a reservoir (500 µl) containing 10 mM $MnSO_4$, 200 mM $(NH_4)_2SO_4$ and 31% PEG 1000. Hair seeding improved reproducibility significantly. In both cases, crystals were transferred to a cryo-solution containing 15% glycerol (mixing reservoir and 80% glycerol stock solution) and flash-frozen in liquid nitrogen.

Data collection and structure solution

Diffraction data were collected at beamline ID23-1 at the European Synchrotron Radiation Facility in Grenoble, France. Diffraction images were processed by MOSFLM (30) and SCALA (31). The structures were solved by molecular replacement using MOLREP (32) with L1-EN (PDB ID: 1vyb) as search model. Automatic model building was done with ARP/wARP (33) to a completeness of 90% for LTx and 98% of LR1. Models were completed manually and structures were refined using REFMAC (34) and COOT (35) iteratively.

Normal mode analysis (NMA)

For Normal mode analysis, the PDB files of L1-EN, LTx, LR1 and TRAS1-EN were provided to the web-based server WEBnm@ (http://www.bioinfo.no/tools/normalmodes) following the standard protocol to calculate and analyze the first six vibrational modes (36).

RESULTS

The crystal structure of L1-EN suggests elements important for DNA target recognition but not for catalysis

We designed variants of L1-EN that fall into three categories (Figure 1). The first category includes point mutations (D145A, T192V, H230A) of catalytic and structurally important residues that are highly conserved within the entire enzyme family. The second category comprises point mutants (R155A, S202A, I204Y) of moderately conserved non-catalytic surface residues expected to affect the accommodation and recognition

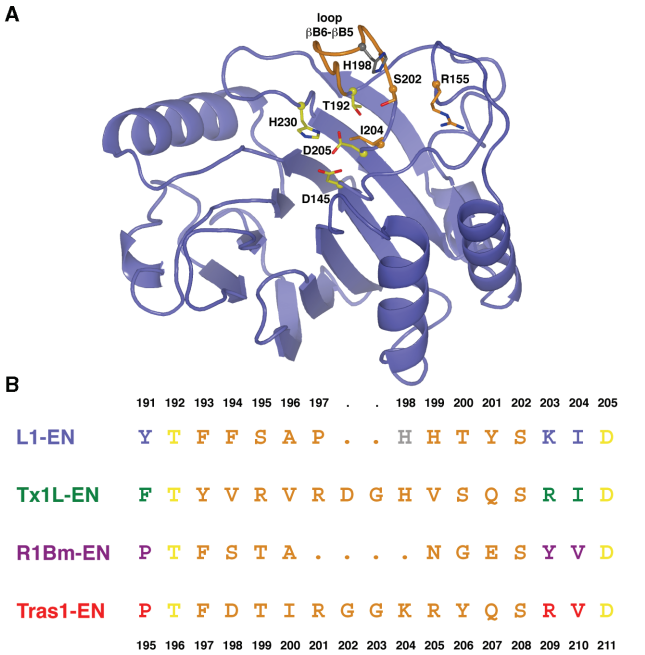


Figure 1. L1-EN point mutants and βB6–βB5 hairpin loop variants. (**A**) Localization of the mutations on the crystal structure of L1-EN. The structure of L1-EN (27) is drawn as ribbons with the backbone of the exchanged loop in orange and with individual point mutants as balls-and-sticks. Yellow: conserved residues, orange: residues potentially contacting DNA, gray: H198. (**B**) Structure-based alignment of the βB5–βB6 hairpin loop. For the chimeric endonucleases LTx and LR1, the respective loop sequences (orange) of Tx1L-EN and R1Bm-EN were grafted onto the L1-EN scaffold between the conserved anchoring residues T192 and S202. For L3G, the loop was replaced by three glycines. The loop sequence of TRAS1-EN is shown for comparison. Numbering is from L1-EN (top, PDB-ID: 1vyb) and TRAS1-EN (bottom, PDB-ID: 1wdu) and color-coding of endonucleases is maintained throughout the article.

of the nucleotide downstream of the scissile bond (Figure 1A and B).

In the third category of L1-EN variants we manipulated the βB6–βB5 hairpin loop, which is positioned to insert into the DNA minor groove with the possibility to read out both sequence and structural parameters (27). It is well suited for a loop-grafting experiment because the anchoring residues T192 and S202 on either side are well conserved among many metal-dependent phosphohydrolases. Therefore, we replaced the entire βB6–βB5 hairpin loop of L1-EN with the corresponding sequences from the R1Bm and Tx1L retrotransposons (Figure 1B). The resulting mutants LR1 and LTx, respectively, were accompanied by the loop deletion variant L3G, where we exchanged the entire loop (including S202) for a linker of three glycines.

L1-EN point mutations and loop grafts affect retrotransposition in cell culture

Initially, the L1-EN variants were tested in the context of a functional, tagged L1 element in a well-established cell culture assay (29,37). We scored successful retrotransposition events by the appearance of G418-resistant HeLa cell colonies, subtracting background activity

Table 1. Comparison of retrotransposition frequencies *in vivo* and plasmid nicking activities *in vitro*

L1-EN variant	Retrotransposition frequency[a], %	Plasmid nicking activity[b], %
wt	100 ± 17.1	100 ± 0.8
LTx	21 ± 2.4	29 ± 2.6
LR1	2 ± 2.3	6 ± 0.8
L3G	0 ± 2.2	10 ± 1.8
D145A	0[c]	3 ± 1.0
R155A	12 ± 3.3	19 ± 3.4
T192V	5 ± 3.0	–
S202A	32 ± 7.8	28 ± 2.2
I204Y	1 ± 1.1	4 ± 1.2
H230A	0	–

[a]Corrected for background activity (≤5%); for details see Supplementary Data.
[b]Normalized to L1-EN (wt) activity, (–) not analyzed.
[c]As a D145A/N147A double mutant.

caused by *trans* complementation or endonuclease-independent retrotransposition.

All variants reduce the frequency of retrotransposition significantly, confirming the relevance of the mutated elements (Table 1). The strongest effects are seen with point mutants D145A, T192V, I204Y and H230A and with loop variants LR1 and L3G. To test whether this is directly related to the ability of the enzyme to recognize and nick target DNA we purified the respective L1-EN variants for assays *in vitro*.

The ability of L1-EN variants to nick plasmid DNA correlates well with the frequency of retrotransposition

Residues T192 and H230 are hydrogen-bonded via D205 (27). These interactions are apparently essential for the structural integrity of L1-EN as the respective mutants were inherently unstable, degraded easily or precipitated rapidly. From the first category only the D145A mutant could be purified as a negative control for catalytic activity.

