Mol Gen Genet (1999) 262: 772–780 © Springer-Verlag 1999

ORIGINAL PAPER

K. Szafranski · G. Glöckner · T. Dingermann
K. Dannat · A. A. Noegel · L. Eichinger
A. Rosenthal · T. Winckler

Non-LTR retrotransposons with unique integration preferences downstream of *Dictyostelium discoideum* tRNA genes

Received: 27 May 1999 / Accepted: 23 July 1999

Abstract Retrotransposable elements are genetic entities which move and replicate within host cell genomes. We have previously reported on the structures and genomic distributions of two non-long terminal repeat (non-LTR) retrotransposons, DRE and Tdd-3, in the eukaryotic microorganism *Dictyostelium discoideum*. DRE elements are found inserted upstream, and Tdd-3 elements downstream, of transfer RNA (tRNA) genes with remarkable position and orientation specificities. The data set currently available from the *Dictyostelium* Genome Project led to the characterisation of two repetitive DNA elements which are related to the *D. discoideum* non-LTR retrotransposon Tdd-3 in both their structural properties and genomic distributions. It appears from our data that in the *D. discoideum* genome tRNA genes are major targets for the insertion of mobilised non-LTR retrotransposons. This may be interpreted as the consequence of a process of coevolution, allowing a viable population of retroelements to transpose without being deleterious to the small microbial host genome which carries only short intergenic DNA sequences. A new nomenclature is introduced to designate all tRNA gene-targeted non-LTR retrotransposons (TREs) in the *D. discoideum* genome. TREs inserted 5′ and 3′ of tRNA genes are named TRE5 and TRE3, respectively. According to this nomenclature DRE and Tdd-3 are renamed TRE5-A and TRE3-A, respectively. The new retroelements described in this study are named TRE3-B (formerly RED) and TRE3-C.

Communicated by G. P. Geogiev

K. Szafranski · G. Glöckner · A. Rosenthal
Institut für Molekulare Biotechnologie, Beutenbergstrasse 11, D-07745 Jena, Germany

T. Dingermann · K. Dannat · T. Winckler (✉)
Institut für Pharmazeutische Biologie,
Universität Frankfurt/M. (Biozentrum), Marie-Curie-Strasse 9, D-60439 Frankfurt/M, Germany
e-mail: winckler@em.uni-frankfurt.de,
Tel.: +49-69-79829648; Fax: +49-69-79829662

A. A. Noegel · L. Eichinger
Institut für Biochemie I, Universität Köln,
Joseph-Stelzmann-Strasse 52, D-50931 Köln, Germany

Key words *Dictyostelium* · tRNA gene · Retrotransposon · Genome evolution · *Dictyostelium* Genome Project

Introduction

Retrotransposable elements move and replicate within genomes by reverse transcription of retroelement-derived RNA into DNA catalyzed by an RNA-directed DNA polymerase activity (reverse transcriptase, RT). Retroelements are divided into two classes based on the RTs they encode. The class of retroviruses and retrovirus-like retrotransposons encode at least two ORFs flanked by LTRs. Non-LTR retroelements typically encode one or two ORFs which are flanked by nonredundant untranslated regions (UTRs) at both ends. Most non-LTR retroelements end with poly(A) or A-rich stretches at their 3′ ends (Eickbush 1992; Boeke 1997). The 5′ UTRs of many non-LTR retroelements encode promoter activities which ensure the initiation of plus strand RNA synthesis (Mizrokhi et al. 1988; Swergold 1990; Minchiotti and Di Nocera 1991; McLean et al. 1993). The mechanism of retrotransposition of non-LTR retroelements is best known for the insect R2 elements and yeast mitochondrion group II introns. For the silkworm R2Bm element it has been shown that RNA transcripts are bound by retroelement-encoded multifunctional RT/endonuclease proteins via their 3′-UTR sequences (Luan and Eickbush 1995). The endonuclease cleaves the target DNA to produce free 3′-OH ends which can be used by the RT as primers for reverse transcription (Xiong and Eickbush 1988; Luan et al. 1993; Feng et al. 1996).

So far two distinct non-LTR retrotransposons, DRE and Tdd-3, have been reported to exist in the *Dictyostelium discoideum* genome (reviewed in Winckler 1998). Both elements insert independently of fixed DNA target sequences, but in a highly position- and orientation-

