2408–2415 *Nucleic Acids Research, 1999, Vol. 27, No. 11* © *1999 Oxford University Press*

Tdd-4, a DNA transposon of *Dictyostelium* that encodes proteins similar to LTR retroelement integrases

David J. Wells*

Program in Molecular Biology, Department of Biology, Utah State University, Logan, UT 84322, USA

Received January 6, 1999; Revised and Accepted April 12, 1999

DDBJ/EMBL/GenBank accession nos+

ABSTRACT

Tdd-4 is the first DNA transposon to be isolated from *Dictyostelium discoideum*. This element was isolated by insertion into a target plasmid. Two classes of elements were identified which include a 3.8 kb version and a 3.4 kb deleted version. Sequence analysis reveals that the 145 bp inverted terminal repeats contain the 5′-TG…CA-3′ conserved terminal dinucleotides found in prokaryotic transposons and integrated LTR retroelement DNA sequences. Tdd-4 open reading frames are assembled by removal of six introns. Introns 1–5 conform to the GT-AG rule, whereas intron 6 appears to be an AT-AA intron. Also, intron 6 undergoes an alternative 5′ splicing reaction. The alternatively spliced region encodes 15 tandem SPXX repeats that are proposed to function as a DNA binding motif. By analogy to other transposons that encode two proteins from the same gene, the full-length Tdd-4 protein is the putative transposase and the truncated Tdd-4 protein is the putative transposition inhibitor. Protein database searches demonstrate Tdd-4 encoded proteins are unique for a DNA element by containing similarities to retroviral/retrotransposon integrases. The putative Tdd-4 transposase contains the same structural relationship as integrases by possessing an N-terminal HHCC motif, a central DDE motif and a C-terminal DNA-binding domain composed of the SPXX motif.

INTRODUCTION

Transposable elements are DNA segments that can move to new chromosomal or extrachromosomal locations in the absence of homology between the donor and recipient sites. Classification of these elements is based, in part, on the orientation of the terminal nucleotide sequence and the intermediate form of the element during the transposition event (1). The LTR retroelements (retroviruses and LTR retrotransposons) possess a long terminal repeat sequence in a direct orientation and transpose via an RNA intermediate prior to synthesis of a double-strand DNA copy. Conversely, prokaryotic and eukaryotic DNA transposons possess a terminal repeat sequence in an inverted orientation (ITR) and remain strictly in the DNA form during transposition.

Despite differences in the intermediate forms and terminal nucleotide structures, similarities have been noted at the nucleotide and amino acid level of these elements. A feature identified from integrated LTR retroelements and several prokaryotic transposons is a conservation of the 5′-TG…CA-3′ terminal dinucleotides. The essential nature of the CA-3′ dinucleotide was demonstrated by mutational studies which impaired strand cleavage or transfer reactions for retroviral DNA (2–4) and prokaryotic transposons (5–7). The enzymes that mediate transposition are the integrases for LTR retroelements and the transposases for DNA elements. A region of similarity, called the DDE motif, has been identified among integrases and transposases by alignment of their amino acid sequences (8–10). The two aspartate and one glutamate residues are proposed to form the catalytic region of the protein through coordination of a Mn^{2+} or Mg^{2+} metal ion (for a schematic representation see 11). The importance of these residues was confirmed by mutations introduced into the DDE coding positions of retroviral integrase (10,12,13), retrotransposon integrase (14), eukaryotic transposase (15) and prokaryotic transposase (16). N-terminal to the DDE motif, another metal-binding region is present in integrase. This zinc finger-like structure has been termed the HHCC motif due to the ordered sequence of histidine and cysteine residues (17). The HHCC motif appears to be fully conserved among integrase proteins, whereas this domain has not been identified in transposase proteins (18,19). C-terminal to the integrase DDE motif, a third domain has been identified which is involved in a non-specific binding to DNA (20–22).

Five transposable elements have been identified in the eukaryotic microorganism *Dictyostelium discoideum*: *skipper* (23); Tdd-1 (24), also called DIRS-1 (25); Tdd-2 (26); Tdd-3 (27); DRE (28). Excluding Tdd-2, each of these elements has been completely sequenced and contains structural features and/or encodes proteins indicative of retroelements. Tdd-2 has not been well characterized but contains an A-rich terminus, suggesting that this element may be a non-LTR retrotransposon. The following report describes the first DNA transposon, Tdd-4, to be discovered in *Dictyostelium* and details the unique features of its encoded proteins which have similarities to the LTR retroelement integrases.