The purified L1-EN variants were first analyzed in a plasmid DNA nicking assay (20), where supercoiled plasmid is converted into the open circle form that runs considerably slower on an agarose gel (Figure 2). Figure 2A shows a side-by-side comparison of the activities of all L1-EN mutants (32 nM) on 2 nM supercoiled plasmid. Under these conditions wild-type L1-EN converts 95% of plasmid DNA into the open circle form. The three point mutants, R155A, S202A and I204Y, show strongly reduced activity, with S202A being affected the least and I204Y the most. The strong effect of I204Y suggests that L1-EN probably binds double-stranded DNA in an orientation that differs from the one seen in the complex with DNaseI (38), because in DNaseI the tyrosine is present and tolerated at this position. This view is supported by the effects of S202A and R155A, which indicate that these moderately conserved amino acids are indeed involved in contacting the nucleotide(s) downstream of the scissile bond, either specifically or non-specifically. For a direct contact with R155A the downstream DNA would have to be distorted or even

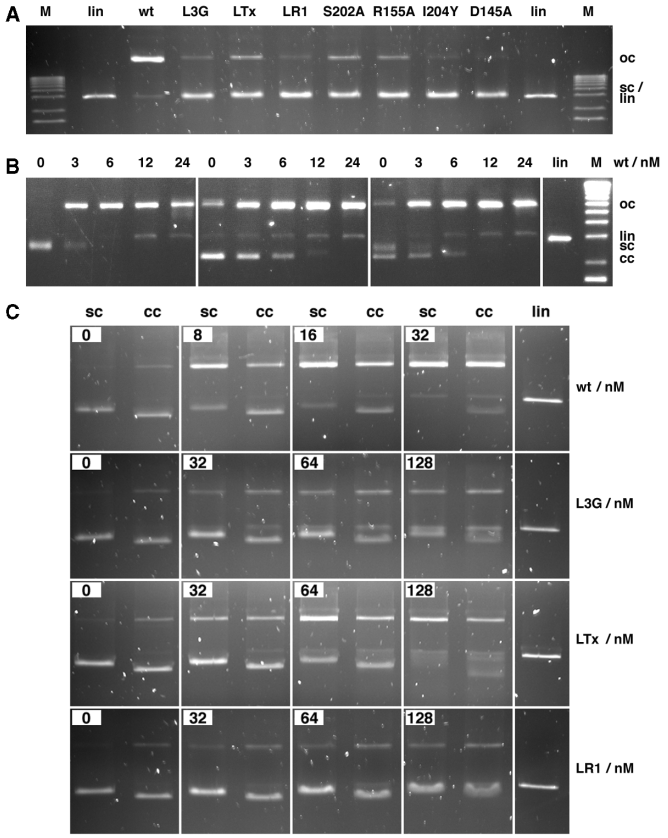


Figure 2. Plasmid nicking activity of L1-EN variants. Experiments were done with wild-type L1-EN (wt), βB5–βB6 hairpin loop variants (L3G, LTx, LR1) and point mutants (S202A, R155A, I204Y, D145A). Supercoiled (sc) plasmid DNA (pBluescript) or relaxed closed circle DNA (cc) was converted into the open circle form (oc) and eventually into linear DNA (lin). Closed circle DNA contains trace amounts of dimer, which runs like open circle DNA both on 1.4% agarose gels (A and C) and 1.0% gels (B). M: DNA size marker X (Roche). (**A**) Relative activity of L1-EN mutants (32 nM) on supercoiled plasmid DNA (2 nM). (**B**) Preference of L1-EN for supercoiled target DNA. Supercoiled and closed circle DNA was nicked by increasing concentrations of L1-EN, either separately (2 nM) or in mutual competition (1 nM each). (**C**) Titration of L1-EN hairpin loop variants and selectivity for supercoiled DNA. LR1 and L3G are significantly less active than LTx and show no more preference for supercoiled DNA.

flipped as in the complex with APE1 (39). Among the loop variants, LTx remains most active, at levels similar to the S202A point mutant. In contrast, LR1 and L3G retain little but still detectable activity (Table 1).

The structural context of the DNA target is important for its recognition by L1-EN (19). When presented with equal amounts of supercoiled and of relaxed, closed circle pBluescript DNA, L1-EN nicks the supercoiled DNA much more efficiently (Figure 2B). Since the βB6–βB5 hairpin loop may well be involved in the recognition of an unusual DNA structure caused by supercoiling, we tested the L1-EN loop variants also in this respect (Figure 2C). While LTx still prefers supercoiled DNA, the very inefficient LR1 shows no detectable preference for supercoiled DNA anymore. The same observation holds true for L3G, where the loop is deleted. This experiment shows that the βB6–βB5 hairpin loop of L1-EN may

be particularly important for reading out the structural context of a potential new retrotransposon integration site.

Finally, there is a good correlation between the nicking activities *in vitro* and the retrotransposition frequencies *in vivo*, indicating that the activity of the endonuclease is limiting over a considerable range (Table 1). Consequently, alterations in the nicking specificity of the endonuclease should lead to changes in integration specificity. To distinguish whether our mutations simply impair catalysis or indeed alter target recognition we verified *in vitro* if and how nicking specificities were affected.

Efficient DNA nicking by L1-EN requires a minimum of 5 bp upstream and 3 bp downstream of the target site

Genomic L1 pre-integration sites have been analyzed statistically and a consensus sequence has been reconstructed. In the 5′ to 3′ direction the substrate strand consists of an upstream tract of four to five strongly conserved thymidines (T-tract) followed downstream by two more moderately conserved adenines, with the integration occurring at the poly(T)-A junction (16–18). In contrast to previous approaches (19) we chose this type of asymmetric target for a DNA oligonucleotide nicking assay (Figure 3).

We designed a DNA duplex consisting of 14 T-A pairs, followed by two A-T pairs and a single clamp of four C-G pairs [Figure 3A (Cwt)]. We find the 5′ labeled substrate strand (the bottom strand in all figures) to be nicked throughout the entire T-tract with very similar relative frequencies and only the first five thymidines are spared. Nicking at the poly(T)-A junction is enhanced not more than 4- to 5-fold [Figure 3B (wt)]. As shown previously (19), the observed nicking patterns result from multiple independent endonucleolytic nicking events and not from a cryptic 3′ to 5′ exonuclease activity of L1-EN.

For a closer analysis of the DNA structural parameters required for efficient nicking, we manipulated the complementary DNA strand (upper strand in all figures). A mismatched adenine (A:C) in position (+1) immediately downstream of the target site diminishes the preference for the poly(T)-A junction, reducing it to the levels observed for nicking within the T-tract [Figure 3A and C (Cim)]. This suggests that, at least during the initial step of recognition of a poly(T)-A junction by L1-EN, this nucleotide position needs to be base-paired properly with an unobstructed minor groove. Mismatching the complete remainder of downstream DNA in addition to position (+1) does not cause any further reduction of nicking efficiency at the poly(T)-A junction [Figure 3A and C (C56m)]. Next, we tested to which degree the complementary strand is required downstream of the target site by deleting an increasing number of nucleotides from the 5′ end. The results show that the complementary strand needs to extend downstream by at least one nucleotide. However, for nicking at the poly(T)-A junction to be preferred over the adjacent T-tract, at least three downstream base pairs are required [Figure 3A and C (C53-, C54-, C55- and C56-)]. Upstream of the target site, L1-EN prefers at least 5 nt to be base-paired. If this is not the case, nicking is significantly reduced [Figure 3A and C (C35-)]. In summary (Figure 3D), our data suggest that preferential recognition of a poly(T)-A junction by L1-EN requires 5 nt upstream that should be base-paired at least close to the target site and 3 bp downstream which are just sufficient to form a short independent stem that does not need to stack on the upstream duplex for stability. Thus, the minor groove at the poly(T)-A junction would be flexible and could easily be widened by external strain on the DNA or simply by the insertion of the βB6–βB5 hairpin loop, pushing the downstream DNA into a position to be contacted by S202 and R155.