specific manner, close to tRNA genes. DRE elements are inserted exclusively 50 ± 4 bp upstream of tRNA genes in *D. discoideum* strains (Marschalek et al. 1989; Hofmann et al. 1991). Full-length DRE elements (DREa) are characterized by a unique modular organization of their untranslated regions (Marschalek et al. 1992b) and are transcribed from both ends (Schumann et al. 1994). The integration specificity of DRE elements resembles that of the LTR retrotransposon Ty3 of *Saccharomyces cerevisiae* (Kim et al. 1998; Sandmeyer 1998), which is known to target to tRNA genes via interaction with RNA polymerase III transcription factors (Kirchner et al. 1995; Connolly and Sandmeyer 1997). The *D. discoideum* Tdd-3 element was originally described as a mobile repetitive DNA element that caused restriction length polymorphisms in the discoidin I genes of several laboratory strains of *D. discoideum* (Poole and Firtel 1984). It was rediscovered when it was found to be associated with tRNA genes (Marschalek et al. 1990), and was subsequently identified as a non-LTR retrotransposon (Winckler et al. 1998). Tdd-3 elements encode two ORFs which are flanked by short UTRs, and all Tdd-3 elements end with oligo(dA) stretches of up to 70 bp in length (Winckler et al. 1998). All genomic Tdd-3 copies analysed so far are inserted about 100 bp downstream of internal B-box promoter motifs or external B-boxes (exB boxes) of tRNA genes (Winckler et al. 1998), suggesting that the RNA polymerase III transcription factor TFIIIC may be involved in the positioning of mobilised Tdd-3 elements. Some Tdd-3 elements are organized in tandem with another repetitive element, RED, with exB boxes being found in the spacer region between the two elements (Marschalek et al. 1990). This may be interpreted as the result of an insertion of a mobilised RED downstream of a B box in a context in which a Tdd-3 was inserted downstream of an exB box of the same tRNA gene. The short DNA sequences available from the 3′ ends of RED elements in RED/Tdd-3 tandems reported in our previous study (Marschalek et al. 1990) did not allow us to draw conclusions about the structural and genomic organization of RED elements. Here we characterize the *D. discoideum* RED element and show that it is a non-LTR retrotransposon related to Tdd-3 in both its structural organization and integration specificity. The sequence information available at this early stage of the *Dictyostelium* Genome Project (for more information on the project see http://www.uni-koeln.de/dictyostelium/ or http://genome.imb-jena.de/dictyostelium/) suggests that a family of non-LTR retrotransposons with insertion specificity for sites downstream of tRNA genes exists in the *D. discoideum* genome.

Materials and methods

DNA isolation and sequencing (*Dictyostelium* Genome Project)

Clones AF067198, AF067199, and AF067200 were isolated from genomic DNA libraries prepared from *D. discoideum* AX2 cells as described (Winckler et al. 1998). *D. discoideum* cells (strain AX4) were grown to mid-log phase, harvested, washed, and allowed to develop into aggregates on phosphate-agar plates at a density of 1×10^8 cells per plate. Aggregates were harvested, cells were mixed with Incert Agarose (Seaplaque) at 40° C, and poured into blocks. The trapped cells were digested with Proteinase K at 50° C for 14 h. Proteinase K and the cell debris were removed by washing the agarose blocks 20 times with TE buffer (10 mM TRIS-HCl, 1 mM EDTA pH 8.0). Blocks were loaded onto a 0.8% LMP agarose gel and the DNA was fractionated by pulsed-field gel electrophoresis (18 h; 160 V, 150 s switch time). The region containing DNA fragments larger than the 90-kb rDNA palindrome was cut out and the agarose was digested with β-agarase (New England Biolabs). The DNA was precipitated with isopropanol, dissolved in H_2O, and used to construct a shotgun library in pUC18 (Craxton 1993). DNA was isolated from the shotgun clones using the Qiagen Turbo extraction kit. Cycle sequencing was done with ABI Big Dye sequencing kits. The sequencing products were separated on ABI 377 sequencing machines.

Consensus construction

Single reads homologous to TRE3 elements were detected using WU-BLASTN 2.0 (Altschul et al. 1990) with scoring parameters M = 6 and N = −18. The identified sequence stretches were locally aligned and assembled using Clustal W 1.7.4 (Thompson et al. 1994) with gap parameters gapopen = 0.35 and gapext = 2.00. Since single sequences were derived from different transposon copies, the alignment contained mismatches, insertions and deletions. Thus, the consensus was built using the majority reads at each position.

Protein alignments and generation of phylogenetic trees

Multiple protein alignments were constructed using Clustal W 1.7.4 with default parameters. Phylogenetic trees were calculated either by the neighbor-joining or the parsimony method provided by PHYLIP 3.5 or the maximum-likelihood method using PUZZLE (Strimmer and von Haeseler 1996).

GenBank accession numbers

DNA sequences reported in this paper have been deposited in GenBank under the following Accession Nos.: AF067198, AF067199, AF067200, AF134169, AF134170, AF134171.

Results

A new nomenclature for tRNA gene-targeted non-LTR retrotransposons in the *D. discoideum* genome

An increasing number of non-LTR retrotransposons is being isolated from the *D. discoideum* genome, all of which appear to be targeted to tRNA genes. We therefore consider it useful to introduce a new nomenclature covering all *D. discoideum* non-LTR retrotransposons. This nomenclature is based on the observed insertion preferences of the retroelements upstream or downstream of tRNA genes. We propose to use the term "tRNA gene-targeted retroelement", TRE. TRE5 designates retroelements that insert upstream (5′) of tRNA genes. The first member of this group discovered, DRE, is renamed TRE5-A. The more diverse group of elements inserted 3′ of tRNA genes described in this paper will be named TRE3-A (Tdd-3), TRE3-B (RED), and TRE3-C (a newly

discovered retroelement, GenBank Accession No. AF134171). Subgroups of the individual elements should be named with numerals, e.g. A.1, A.2 and so forth). Hence DREa and DREb (Marschalek et al. 1992a) are renamed TRE5-A.1 and TRE5-A.2, respectively.