*Correspondence should be addressed to present address: Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, 903 South Fourth Street, Hamilton, MT 59840, USA. Tel: +1 406 363 9267; Fax: +1 406 363 9371; Email: dwells@atlas.niaid.nih.gov

+U57081, U63593–U63608

MATERIALS AND METHODS

Strains and culture conditions

All analyses in this study were performed with the *D.discoideum* AX4 strain. Axenic cultures were grown in HL5 medium (29) and non-axenic cultures were grown on DM agar (30) with *Escherichia coli* B/r as a food source. The bacterial recipient strains used for DNA constructs were *E.coli* DH5a and SURE (Stratagene).

DNA sequencing

DNA sequences were generated by the dideoxy chain termination method using cloned and subcloned DNA fragments and exonuclease III deletion versions of subcloned fragments. The Tdd-4 element from p93d1 was the only copy to be fully sequenced. Other copies were acquired by PCR amplification of Tdd-4 elements contained in a *D.discoideum* yeast artificial chromosome genomic library (kindly provided by W. Loomis, University of California, San Diego, CA) or by excising and eluting appropriately sized AX4 restriction-digested genomic DNA fragments from agarose gels. The DNA sequence for the Tdd-4 element from p93d1 was generated by manual sequencing. All subsequent sequencing was performed by the Utah State University Biotechnology Center using an automated sequencer. The complete nucleotide sequence of Tdd-4 is available under GenBank accession no. U57081. DNA sequences flanking Tdd-4 elements are under accession nos U63593–U63608. Oligonucleotide sequences mentioned in this study are as follows. Mismatches with the Tdd-4 sequence are in bold and were introduced to create the identified restriction site for cloning amplified fragments. OUD4, 5′-GGATGGTTTAGAATGTG (positions 2503–2487); OUD6, 5′-GATGGAATCAATAGCC (positions 1907–1922); OUD8, 5′-GATTATTAGTTCTATATAAG (positions 160–179); OUD9, 5′-CGGTATCTACAATTGGAGC (positions 862–844); OUD15, 5′-CAAAAAACAAATAAAG (positions 119–134 and 3720–3705); OUD23, 5′-CTAGCTTTG**G**TACCTCTGGTGCATTTG (*Kpn*I) (positions 3371–3345); OUD25, 5′-TAGAATGGTA**C**CACTCTGGGAAG (*Kpn*I) (positions 590–612); OUD30, 5′-TATTTTC**CCGGG**CCCAAAATG AAACAAG (*Sma*I) (positions 306–333); OUD22, 5′-CACTACGCTACTG**C**AGTGAAATTAAATAATAAAATTCACTACGC (*Pst*I) (positions 81–38 and 3759–3802).

DNA and RNA isolation and analysis

Dictyostelium discoideum AX4 total nuclear DNA from non-axenic cultures was isolated (31) and restriction enzyme-digested fragments were size fractionated in 0.8% agarose gels prior to transfer onto nitrocellulose membranes by the Southern blotting procedure (32). Total RNA, used for reverse transcription, from AX4 axenic cultures was isolated using TriReagent (Molecular Research Inc.). All hybridizations for detection of Tdd-4 elements used a [^{32}P]dATP random labeled 2.15 kb *Nsi*I–*Sca*I Tdd-4 DNA fragment.

Reverse transcription PCR

Reverse transcription reactions were performed with 2 µg of total RNA, 25 pmol of OUD9 or OUD23 and 0.5 mM each deoxynucleoside triphosphate in a reverse transcriptase buffer (50 mM Tris–HCl, pH 8.3, 75 mM KCl, 5 mM $MgCl_2$, 10 mM dithiothreitol). These mixtures were heated to 70°C and slowly cooled to room temperature. Moloney murine leukemia virus reverse transcriptase (200 U; Promega) was added and samples were incubated for 30 min at 37°C. Aliquots from reverse transcription reactions were amplified by PCR using OUD30/OUD9 and OUD25/OUD23 primer pairs.

Inverse PCR

Inverse PCR (33) was employed to obtain DNA fragments that flank Tdd-4 elements. Template DNA for PCR was prepared from restriction enzyme-digested genomic AX4 DNA or *D.discoideum* DNA contained in a yeast artificial chromosome. Following a screening of restriction enzymes that do not contain sites within Tdd-4, five enzymes (*Bcl*I, *Bgl*II, *Csp*45I, *Eco*RV and *Nde*I) were chosen which generate appropriately sized fragments (~4–7 kb) when analyzed by Southern blot. Digested DNA was fractionated on agarose gels and desired sized fragments were excised, eluted and circularized with T4 DNA ligase. Template DNA was amplified by PCR using the OUD22 primer. This primer recognized both ITRs of a Tdd-4 element and allowed the terminal 37 bases from each ITR to be amplified. Once sequenced, these 37 bases served to verify that the amplified fragment contained the Tdd-4 element and showed the boundaries between the Tdd-4 sequence and the flanking sequence.