The protruding hairpin loop of L1-EN is crucial for recognition of the DNA target structure

The relative enzymatic activities of the three L1-EN loop variants are similar in the plasmid DNA nicking and duplex DNA nicking assays with LTx being the most active and LR1 being the least active (Figures 2C and 3B). LTx still nicks T-tract DNA, but the preference for the poly(T)-A junction has disappeared. We conclude that the βB6–βB5 loop of LTx is less well suited to recognize a poly(T)-A junction, although it does functionally replace the βB6–βB5 loop of L1-EN to a large degree. In sharp contrast, LR1 and L3G do not show any significant endonucleolytic activity, even at the highest concentrations (Figure 3B).

To extend the analysis of the respective nicking profiles we designed long DNA oligonucleotides (Dwt and Dhy) with more sequence variation (Figure 4). Dwt contains the genomic target sequences of human L1-EN, Tx1L-EN and R1Bm-EN on a single DNA duplex. This design assures that the potential target sites are present at equal concentrations and compete for the respective endonuclease under identical conditions. Dhy is identical, except that the upstream sequences of the respective target sites have been replaced by T-tracts (Figure 4A). In addition to L1-EN, we also used wild-type Tx1L-EN (wTx) as a positive control in this assay (13). Figure 4B demonstrates the difference in nicking specificity between the sequence-specific Tx1L-EN and the promiscuous L1-EN. Tx1L-EN nicks almost exclusively at the expected target site, after nucleotide 29 on Dwt and to a lesser extent on the corresponding hybrid site on Dhy. L1-EN nicks preferentially at the poly(T)-A junctions on Dwt or Dhy, but also within extended T-tracts and non-canonical sequences like after nucleotides 19 and 20. This gives rise to characteristic and reproducible nicking profiles (Figure 4C).

The nicking profile of the LTx chimera deviates from both the L1-EN and the Tx1L-EN pattern. LTx does not recognize the Tx1L integration site (nucleotide 29 on Dwt). On Dhy there is no specific nicking at this position either, despite the upstream T-tract that was introduced and expected to fit the L1-EN scaffold (Figure 4B). This suggests that one cannot simply combine and exchange upstream (L1-EN scaffold) and downstream (βB6–βB5 loop) recognition elements in a modular fashion to generate a desired target specificity. In comparison to

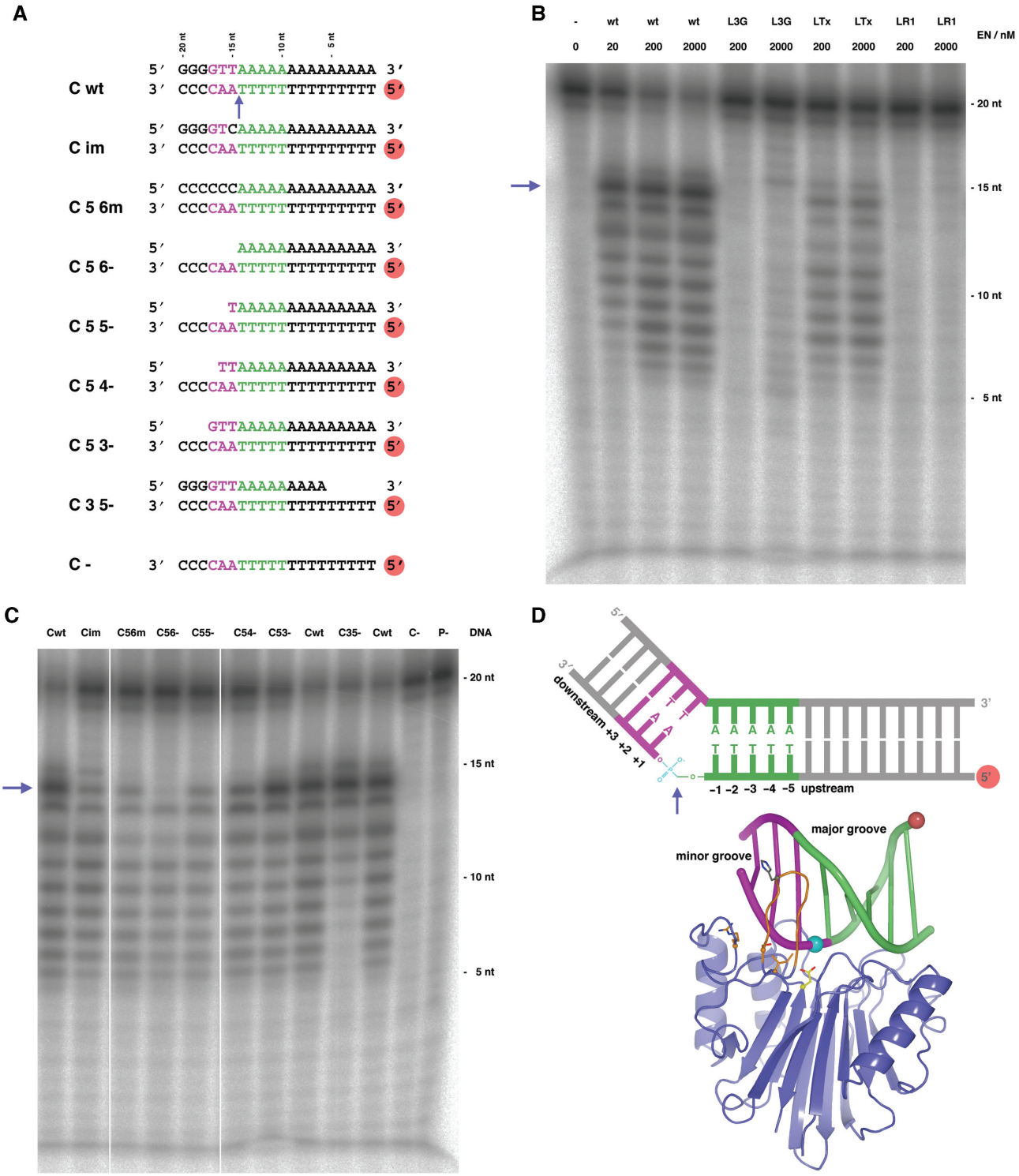
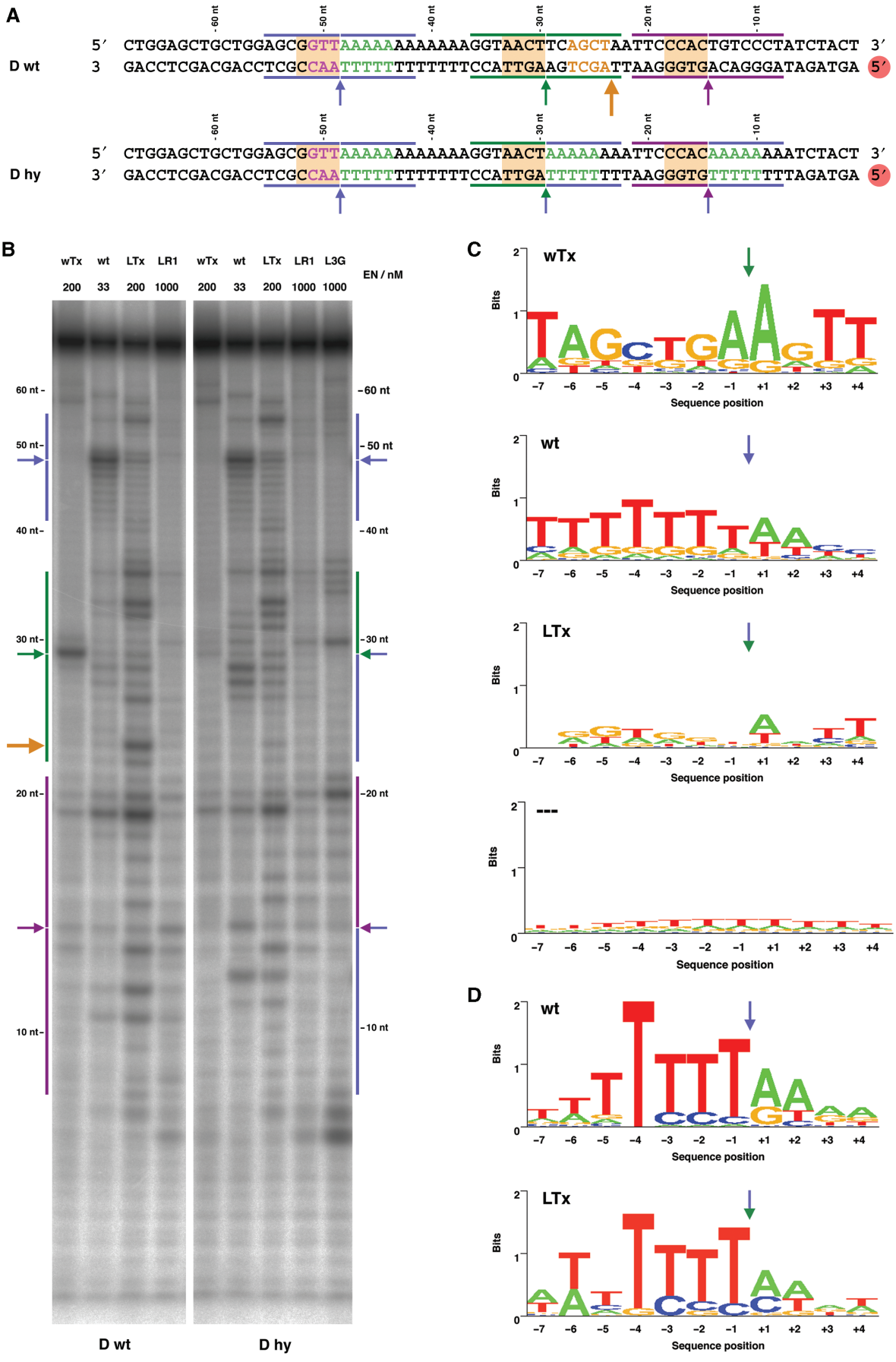