Structure of the TRE3-B (RED) element

We have recently shown that the repetitive DNA element TRE3-A (Tdd-3) from *D. discoideum* is a non-LTR retrotransposon with a unique specificity for integration into the host genome downstream of tRNA genes (Winckler et al. 1998). Short TRE3-B sequences (≤500 bp) had previously been identified as repetitive DNA and isolated through their association with TRE3-A elements in tandem arrays (Marschalek et al. 1990). Using the deduced protein sequence of TRE3-A ORF2 as a query sequence in database searches we noticed that the C-terminal one-third of TRE3-A resembled a corresponding part of TRE3-B (formerly known as RED) (Marschalek et al. 1990; Hitt et al. 1994). The similarity of TRE3-A ORF2 to the TRE3-B-encoded ORF prompted us to determine the full-length DNA sequence of a consensus TRE3-B element and to compare its structural and insertion characteristics with those of TRE3-A. The *Dictyostelium* Genome Project is now rapidly adding to the previously known repetitive DNA sequences in the *D. discoideum* genome. Hence, the complete structure of the TRE3-B retroelement described below is based primarily on data from the *Dictyostelium* Genome Project. For example, the consensus DNA sequence for the full-length TRE3-B element was assembled using multiple sequence reads from the *Dictyostelium* Genome Project. A consensus 5287-bp TRE3-B element contains a 136-bp 5′ UTR and a 110-bp 3′ UTR, which flank two ORFs organized in different reading frames (Fig. 1A). The structural organization of TRE3-B elements resembles that of TRE3-A, which is also reflected in the similarity of their encoded proteins (discussed below). The +1 nucleotide of a full-length TRE3-B was defined by comparison of several individual clones containing random TRE3-B-flanking genomic sequences. According to these data the consensus 5′ end of TRE3-B reads 5′-ACAAAGAAATAAACG...-3′. The 5′ UTR of TRE3-B is completely unrelated to TRE3-A or TRE3-C. Comparison of the DNA sequences of several clones containing TRE3-B 3′-flanking genomic sequences allowed us to define the consensus 3′ end of an TRE3-B element as 5′-...GTTTATTTATATC-3′, which is

A

5' UTR
3' UTR
tDNA
$(A)_{20\text{-}70}$
EN RT HC
ORF1

B

TRE3-B (RED)

```
      M  S  S  K  I  N  F  I  T  W  N
+1  ACTGGTGCTGCCAAACCTAGTAAGGTATGAGCTCAAAAATCAACTTTATAACATGGAAT
+2  ACTGGTGCTGCCAAACCTAGTAAGGTATGAGCTCAAAAATCAACTTTATAACATGGAAT
    T  G  A  A  K  P  S  K  V  *
```

TRE3-C

```
      M  E  Q  L  K  L  L  L  W  N
+2  TTTCCAATTGCCAGTAGTATTAGAAAATAATGGAACAATTAAAATTATTACTATGGAAC
+3  TTTCCAATTGCCAGTAGTATTAGAAAATAATGGAACAATTAAAATTATTACTATGGAAC
    F  P  I  A  S  S  I  R  K  *
```

TRE3-A (Tdd-3)

```
      M  V  V  I  K  I  N  Q  W  N
+2  TCCCAAGAGGAGATCGATATCATCTCATAATGGTAGTAATCAAAATCAACCAATGGAAT
+3  TCCCAAGAGGAGATCGATATCATCTCATAATGGTAGTAATCAAAATCAACCAATGGAAT
    D  Q  E  E  I  D  I  I  S  *  W  *  *
```

Fig. 1A, B Comparison of ORF1/ORF2 overlap regions in TRE3 elements. **A** Schematic representation of the structures of TRE3 elements. EN, endonuclease domain; RT, reverse transcriptase domain; HC, histidine-cysteine-rich domain. The *black triangle* represents a tRNA gene together with its transcription orientation. **B** Overlap regions between ORF1 and ORF2 given for TRE3-B, TRE3-C, and TRE3-A. Translation start codons and translation stop codons are shown as *black boxes* and *grey boxes*, respectively. The deduced amino acid sequences of ORF1 and ORF2 are given *below* and *above* the DNA sequences, respectively

followed by a homopolymeric A stretch of approximately 20 bp in length. The TRE3-B element contains a 110-bp 3′ UTR downstream of ORF2 which includes three adjacent AATAAA consensus polyadenylation sites.

In a full-length TRE3-B element ORF1 and ORF2 are organised in different reading frames, with a 4-bp overlap. An AUG translation start codon is located at the beginning of ORF2 which overlaps with the UGA stop codon of ORF1 (Fig. 1B) leading to a −1 frameshift. Similar arrangements of overlapping stop and start codons but with a frameshift of +2 are also present at the ORF1/ORF2 junctions of TRE3-A and TRE3-C (Fig. 1B). The ORF1 of the consensus TRE3-B element encodes a protein of 402 amino acids. This protein is completely unrelated to any other proteins in the SwissProt database, including ORF1 proteins from TRE3-A and TRE5-A. ORF2 predicts a protein of 1275 amino acids. Comparison with known non-LTR retrotransposons revealed the presence of a putative RT located in the central part of ORF2. The RT domain is the most characteristic feature of retroelements. The amino acid sequences of RTs have been used to establish phylogenetic relationships among retroelements (Xiong and Eickbush 1990; Eickbush 1992). We used the TRE3-B-encoded RT domain as the query sequence in TBLASTN searches of the GenBank and *Dictyostelium* Genome Project databases to search for RT homologs. The results from GenBank searches verified that TRE3-B was a non-LTR retroelement whose RT sequence is only distantly related to RTs encoded by LTR retrotransposons and retroviruses. As expected, the TRE3-B RT domain showed the highest homology to RTs of *D. discoideum* non-LTR retrotransposable elements and was particularly similar to TRE3-A.