Figure 3. Characterization of the DNA target of L1-EN. (**A**) DNA substrate duplexes containing mismatches or single-strand deletions of the complementary (top) strand. Three base pairs (magenta) downstream of the poly(T)-A junction (blue arrow) and 5 bp (lime) upstream are highlighted. Red circle: 5' end (labeled) of the substrate strand. (**B**) Activity and target recognition of L1-EN hairpin loop variants. DNA duplexes (Cwt, 180 nM) were titrated with increasing concentrations of L1-EN and L1-EN hairpin loop variants. Products were analyzed on autoradiographs of denaturing polyacrylamide gels. Blue arrow: poly(T)-A junction. (**C**) Substrate requirements of L1-EN. L1-EN (160 nM) was used to nick substrates from (A). Lane (P-): Cwt without L1-EN protein. (**D**) Model for DNA target recognition by L1-EN. Top: Scheme of target DNA including the consensus L1 integration sequence. Bottom: 3D model adapted from (27) with L1-EN represented as in Figure 1. The upstream DNA duplex (T-tract geometry, lime) is thought to be contacted by the L1-EN protein scaffold, while the orientation and flexibility of the downstream DNA duplex (magenta) is probed by the insertion of the βB5–βB6 hairpin loop (orange) into the widened minor groove at the poly(T)-A junction.


A
D wt
D hy
B
wTx wt LTx LR1 wTx wt LTx LR1 L3G
EN / nM
200 33 200 1000 200 33 200 1000 1000
60 nt
50 nt
40 nt
30 nt
20 nt
10 nt
D wt
D hy
C
wTx
wt
LTx

Bits
Sequence position
D
wt
LTx

L1-EN, LTx loses the preference for the poly(T)-A junction. Although many nicking sites remain the same for both enzymes, the relative nicking frequencies change. As a result, the LTx nicking profile is clearly distinguishable from the L1-EN pattern (Figure 4B and C, Supplementary Table 1).

Additionally, LTx also nicks novel sites that are recognized neither by L1-EN, nor by Tx1L-EN. This is illustrated by the frequent nick of LTx after nucleotide 23 on Dwt, where the downstream sequence (5′ AGCT 3′) resembles the Tx1L-EN target sequence (5′ AGTT 3′) downstream of nucleotide 29. In this particular case, the sequence of the LTx βB6–βB5 hairpin loop may play a role in the recognition of downstream DNA (Figure 4A and B).

In clear contrast to LTx, the βB6–βB5 loop of LR1 cannot functionally replace the βB6–βB5 loop of L1-EN, as replacement results in a low nicking activity. With respect to specificity, LR1 rather seems to avoid T-tracts and produces a very distinct nicking pattern that is quite similar to the one from the loop deletion variant L3G (Figure 4B, Supplementary Figure 1). The prominent nick of LR1 on Dwt after nucleotide 14 [the preferred nicking site for R1Bm-EN (14)] does not seem to be specific for LR1, since it is present also with L3G (data not shown), L1-EN (Figure 4B) and other variants (Supplementary Figure 1).

Loop grafting has thus produced chimeric endonucleases with altered and novel nicking preferences. In contrast, all analyzed point mutants display nicking patterns that are identical to those of L1-EN (Supplementary Figure 1). They lose activity to various degrees, but maintain specificity. Contrary to the hairpin loop these residues seem to play a rather passive role in contacting an unusual or bendable DNA structure and they might need to be replaced simultaneously to cause any significant effect on nicking specificity.

Requirements for genomic integration of L1 elements are more stringent than requirements to nick target DNA

To test whether the altered nicking specificity of the LTx endonuclease is reflected by an altered integration site preference of the respective L1 variant, we determined the genomic pre-integration sequences from several G418-resistant HeLa cell clones obtained in the cell culture assay (16). Comparison of the *in vitro* nicking profiles to the integration site consensus sequences confirms that for the wild-type L1 element, the nicking specificity of the endonuclease and integration site selection match. However, in the case of LTx, they differ significantly. Like L1, the chimeric LTx element prefers to integrate into locations with a T-tract upstream of the nicking site and only a subset of nicking sites appears to be used for integration (Figure 4C and D).

This is very interesting as it points to additional constraints for L1 retrotransposon targeting other than the DNA nicking specificity of the endonuclease that we assayed on straight DNA duplexes *in vitro*. *In vivo* the poly(T)-A junction may be preferentially recognized in a pre-bent conformation and hence the rigidity of the T-tract could play a much more important role than in the *in vitro* assay. Furthermore, there may be additional contributions for a successful genomic integration, such as base pairing between the 3′ ends of retrotransposon RNA and target site DNA. As a consequence, only a subset of nicked sites would allow for efficient initiation of target-primed reverse transcription.

Loop grafting results in beta-hairpin loops of similar orientation and does not perturb the rest of the L1-EN structure

The distinct effects of the exchanged βB6–βB5 hairpin loop sequences on DNA target recognition and hence nicking specificity is intriguing and may largely relate to the respective structures (Figure 5). We therefore determined the crystal structures of LTx (Figure 5C and D) and LR1 (Figure 5E and F) at 2.3 and 1.8 Å resolution, respectively (Table 2) and compared them to the existing structure of L1-EN (Figure 5A and B) (27). According to an analysis with the program ESCET (40), the common scaffold and catalytic center of the three enzyme variants are essentially unchanged (Figure 5G and H), despite some variance in the crystal packing. The exchanged βB6–βB5 loop sequences are well ordered in both variants, forming protruding beta-hairpins as in wild-type L1-EN, and their orientation is similar.

The backbone of the LR1 hairpin loop superimposes well onto the backbone of the L1-EN hairpin loop (Figure 5G and H). Since the βB6–βB5 hairpin loop of LR1 is two amino acids shorter it lacks the tip (P197 and H198 of L1-EN) that bends towards the minor groove of a putative DNA substrate (Figure 3D). Furthermore, residue T200 of L1-EN is replaced by a glycine in LR1, eliminating an additional possibility of LR1 to interact with the substrate. Finally, the LR1 hairpin loop lacks the positive charges of the L1-EN and LTx hairpin loops that might mediate initial contacts with the negatively charged DNA backbone (Figure 5F). The backbone of the LTx hairpin loop is twisted slightly with respect to the

Figure 4. Target specificity of L1-EN mutants. (**A**) DNA multi-substrate duplexes. Dwt contains wild-type target sites (arrows with seven flanking nucleotides marked by horizontal lines) that are ideal for L1-EN (blue), Tx1L-EN (dark green) and R1Bm-EN (purple). Dhy contains hybrid target sites designed for nicking by LTx (dark green/blue) and LR1 (purple/blue), where the seven upstream base pairs of the ideal target sites of Tx1L-EN and R1Bm-EN, respectively, are replaced by a T-tract. Upstream and downstream base pairs important for recognition by the scaffold and the βB5–βB6 hairpin loop of L1-EN are colored lime and magenta, respectively. Nucleotides on the marked target sites that are thought to be in the reach of the various βB5–βB6 hairpin loops are on an orange background. The major novel target site of LTx on Dwt is marked by an orange arrow with the downstream nucleotides highlighted in orange. Red circle: 5' end (labeled) of the substrate strand. (**B**) Specificity of L1-EN βB5–βB6 hairpin loop variants. DNA duplexes (180 nM) were nicked by the indicated amounts of endonuclease. Products were analyzed on autoradiographs of denaturing polyacrylamide gels. Colors and symbols are as in (A). (**C**) Sequence logos representing nicking profiles. (—), hypothetical logo obtained by assuming random nicking of Dwt. (**D**) Sequence logos representing genomic pre-integration site consensus sequences. Top, $n = 35$, from (16). Bottom, $n = 14$. For details see Supplementary Table 1.