Identification and structure of the TRE3-C element

Using the RT domain of TRE3-A ORF2 in a homology search of the data from the *Dictyostelium* Genome Project we identified a new retroelement in the *D. discoideum* genome, TRE3-C (GenBank Accession No. AF134171), whose ORF2 sequence was most closely related to TRE3-A and TRE3-B elements (Fig. 2). This element also encodes reading frames with overlapping stop/start codons at the ORF1/ORF2 boundary (Fig. 1B). TRE3-C contains a 160-bp 3′ UTR including two polyadenylation sites and ends with a homopolymeric d(A) stretch.

Three out of six individual sequences at the 5′ end of TRE3-C were similar to the sequence of the repetitive element Tdd-2 (Poole and Firtel 1984). Three other sequences were found to be truncated around position +200 in the poly(A) stretch located before ORF1. Due to this overlap between part of the TRE3-C 5′ sequences and published Tdd-2 sequence fragments we propose that TRE3-C represents the full-length version of the previously described Tdd-2 element.

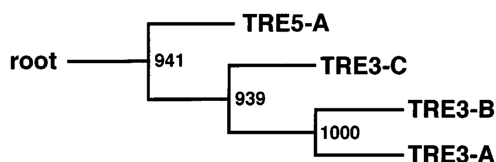

Fig. 2 Phylogenetic analysis of ORF2 sequences. The tree shows the results of the parsimony method. The tree was rooted using the ORF2 sequence of human L1. The numbers represent bootstrap confidence levels from 1000 bootstrap replicates. Other methods used resulted in exactly the same branching pattern but different bootstrap values

Phylogenetic relationships among *D. discoideum* TREs

The deduced ORF2 proteins of all *D. discoideum* TREs were aligned using Clustal W, and a phylogenetic tree was calculated (Fig. 2). TRE3-A, TRE3-B and TRE3-C were grouped on one branch of the resulting tree and were separated from TRE5-A – which appears plausible considering the unique modular structure of TRE5-A UTRs and the TRE5-A integration specificity. We also used the RT sequences corresponding to amino acids 484-768 of TRE3-B ORF2 to perform alignments and phylogenetic analysis of the corresponding regions of non-LTR retrotransposons from other organisms. In the resulting phylogenetic trees the *D. discoideum* TREs were grouped on one branch in the subclass defined by insect R2 and human L1 elements, which is in agreement with previous phylogenetic trees including the TRE5-A RT domain (Xiong and Eickbush 1990; Eickbush 1992) (data not shown).

The HC motif in ORF2 of TRE3-A is a region C-terminal to the RT domain, which has several cysteine and histidine residues in a characteristic arrangement ($H-X_2-C-X_6-H-X_3-C-X_2-C-X_{10}-H-X_4-C$, where X is any amino acid) (Winckler et al. 1998). This domain is also present in ORF2 of TRE3-B (amino acids 1042-1075). A conserved cysteine-rich motif $C-X_{1-3}-C-X_{7-10}-H-X_4-C$ found in non-LTR retrotransposons of many species (Jakubczak et al. 1990) is part of the HC domains of TRE3-B and TRE3-C (not shown).

The 5′ region of ORF2 in many non-LTR retroelements encodes a putative endonuclease (EN) domain similar to apurinic/apyrimidinic (AP) endonucleases (Martín et al. 1995; Feng et al. 1996). As in the TRE3-A and TRE5-A ORF2, a region similar to EN proteins is also found in the ORF2 proteins encoded by TRE3-B and TRE3-C (Fig. 3). Although the overall similarity between EN proteins is rather small, all amino acids shown to be crucial for EN activity (Feng et al. 1996) are conserved in the predicted EN proteins encoded by the *D. discoideum* TRE3 elements.

Distribution of TRE3 copies in the *D. discoideum* genome

Additional data from the *Dictyostelium* Genome Project allowed us to determine the genomic distribution of

```
TRE3-A          MVVIKINQWNVRGW (21) LALTIVNELN (22) GIGILNHNNQ (24) TTRILAIYAPAQP (28)
TRE3-B           MSSK.NFIT..C..F (28) PTIS.LT... (20) .N.VIII.HN (27) PIG..G.....SN (30)
TRE3-C            .EQL.LLL.....N (17) TQ.ALLT.T. (22) .T..AIE.RD (26) SIN..L.....TI (21)
TRE5-A  (05) KTIKKNT.R.GV...Q.S (17) .DAALLT.T. (26) .Q.VSQIIIN (24) QIKCTT.....KS (17)
L1Hs         MTGSNSH.T.LTL.IN.L (17) PSVCCIQ.TH (25) KA.VAILVSD (27) ELT..N....NTG (18)
DdApeA  (98) KNVEENQM..IS...A.F (16) PDVLCLQ.TK (28) .T.V.TKKKP (24) QFY.VNT.I.NAG (29)

TRE3-A       IAIIAGDFNCLDFNDN (39) RPTFSRTIHNTNNNLTRI-LERRLDRIYLN (22) LSDHNFLSTTFTL
TRE3-B       VS........IH-... (11) YQHHLIDQGEFK.TP.HT-HGN......TQ (19) .....PI...ISS
TRE3-C       HQ.......NNHDCNS (10) DQDMLLDTGIEE.TP.FPRSMK......CH (18) K...FPITI.IQT
TRE5-A       SD..T....VDCSV.. (20) -------------GI.FPRNKSTI..VFVS (17) K....MVIIELKI
L1Hs         HTL.M....TPLSTLD (35) EY..FSAP.H.---------YSKI.H.VGS (16) ....SAIKLELRI
DdApeA       PI.WC..L.VAHTEID (42) SY..WSYLGGGRSKNVGW----...YFVVS (18) G...CPIGVVVD.
```