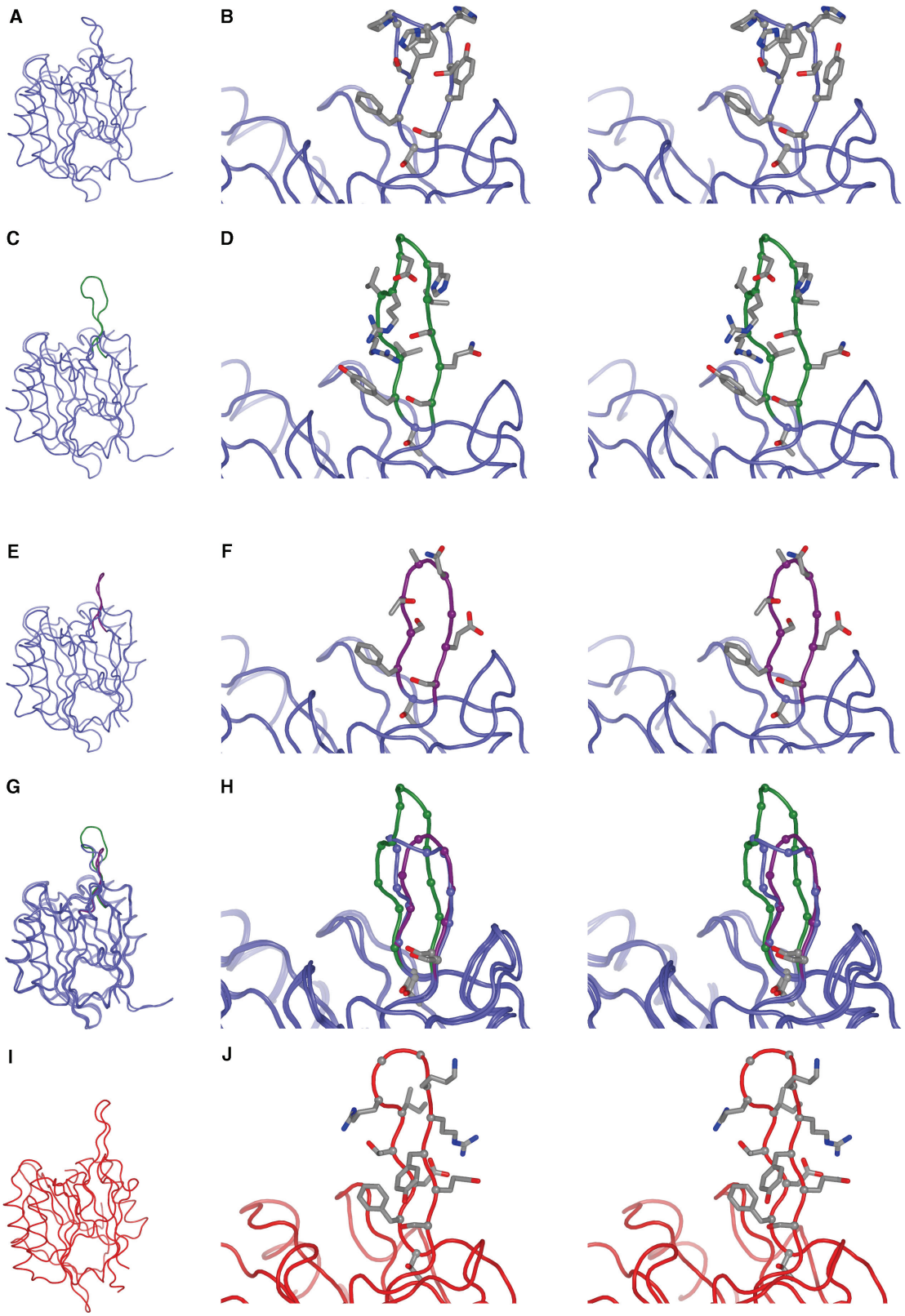


Figure 5. Crystal structures of L1-EN βB5–βB6 hairpin loop variants. The structures of L1-EN and of the two chimeras LTx and LR1 are compared to each other and to the structure of TRAS1-EN. (**A** and **B**) L1-EN (blue). (**C** and **D**) LTx (dark green/blue). (**E** and **F**) LR1 (purple/blue). (**G** and **H**) Superposition of (A), (B) and (C) illustrating differences in size and orientation of the βB5–βB6 hairpin loop. (**I** and **J**) TRAS1-EN (red). Structures are represented as tubes and seen from the side (A, C, E, G, I) or from the front (stereo) zooming in on the loop region (B, D, F, H, J). Side chains of the hairpin loops are shown as balls-and-sticks with carbons in gray, oxygens in red and nitrogens in blue. In (H) side chains are omitted except for T192 and S202.

Table 2. Data collection and refinement statistics

	LTx	LR1
Data collection		
Resolution, Å	2.3	1.8
Space group	$P2_12_12_1$	$C222_1$
Cell dimensions, Å	a = 54.7	a = 58.6
	b = 70.1	b = 67.6
	c = 130.2	c = 128.3
R_{merge}, %[a]	11.2 (48.8)	7.8 (44.2)
Completeness, %[a]	99.8 (100.0)	96.4 (98.6)
$I/\sigma(I)$[a]	8.8 (2.3)	11.9 (2.7)
Number of reflections		
Unique observed	22 873	23 062
Total measured	79 345	88 132
Refinement		
R_{cryst}, %	21.6	18.5
R_{free}, %	26.8	22.2
Number of		
Molecules in asymmetric unit	2	1
Atoms	3948	2039
Ions	6	3
Glycerol molecules	–	1
Water molecules	159	185
Ramachandran plot		
Most favored regions, %	88.5	91.0
Allowed regions, %	10.2	8.5
Generously allowed regions, %	1.4	0.5
R.m.s.d. from ideal geometry		
Bond lengths, Å	0.018	0.013
Bond angles,°	1.81	1.4

[a]Values in parentheses correspond to those in the outer resolution shell (1.89–1.8 Å and 2.4–2.3 Å for LR1 and LTx, respectively).

βB6–βB5 hairpin loop of L1-EN, especially at the distal end (Figure 5G and H). There, the RDGH sequence of Tx1L-EN (Figure 1B) replaces P197 and H198 of L1-EN, forming a more extended tip with side chains that could all make favorable DNA contacts (Figure 5D).

The structure of the catalytic center is not perturbed by the exchange of the loop sequence, suggesting that the mechanism of phosphodiester hydrolysis is not affected directly. It therefore seems likely that certain properties of the βB6–βB5 hairpin loop itself are causing the observed differences in activity and specificity. As the largest structural differences between the three loop variants are at the tip of the loop, we created a point mutation (H198A) in this region of L1-EN. However, the mutation only reduces activity, but does not change target specificity (Supplementary Figure 1). Together with the observation that the chimeric LTx still nicks T-tract DNA despite an entirely different loop sequence this argues against the requirement of sequence-specific protein–DNA contacts. The initial affinity between L1-EN and its target may therefore be based on passive, non-specific contacts resulting from simple complementarity between the shapes of the βB6–βB5 hairpin loop of the endonuclease and the minor groove of the DNA. According to this model, the LR1 hairpin loop is just too short to reach the minor groove properly, explaining why the nicking pattern resembles that of L3G, where the loop is missing entirely.

An additional property of the βB6–βB5 hairpin loop that may be relevant for target selectivity and that would not become obvious from a static crystal structure is its dynamic behavior in the course of the catalytic nicking cycle. In a crucial initiation step, a flexible βB6–βB5 hairpin loop may be needed to probe the dynamics of the minor groove at the junction of the two non-stacking DNA stems.

Normal mode analysis indicates different flexibilities of the grafted hairpin loops

Normal mode analysis is a powerful molecular modeling approach that is particularly suited for calculating slow, large-scale movements within proteins, which would be too expensive computationally for full-scale molecular dynamics simulations. We used the web-based server WEBnm@ (36) to analyze the C-alpha chains of L1-EN, LTx and LR1. As an additional reference we included TRAS1-EN, which is encoded by the telomere-specific APE-type retrotransposon TRAS1 from *B. mori*. Its structure (Figure 5I and J) is characterized by a βB6–βB5 beta-hairpin loop that, like Tx1L-EN, contains eleven residues (41). We calculated the respective average deformation energies of the lowest vibrational mode and also plotted the normalized squared atomic displacements along the sequence of each protein (Supplementary Figure 2).

Low deformation energies indicate that large regions of the protein, possibly domains, can be displaced. For the relatively inactive LR1 we obtain the highest deformation energy (4345), which decreases with increasing loop size via L1-EN (1290) to LTx (684) and TRAS1-EN (510). Furthermore, we clearly identify the βB6–βB5 hairpin loop as the most flexible region in each protein, with a big difference in the extent of the atomic displacement between LR1 and the other three proteins. Taken together, these calculations suggest that an additional reason for the low activity and altered specificity of LR1 in our assays is the missing flexibility of the hairpin loop that is potentially required during the catalytic cycle to lock the DNA target in a suitable position for effective binding and subsequent hydrolysis of the phosphodiester bond.

DISCUSSION

The structural context of the DNA target is highly important for efficient nicking by L1-EN

DNA target specificity of L1-EN has been studied before with plasmid DNA (20) and with special DNA duplexes that contained a symmetric junction of two T-tracts (19). The present study confirms such junctions to be ideal nicking substrates for L1-EN and corroborates the importance of the DNA structure for molecular recognition. We extend the previous analyses to asymmetric DNA targets and determine minimal substrate requirements for the flanking upstream and downstream sequences. Furthermore, we look at the nicking specificity of L1-EN on more general DNA substrates and compare it to the integration specificity of L1 elements *in vivo*.

We find that with unstrained duplex DNA, L1-EN requires a minimum of 5 bp upstream and 3 bp downstream of the target site for efficient target recognition. On the upstream duplex L1-EN recognizes mainly the T-tract

(A-tract) geometry (19) that is primarily characterized by its very narrow minor groove (42). Downstream, the 3 bp are just enough to form an independent stem. In the case of a T-A junction following the T-tract (poly(T)-A junction), the downstream adenine is not stacked on the upstream thymidine (42) and thus, the downstream stem can more easily be bent away with an associated widening of the minor groove. Most likely, this local flexibility is a feature that is recognized by L1-EN in addition to the narrow minor groove of the T-tract, leading to the enhanced nicking efficiency observed at the junction. On a strained substrate such as supercoiled plasmid DNA, the difference between cleaving T-tract DNA and a poly(T)-A junction would probably be even more pronounced. The torsional strain might widen the minor groove at the junction even further and facilitate the structural recognition of the DNA target.

Although the structure of L1-EN would allow the accommodation of a flipped nucleotide at position (+1) downstream of the scissile bond (27), we do not find any evidence for the base-specific recognition of such a nucleotide. At least for the initial target recognition the nucleotide needs to be part of a downstream stem. However, this does not rule out the possibility that the flexibility (or 'flippability') of the nucleotide is required in consecutive steps of the integration process.

L1 integration specificity is influenced by additional factors

In conclusion, L1-EN recognizes structural features of the DNA target rather than specific nucleotides in the sequence. The 5′ TTTT/AA 3′ integration site consensus sequence may fulfill these structural requirements in an ideal way, but many alternative sequences seem to have similar structural features and are nicked *in vitro*. The requirements for integration seem stricter than the requirements for nicking. This indicates that although the nicking specificity of the endonuclease is the primary determinant for integration site selection it may not be the only one (9). Additional specificity factors could influence the choice of nicking site in the first place (co-targeting factors) or select among already nicked sites the ones that are suitable for integration (post-nicking factors). The latter possibility is favored by reports of endonuclease-independent retrotransposition (43) and L1-induced chromosomal breaks (8).

Structure and dynamics of the βB6–βB5 beta-hairpin loop are more important for activity and specificity of L1-EN than sequence

During DNA target site recognition, the conformational space available to the downstream DNA duplex is probed by the insertion of the βB6–βB5 beta-hairpin loop of L1-EN into the minor groove at a poly(T)-A junction, according to the presented model (Figure 3D). The presence of the loop is important for nicking activity and both nicking activity and target specificity are very sensitive to structural changes of the loop, especially at its tip. Similar to the situation in TRAS1-EN (41) a deletion of the tip (LR1) or of the entire loop (L3G) results in an altered specificity and much reduced activity. To examine the importance of the amino acid sequence we exchanged residue H198 in the tip of the loop, which had no impact on the nicking pattern. Even the substitution of the entire loop with a different sequence and an extended reverse turn (LTx) was tolerated rather well. This suggests that the conformational flexibility of the beta-hairpin loop probing the DNA minor groove may be much more important than its sequence, especially if target recognition proceeds via the structural flexibility of the DNA at the poly(T)-A junction. This hypothesis is supported by the presented Normal mode analysis. The βB6–βB5 hairpin loop of LTx may be able to functionally replace the βB6–βB5 hairpin loop of L1-EN because it is flexible enough to insert partially into the minor groove of many L1-EN targets to probe the conformational space of the downstream duplex. The βB6–βB5 hairpin loop of LR1 may be too rigid for this function. In its natural context on R1Bm-EN (14) it may only be required as a counter bearing for the target DNA, which would then be probed sequence specifically from the side of the major groove by a unique extension of surface loop βB4–αB2, predicted for R1Bm-EN (27).

Can novel integration specificities be engineered?

The L1 retrotransposon bears considerable potential as a genetic tool (44). It can be delivered to cells by an adenovirus vector (45) and its suitability for *in vivo* mutagenesis has recently been demonstrated with a synthetic, highly active mouse L1 element called ORF*eus* (46). The application of similar L1 retrotransposons for gene delivery into defined genomic locations requires engineering of the endonuclease target specificity as one of the most crucial steps. This appears feasible since there are many natural APE-type non-LTR retrotransposon endonucleases with distinct target specificities that all share the same protein scaffold and the same catalytic site (9,27).

Loop grafting experiments have been shown to mimic evolutionary processes (47), allowing novel specificities to be engineered (48,49). The analysis of the presented L1-EN βB6–βB5 hairpin loop variants shows that the respective grafting experiments worked successfully from a structural point of view and that other surface loops may be manipulated in a similar way in the future. From a functional point of view, we could show that the DNA nicking profile of L1-EN is quite sensitive to structural changes of the studied loop and that novel specificities can indeed be acquired. For further improvements high-resolution structures of retrotransposon endonucleases in complex with their respective DNA targets would be of great help.

Finally, the apparent existence of additional targeting factors poses further challenges and opportunities for the engineering of novel integration specificities. One such factor may be the contribution from complementary bases between the 3′ end of retrotransposon RNA and the 3′ end of nicked genomic DNA. Tools like the LTx variant will allow us to investigate these effects in the future.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Retrotransposition frequencies and genomic pre-integration sites were determined in the Schumann lab at the Paul-Ehrlich-Institut, biochemical and structural analyses were done at the Netherlands Cancer Institute. We thank Drs J. V. Moran, and D. Carroll for providing plasmids pCEP4/L1.3/ColE1/*mneoI*$_{400}$ and pE1EN. Special thanks go to Jef D. Boeke, Greg Cost and Titia K. Sixma for helpful discussions. We also thank beam line scientists at the ESRF for assistance with data collection. This research was supported by a personal VIDI grant (NWO-CW 700.54.427) to O.W. by the Dutch National Science Organization (NWO) and in part by grants Schu1014/2-1, Schu1014/2-2, Schu1014/2-3 and Schu1014/2-4 of the *Deutsche Forschungsgemeinschaft* (DFG) to G.G.S. K.R. was funded by grant NWO-CW 700.51.012 awarded to A.P. and O.W. Coordinates and structure factors for LTx and LR1 have been deposited with the PDB, accession numbers 2v0r and 2v0s, respectively. Funding to pay the Open Access publication charges for this article was provided by the VIDI grant (NWO-CW 700.54.427).

Conflict of interest statement. None declared.