Fig. 3 Alignment of EN domains of TRE elements. Alignment of the EN domains was performed according to Feng et al. (1996). Dots indicate amino acids identities to the TRE3-A sequence. *Dashes* represent gaps introduced to achieve maximum identity. The numbers of amino acids spacing the aligned subdomains are given in parentheses. DdApeA is the sequence of a *D. discoideum* AP endonulease (Freeland et al. 1996)

TRE3-A and TRE3-B elements in more detail than was previously possible (Winckler et al. 1998) (summarised in Table 1). Due to the random nature of the cloning method the numbers of the individual TRE-tRNA junctions reflect the frequency of the different transposons in the genome. At present 27 DNA fragments containing TRE3-A elements with 5′-flanking genomic DNA sequences have been isolated. This is about the number of near full-length TRE3-A elements expected from Southern analysis (T. Winckler, unpublished). Of the 27 isolated TRE3-A copies 21 were found inserted downstream of B-box promoter motifs within the corresponding tRNA genes, and six elements apparently inserted downstream of exB boxes (Table 1). The latter include tandem TRE3-B/TRE3-A elements, which in most cases have exB boxes in the spacer regions between the two retroelements (Marschalek et al. 1990). The distance from either B boxes or exB boxes varied between 62 and 133 bp (average 105 bp). We found that 19 of the 27 TRE3-A copies contained the consensus 5′ sequence of TRE3-A and should thus be derived from full-length elements.

We have so far isolated 11 genomic DNA fragments containing TRE3-B sequences with 5′-flanking sequences (Table 1). Five of these TRE3-B copies were full length and four elements were missing up to 4 bp at their 5′ ends following insertion. DNA fragments carrying 5′-flanking sequences contained B boxes and/or exB boxes in the TRE3-B-flanking regions. The distance of individual TRE3-B 5′ ends from B boxes ranged from 40 to 98 bp (average 67 bp).

From the frequency at which TRE3-C sequences have emerged from the Genome Project we conclude that this element is present in relatively low copy number in the *D. discoideum* genome. We have so far isolated 12 TRE3-C copies which inserted 75-125 bp downstream of tRNA genes (Table 1), suggesting that mobilised TRE3-C elements are targeted to tRNA genes like TRE3-A and TRE3-B.

Integration of retroelements is catalyzed by endonucleases (integrases) which usually cleave both strands of the target DNA sequence via a staggered cut. Cellular DNA enzymes are thought to repair the single-stranded target sequences after insertion of the retroelements, resulting in characteristic target site duplications of variable lengths. TRE3-A and TRE5-A are flanked by target site duplications of 9-10 bp and 12-16 bp, respectively (Poole and Firtel 1984; Marschalek et al. 1993). Two isolated clones (AF067199 and AF067200) contained genomic TRE3-B-flanking sequences on both sides of the retroelement. Comparison of the TRE3-B-flanking DNA sequences did not show target-site duplications, suggesting that the TRE3-B-specific integrase either produces blunt-ended DNA instead of staggered cuts, or deletes the overhanging nucleotides resulting from a staggered cut.

Discussion

Structures of TRE3 elements

In a previous study we described a repetitive DNA element in the *D. discoideum* genome, TRE3-A, which shows a remarkable integration specificity for sites downstream of tRNA genes (Winckler et al. 1998). The structural properties of TRE3-A, namely the organization of ORFs and the nature of the encoded proteins, were indicative of a non-LTR retrotransposon. To our knowledge TRE3-A was the first non-LTR retrotransposon to be described that shows strict integration specificity for sites downstream of tRNA genes. In this study we characterised two new non-LTR retrotransposable elements in the *D. discoideum* genome, TRE3-B and TRE3-C, whose structural features support a relationship to TRE3-A elements and which also insert downstream of tRNA genes in the *D. discoideum* genome.

All TREs encode two ORFs which are organized in separate reading frames. Retroviral *gag* and *pol* ORFs are encoded in separate reading frames and are produced by ribosomal frameshifting as polyprotein precursors from single transcripts (Varmus and Brown 1989). Retrovirus-encoded proteinases cleave the precursor polyproteins into functional enzymes such as integrase and RT. The low efficiency of ribosomal frameshifting accounts for high-level production of the *gag* capsid proteins and low-level production of regula-

Table 1 Compilation of tRNA genes associated with TRE3 elements

Retroelement	tRNA gene	Distance (bp)[a] from		5′ truncation (bp)	Accession No./ clone name[b]	Reference
		B box	exB box			
TRE3-A	Ile (AAU)	111	56	3945	U46204	Vithalani et al. (1996)
TRE3-A	Cys (ACA)	84	–	2001	AF133115	Winckler et al. (1998)
TRE3-A	Val (UAC)	104	–	–	AF133116	Winckler et al. (1998)
TRE3-A	Lys (UUU)	124	71	4492	X59577	Hofmann et al. (1991)
TRE3-A	Val (UAC)	106	25	–	X03499	Hofmann et al. (1991)
TRE3-A	Glu (UUC)	108	57	2	M24566	Hofmann et al. (1991)
TRE3-A	Glu (UUC)	131	82	4	M24567	Hofmann et al. (1991)
TRE3-A	Val (AAC)	89	41	–	M24053	Hofmann et al. (1991)
TRE3-A	Tyr (GUA)	115	66	–	X53447	Hofmann et al. (1991)
TRE3-A	Ala (AGC)	85	35	–	X53444	Hofmann et al. (1991)
TRE3-A	Asn (GUU)	150	99	–	X53443	Hofmann et al. (1991)
TRE3-A	Arg (ACG)	109	53	13	X43438	Hofmann et al. (1991)
TRE3-A	Arg (ACG)	102	48	4935	X59561	Hofmann et al. (1991)
TRE3-A	Met (CAU)	129	68	–		Poole and Firtel (1984)
TRE3-A	Ser (UGA)	100	–	–	JAX4a34b06	
TRE3-A	Asp (GUC)	94	44	–	JAX4a109d05	
TRE3-A	Thr (AGU)	90	43	–	JAX4a35h11	
TRE3-A	Thr (AGU)	133	79	–	JAX4a69b10	
TRE3-A	Leu (UAA)	165	114	–	JAX4a186a09	
TRE3-A	?	97	45	–	JAX4a212b08	
TRE3-A	Arg (UCU)	62	–	25	JAX4a77f03	
TRE3-A	Leu (CAA)	101	46	–	JAX4a135d05	
TRE3-A	Thr (UGU)	130	78	–	JAX4a20e08	
TRE3-A	Ser (UGA)	130	–	–	JAX4b24d09	
TRE3-A	(TRE3-B)	–	86	–		Hofmann et al. (1991)
TRE3-A	(TRE3-B)	–	73	–		Hofmann et al. (1991)
TRE3-A	(TRE3-B)	–	86	–		Hofmann et al. (1991)
TRE3-B	Lys (UUU)	98	27	–	JAX4a218e03	
TRE3-B	Met (CAU)	82	27	4	JAX4a106b09	
TRE3-B	Gly (GCC)	103	51	–	JAX4a117a10	
TRE3-B	Ile (AAU)	67	14	4	JAX4a143g01	
TRE3-B	Thr (AGU)	40	–	3	JAX4a88h08	
TRE3-B	Phe (GAA)	70	20	–	U46205	Vithalani et al. (1996)
TRE3-B	Gly (GCC)	61	9	4	JAX4a45h01	
TRE3-B	Val (AAC)	44	–	–	AF067199	This study
TRE3-B	Val (AAC)	58	5	4174	AF067200	This study
TRE3-B	Gln (UUG)	82	31	4998	JAX4d01d09	
TRE3-B	Arg (UCU)	106	54	5136	X59563	Hofmann et al. (1991)
TRE3-C	Met (CAU)	?	?	–	JAX4a16a06	
TRE3-C	Pro (UGG)	98	44	223	JAX4b13e08	
TRE3-C	Asp (GUC)	75	–	229	JAX4a137h06	
TRE3-C	Pro (UGG)	116	–	230	JAX4b11a07	
TRE3-C	Asp (GUC)	58	–	651	JAX4d04f09	
TRE3-C	Thr (AGU)	135	87	2099	JAX4a121b08	
TRE3-C	Gln (UUG)	99	–	3300	JAX4a170e04	
TRE3-C	Lys (UUU)	95	38	3679	X59576	Hofmann et al. (1991)
TRE3-C	Asp (GUC)			3701	JAX4d03g04	
TRE3-C	Pro (UGG)			3962	JAX4a18f07	
TRE3-C	Asp (GUC)	65	–	4008	JAX4b23h11	
TRE3-C	Cys (GCA)	125	–	4272	JAX4b07h04	

[a] Consensus sequences for B boxes and exB boxes are GTTCRANNC (R = purine) (Marschalek and Dingermann 1991)
[b] JAX clones are from the *Dictyostelium* Genome Project database

tory *pol* proteins. Proteinase homologs are not detected in non-LTR retrotransposons, suggesting that ORF2 proteins may be produced as separate proteins rather than as ORF1/ORF2 fusion proteins. Many non-LTR retroelements encode two non-overlapping ORFs, and experimental evidence suggests that the transcripts function as bicistronic mRNAs (Ilves et al. 1992; McMillan and Singer 1993; Bouhidel et al. 1994). Translation of ORF2 from bicistronic mRNAs may be accomplished by termination-reinitiation or by internal initiation at a specific AUG codon at the 5′ ends of ORF2. As for the *D. discoideum* TRE3 elements, a mechanism of termination-reinitiation is likely to function in ORF2 translation, because such a translation mechanism is favored when the AUG codon of the second ORF is very close to the stop codon of the first ORF (Jacks et al. 1988). This is undoubtedly true for the TRE3 elements, where the start and stop codons overlap

(Fig. 1). Support for a termination-reinitiation mechanism rather than frameshifting in ORF2 translation from TRE3 transcripts comes from the observation that neither of the *D. discoideum* TRE3 elements contain signals for ribosomal frameshifting or stem-loop structures (Jacks et al. 1988) at the site of the ORF1/ORF2 overlap. For TRE3-A, we noticed two variants of DNA sequence covering the ORF1 stop codon (Fig. 1B, bottom): TAATGGTAGTAA and TAATAGTAGTAA. The latter sequence does not contain the methionine translation start of ORF2, placing the next possible translation initiation site some 500 bp downstream. If termination-reinitiation is assumed as the mechanism of ORF2 translation, the TRE3-A elements that display the TGG to TAG mutation may produce an aberrant ORF2 protein or none at all.

The RT domains of all known *D. discoideum* non-LTR retrotransposons were subjected to phylogenetic analysis. Based on these data we conclude that the analyzed retroelements diverged from a common ancestor, such that TRE3-B and TRE3-A elements separated from TRE3-C and TRE5-A earlier than from each other. The separation of the *D. discoideum* non-LTR retroelements into two groups defined by TRE3-A and TRE5-A is supported by the different insertion specificities developed by the two groups of retroelements and by the unique modular structure of TRE5-A UTRs. TRE3-C appears to be a distantly related member of the TRE3-A/TRE3-B group, as suggested by the phylogenetic tree of the RT domains (Fig. 2) and the observation that TRE3-C lacks the HC domain contained in TRE3-A and TRE3-B ORF2.