REFERENCES

1. Bestor,T.H. (2003) Cytosine methylation mediates sexual conflict. *Trends Genet*., **19**, 185–190.
2. Brosius,J. (2003) The contribution of RNAs and retroposition to evolutionary novelties. *Genetica*, **118**, 99–116.
3. Eickbush,T.H. and Malik,H.S. (2002) *Mobile DNA II*, Vol. 49. ASM Press,Washington, D.C., USA, pp. 1111–1144.
4. Han,J.S. and Boeke,J.D. (2005) LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays*, **27**, 775–784.
5. Kazazian,H.H.Jr (2004) Mobile elements: drivers of genome evolution. *Science*, **303**, 1626–1632.
6. Cost,G.J., Feng,Q., Jacquier,A. and Boeke,J.D. (2002) Human L1 element target-primed reverse transcription in vitro. *EMBO J*., **21**, 5899–5910.
7. Luan,D.D., Korman,M.H., Jakubczak,J.L. and Eickbush,T.H. (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.
8. Gasior,S.L., Wakeman,T.P., Xu,B. and Deininger,P.L. (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol*., **357**, 1383–1393.
9. Zingler,N., Weichenrieder,O. and Schumann,G.G. (2005) APE-type non-LTR retrotransposons: determinants involved in target site recognition. *Cytogenet. Genome Res*., **110**, 250–268.
10. Dlakic,M. (2000) Functionally unrelated signalling proteins contain a fold similar to Mg2+-dependent endonucleases. *Trends Biochem. Sci*., **25**, 272–273.
11. Xiong,Y. and Eickbush,T.H. (1988) The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. *Mol. Cell. Biol*., **8**, 114–123.
12. Garrett,J.E., Knutzon,D.S. and Carroll,D. (1989) Composite transposable elements in the Xenopus laevis genome. *Mol. Cell. Biol*., **9**, 3018–3027.
13. Christensen,S., Pont-Kingdon,G. and Carroll,D. (2000) Target specificity of the endonuclease from the Xenopus laevis non-long terminal repeat retrotransposon, Tx1L. *Mol. Cell. Biol*., **20**, 1219–1226.
14. Feng,Q., Schumann,G. and Boeke,J.D. (1998) Retrotransposon R1Bm endonuclease cleaves the target sequence. *Proc. Natl Acad. Sci. USA*, **95**, 2083–2088.
15. Dombroski,B.A., Mathias,S.L., Nanthakumar,E., Scott,A.F. and Kazazian,H.H. Jr (1991) Isolation of an active human transposable element. *Science*, **254**, 1805–1808.
16. Gilbert,N., Lutz-Prigge,S. and Moran,J.V. (2002) Genomic deletions created upon LINE-1 retrotransposition. *Cell*, **110**, 315–325.
17. Symer,D.E., Connelly,C., Szak,S.T., Caputo,E.M., Cost,G.J., Parmigiani,G. and Boeke,J.D. (2002) Human L1 retrotransposition is associated with genetic instability in vivo. *Cell*, **110**, 327–338.
18. Szak,S.T., Pickeral,O.K., Makalowski,W., Boguski,M.S., Landsman,D. and Boeke,J.D. (2002) Molecular archeology of L1 insertions in the human genome. *Genome Biol*., **3**, research0052.
19. Cost,G.J. and Boeke,J.D. (1998) Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry*, **37**, 18081–18093.
20. Feng,Q., Moran,J.V., Kazazian,H.H. Jr and Boeke,J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.
21. Bogerd,H.P., Wiegand,H.L., Hulme,A.E., Garcia-Perez,J.L., O'Shea,K.S., Moran,J.V. and Cullen,B.R. (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl Acad. Sci. USA*, **103**, 8780–8785.
22. Muckenfuss,H., Hamdorf,M., Held,U., Perkovic,M., Löwer,J., Cichutek,K., Flory,E., Schumann,G.G. and Münk,C. (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem*., **281**, 22161–22172.
23. Yang,N. and Kazazian,H.H.Jr (2006) L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol*., **13**, 763–771.
24. Yoder,J.A., Walsh,C.P. and Bestor,T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet*., **13**, 335–340.
25. Takahashi,H. and Fujiwara,H. (2002) Transplantation of target site specificity by swapping the endonuclease domains of two LINEs. *EMBO J*., **21**, 408–417.
26. Kojima,K.K. and Fujiwara,H. (2003) Evolution of target specificity in R1 clade non-LTR retrotransposons. *Mol. Biol. Evol*., **20**, 351–361.
27. Weichenrieder,O., Repanas,K. and Perrakis,A. (2004) Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon. *Structure*, **12**, 975–986.
28. Zou,P., Gautel,M., Geerlof,A., Wilmanns,M., Koch,M.H. and Svergun,D.I. (2003) Solution scattering suggests cross-linking function of telethonin in the complex with titin. *J. Biol. Chem*., **278**, 2636–2644.
29. Wei,W., Morrish,T.A., Alisch,R.S. and Moran,J.V. (2000) A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. *Anal. Biochem*., **284**, 435–438.
30. Leslie,A.G.W. (1992) 'MOSFLM'. *Joint CCP4 and ESF-EACBM Newsletter*, **26.**
31. Evans,P.R. (1997) 'SCALA'. *Joint CCP4 and ESF-EACBM Newsletter*, **33**, 22–24.
32. Vagin,A. and Teplyakov,A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst*., **30**, 1022–1025.
33. Cohen,S.X., Morris,R.J., Fernandez,F.J., Ben Jelloul,M., Kakaris,M., Parthasarathy,V., Lamzin,V.S., Kleywegt,G.J. and Perrakis,A. (2004) Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. Sect. D Biol. Crystallogr*., **60**, 2222–2229.
34. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr*., **53**, 240–255.
35. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr*., **60**, 2126–2132.

36. Hollup,S.M., Salensminde,G. and Reuter,N. (2005) WEBnm@: a web application for normal mode analyses of proteins. *BMC Bioinformatics*, **6**, 52.
37. Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H.Jr (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.
38. Suck,D., Lahm,A. and Oefner,C. (1988) Structure refined to 2A of a nicked DNA octanucleotide complex with DNase I. *Nature*, **332**, 464–468.
39. Mol,C.D., Izumi,T., Mitra,S. and Tainer,J.A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature*, **403**, 451–456.
40. Schneider,T.R. (2000) Objective comparison of protein structures: error-scaled difference distance matrices. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **56**, 714–721.
41. Maita,N., Anzai,T., Aoyagi,H., Mizuno,H. and Fujiwara,H. (2004) Crystal structure of the endonuclease domain encoded by the telomere-specific long interspersed nuclear element, TRAS1. *J. Biol. Chem.*, **279**, 41067–41076.
42. Stefl,R., Wu,H., Ravindranathan,S., Sklenar,V. and Feigon,J. (2004) DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. *Proc. Natl Acad. Sci. USA*, **101**, 1177–1182.
43. Morrish,T.A., Gilbert,N., Myers,J.S., Vincent,B.J., Stamato,T.D., Taccioli,G.E., Batzer,M.A. and Moran,J.V. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.*, **31**, 159–165.
44. Ostertag,E.M. and Kazazian,H.H. Jr (2001) Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.*, **35**, 501–538.
45. Soifer,H.S. and Kasahara,N. (2004) Retrotransposon-adenovirus hybrid vectors: efficient delivery and stable integration of transgenes via a two-stage mechanism. *Curr. Gene Ther.*, **4**, 373–384.
46. An,W., Han,J.S., Wheelan,S.J., Davis,E.S., Coombes,C.E., Ye,P., Triplett,C. and Boeke,J.D. (2006) Active retrotransposition by a synthetic L1 element in mice. *Proc. Natl Acad. Sci. USA*, **103**, 18662–18667.
47. Aharoni,A., Gaidukov,L., Khersonsky,O., Gould,S., Roodveldt,C. and Tawfik,D.S. (2005) The 'evolvability' of promiscuous protein functions. *Nat. Genet.*, **37**, 73–76.
48. Jones,P.T., Dear,P.H., Foote,J., Neuberger,M.S. and Winter,G. (1986) Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature*, **321**, 522–525.
49. Park,H.S., Nam,S.H., Lee,J.K., Yoon,C.N., Mannervik,B., Benkovic,S.J. and Kim,H.S. (2006) Design and evolution of new catalytic activity with an existing protein scaffold. *Science*, **311**, 535–538.