Many non-LTR elements have internal promoters located within the first few hundred base pairs which control transcription of the elements (Eickbush 1992). Since the transcription of retrotransposons into plus-strand RNAs is the first step in their retrotransposition, it is assumed that only those retroelements with complete 5′ ends and intact ORFs can actively transpose (Boeke 1997). As for the TRE3-A element, we know that a promoter is located within the first 350 bp of the retroelement, which is capable of supporting reporter gene expression in *D. discoideum* cells (T. Winckler, unpublished data). TRE3-A-derived transcripts but no TRE3-B transcripts are detectable by reverse-transcription PCR of RNA isolated from growing *D. discoideum* cells (Winckler et al. (1998) and unpublished results). We know from the analysis of isolated genomic clones, e.g. 9.10 (AF067198), that mobilised TRE3-A elements can destroy the 5′ ends of TRE3-B elements by inserting downstream of a tRNA gene that is already associated with a TRE3-B element. TRE3-A elements can insert about 100 bp downstream of either tRNA gene-internal B boxes or exB boxes, the latter being located 30-40 bp downstream of many *D. discoideum* tRNA genes. If it inserts downstream of an exB box, a TRE3-A element may come to lie 50-100 bp into the 5′ regions of TRE3-B elements and disrupt the TRE3-B promoter presumed to be localized at this site. The disruption of TRE3-B 5′ ends by TRE3-A insertion may account for part of the transcriptionally silent TRE3-B population. Currently, we cannot decide whether the isolated 5′ consensus ends of TRE3-B copies can serve as functional promoters for TRE3-B transcription.

Two subfamilies of tRNA gene-targeted non-LTR retrotransposons in *D. discoideum*

All non-LTR retrotransposons currently known to exist in the *D. discoideum* genome are positioned close to tRNA genes. Two “subfamilies” of tRNA gene-associated retrotransposons can be defined by their insertion preferences for sites upstream or downstream of tRNA genes. This classification is supported by differences in the molecular structures of the members of the two groups. Whereas TRE5-A elements have modular UTRs, the retroelements specific for insertion 3′ of tRNA genes have rather simple 5′ UTRs and 3′ UTRs. Both groups of retroelements have in common that integration occurs independently of a fixed target sequence and is strictly orientation specific. This points to specific protein-protein interactions being involved in determining the sites of integration. By analogy to RNA polymerase III transcription factor-guided insertion of yeast Ty3 retrotransposons (Kirchner et al. 1995) one can imagine that TFIIIC bound to B-box or exB-box motifs may be involved in targeting TRE3 elements downstream of tRNA genes. Protein interactions upstream of tRNA genes for specific insertion of TRE5-A probably require TFIIIB to be bound to the tRNA gene in addition to TFIIIC.

The replication of “selfish” retroelements within a host cell genome, accompanied by random integration of mobilised retroelements, may cause problems as a result of insertion mutagenesis of the host genome, eventually leading to decreased evolutionary fitness or cell death. Clustering of retroelements adjacent to tRNA genes has been assumed to be the consequence of a process of coevolution of host cells with retroelements that allows populations of retroelements to remain mobile without having harmful consequences for the hosts (Voytas and Boeke 1993; Voytas 1996; Kim et al. 1998). The integration of retroelements near tRNA genes may circumvent the problem of insertion mutagenesis of structural genes, since the regions flanking tRNA genes are usually devoid of coding regions. The large number of tRNA genes within the genome offers multiple “safe” landing sites for mobilized retroelements. In this context it may be relevant to note that tRNA gene-associated integration of retroelements has as yet only been described in eukaryotic organisms that have small genomes with short intergenic regions, such as *D. discoideum* and yeasts.

The data presented here are expected to have an impact on the ongoing *Dictyostelium* Genome Project, which has in turn provided valuable data for the characterization of the retroelements described in this

study. From the copy numbers of the individual TREs in the *D. discoideum* genome we can estimate that about 4% of the genomic DNA consists of non-LTR retroelements. Detailed knowledge of the molecular structure and genomic distribution of these elements will provide landmarks for the assembly of the whole genome sequence. On the other hand, we can estimate that at least two-thirds of the tRNA genes in the *D. discoideum* genome are accompanied by one or more non-LTR retroelements. Hence searching the accumulating *D. discoideum* genome sequence information for tRNA genes will identify known or new TREs, and, conversely, analysis of TREs will help to identify new tRNA genes.

Acknowledgements We thank D. Pätsch, C. Tschepke, S. Richter, and R. Müller for expert technical assistance. This work was supported by a grant from the Deutsche Forschungsgemeinschaft to T.D. and T.W. The German part of the *D. discoideum* Genome Project is carried out by the Institute of Biochemistry I in Cologne, and the Genome Sequencing Centre in Jena. It is supported by the Deutsche Forschungsgemeinschaft (No 113/10–1 and 10–2).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Boeke JD (1997) LINEs and Alus – the polyA connection. Nature Genet 16:6–7

Bouhidel K, Terzian C, Pinon H (1994) The full-length transcript of the I factor, a LINE element of *Drosophila melanogaster*, is a potential bicistronic RNA messenger. Nucleic Acids Res 22:2370–2374

Connolly CM, Sandmeyer SB (1997) RNA polymerase III interferes with Ty3 integration. FEBS Lett 405:305–311

Craxton M (1993) Cosmid sequencing. Methods Mol Biol 23:149–167

Eickbush TH (1992) Transposing without ends: the non-LTR retrotransposable elements. New Biol 4:430–440

Feng QH, Moran JV, Kazazian HH, Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905–916

Freeland TM, Guyer RB, Ling AZ, Deering RA (1996) Apurinic/apyrimidinic (AP) endonuclease from *Dictyostelium discoideum*: cloning, nucleotide sequence and induction by sublethal levels of DNA damaging agents. Nucleic Acids Res 24:1950–1953

Hitt AL, Lu TH, Luna EJ (1994) Ponticulin is an atypical membrane protein. J Cell Biol 126:1421–1431

Hofmann J, Schumann G, Borschet G, Gosseringer R, Bach M, Bertling WM, Marschalek R, Dingermann T (1991) Transfer RNA genes from *Dictyostelium discoideum* are frequently associated with repetitive elements and contain consensus boxes in their 5′-flanking and 3′-flanking regions. J Mol Biol 222:537–552

Ilves H, Kahre O, Speek M (1992) Translation of the rat LINE becistonic RNAs in vitro involves ribosomal reinitiation instead of frameshifting. Mol Cell Biol 12:4242–4248

Jacks T, Madhani HD, Masiarz FR, Varmus HE (1988) Signals for ribosomal frameshifting in the Rous Sarcoma Virus *gag-pol* region. Cell 55: 447–458

Jakubczak JL, Xiong Y, Eickbush TH (1990) Type I (R1) and type II (R2) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. J Mol Biol 212:37–52

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res 8:464–478

Kirchner J, Conolly CM, Sandmeyer SB (1995) Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science 267:1488–1491

Luan DD, Eickbush TH (1995) RNA template requirements of target DNA-primed reverse transcription by the R2 retrotransposable element. Mol Cell Biol 15:3882–3891

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72:595–605

Marschalek R, Dingermann T (1991) Structure, organization and function of transfer RNA genes from the cellular slime mold *Dictyostelium discoideum*. Adv Gene Technol 2:103–143

Marschalek R, Brechner T, Amon-Böhm E, Dingermann T (1989) Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. Science 244:1493–1496

Marschalek R, Borschet G, Dingermann T (1990) Genomic organization of the transposable element Tdd-3 from *Dictyostelium discoideum*. Nucleic Acids Res 18:5751–5757

Marschalek R, Hofmann J, Schumann G, Dingermann T (1992a) Two distinct subforms of the retrotransposable DRE element in NC4 strains of *Dictyostelium discoideum*. Nucleic Acids Res 20:6247–6252

Marschalek R, Hofmann J, Schumann G, Gösseringer R, Dingermann T (1992b) Structure of DRE, a retrotransposable element which integrates with position specificity upstream of *Dictyostelium discoideum* tRNA genes. Mol Cell Biol 12:229–239

Marschalek R, Hofmann J, Schumann G, Bach M, Dingermann T (1993) Different organization of the transfer RNA-gene-associated repetitive element, DRE, in NC4-derived strains and in other wild-type *Dictyostelium discoideum* strains. Eur J Biochem 217:627–631

Martín F, Marañón C, Olivares M, Alonso C, López C (1995) Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the Ape family of DNA repair enzymes. J Mol Biol 247:49–59

McLean C, Bucheton A, Finnegan DJ (1993) The 5′ untranslated region of the I factor, a long interspersed nuclear element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression. Mol Cell Biol 13:1042–1050

McMillan JP, Singer MF (1993) Translation of the human LINE-1 element, L1Hs. Proc Natl Acad Sci USA 90:11533–11537

Minchiotti G, Di Nocera P (1991) Convergent transcription initiates from oppositely oriented promoters within the 5′ end regions of *Drosophila melanogaster* F elements. Mol Cell Biol 11:5171–5180

Mizrokhi LJ, Georgieva SG, Ilyin YV (1988) Jockey, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell 54:685–691

Poole SJ, Firtel RA (1984) Genomic instability and mobile genetic elements in regions surrounding two discoidin I genes of *Dictyostelium discoideum*. Mol Cell Biol 4:671–680

Sandmeyer S (1998) Targeting transposition: at home in the genome. Genome Res 8:416–418

Schumann G, Zündorf I, Hofmann J, Marschalek R, Dingermann T (1994) Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the *Dictyostelium* Repetitive Element, from *Dictyostelium discoideum*. Mol Cell Biol 14:3074–3084

Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Swergold GD (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol 10:6718–6729

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Varmus H, Brown P (1989) Retroviruses. In: Berg DE, Howe MM (eds) Mobile DNA. American Society of Microbiology, Washington DC, pp 53–108

Voytas DF (1996) Retroelements in genome organization. Science 274:737–738

Voytas DF, Boeke JD (1993) Yeast retrotransposons and tRNAs. Trends Genet 9:421–427

Winckler T (1998) Retrotransposable elements in the *Dictyostelium discoideum* genome. Cell Mol Life Sci 54:383–393

Winckler T, Tschepke C, de Hostos EL, Jendretzke A, Dingermann T (1998) Tdd-3, a tRNA gene-associated poly(A) retrotransposon from *Dictyostelium discoideum*. Mol Gen Genet 257:655–661

Xiong Y, Eickbush TH (1988) Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. Cell 55:235–246

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362