Database searches and computer analysis

DNA and protein sequences were analyzed with the GCG package (Madison, WI). Protein database searches utilized BLAST for searching GenBank or BLITZ for searching SwissProt.

RESULTS

Isolation of Tdd-4

The discovery of the sixth *D.discoideum* transposable element resulted from experiments involving the analysis of *Dictyostelium* plasmids (J.Hughes, personal communication). Studies were conducted to determine whether a region of the *D.discoideum* plasmid, Ddp1, could confer maintenance to a shuttle vector construct. The p88d1 shuttle vector (34) was composed of the prokaryotic vector pGEM-3Z, a 2 kb *Hin*dIII fragment reported to contain the Ddp1 origin of replication and a gene encoding resistance to the antibiotic G418; this vector was used in a maintenance assay to analyze its stability. Following transformation with p88d1, two distinct G418 phenotypes were observed. One population of transformants was unable to grow in the presence of G418 and the other population of transformants was resistant to this antibiotic. Restriction digestion of vectors isolated from *E.coli* transformed with the G418-sensitive or G418-resistant DNA showed that the vector from the sensitive transformant remained unchanged, whereas the vector obtained from the resistant transformant was ~4 kb larger than its parental vector, p88d1. Sequence analysis of p93d1, the vector obtained from the resistant transformant, revealed a DNA segment, designated Tdd-4, inserted into an upstream region of the *D.discoideum* actin 6 promoter which drives expression of the G418 resistance gene. Based on the DNA sequence of the actin 6 promoter (35) in the G418 resistance gene, the flanking sequence from the Tdd-4 element in p93d1 showed the element inserted ~530 bp upstream of the G418 translation start site.

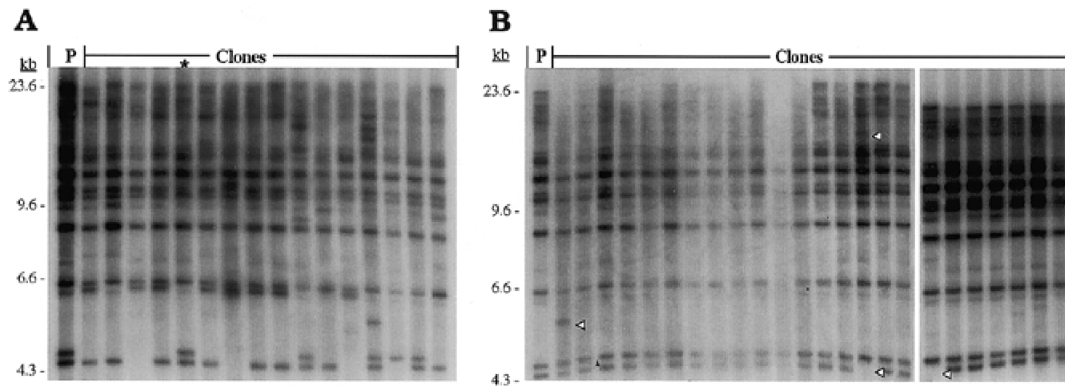

Figure 1. Tdd-4 RFLP analysis from AX4 clones. (**A**) Sixteen clones were derived from the parental AX4 line (P) following a few hundred generations of growth and analyzed by Southern blot using *Eco*RV-digested genomic DNA and a 2.15 kb *Nsi*I–*Sca*I Tdd-4-specific probe. (**B**) The strain marked by the asterisk in (A) was used as a parent and 24 clones were derived following approximately 20 generations of growth. Arrows indicate elements that have transposed.

Mobility of Tdd-4

The p88d1 vector maintenance assay suggested Tdd-4 transposed into p93d1. Additional evidence for the movement of Tdd-4 was obtained from restriction fragment length polymorphism (RFLP) studies. Sixteen clones of normal developmental phenotype were single cell-derived from a parental *D.discoideum* AX4 line following a few hundred generations of growth. Genomic DNA from the parental line and the 16 clones was digested with *Eco*RV and probed with a 2.15 kb *Nsi*I–*Sca*I fragment derived from p93d1. Figure 1A shows the pattern of bands hybridizing to Tdd-4. Almost every clone displayed a restriction fragment pattern variability when compared to the parental line. Specifically, the two bands located near the 4.3 kb size marker display a high variability with only four of the 16 clones possessing the original banding pattern. To obtain additional data for the mobility of Tdd-4, the clone marked by the asterisk in Figure 1A was grown to give rise to 24 clones of normal developmental phenotype. Twenty-four new clones were obtained from single cell-derived colonies of the new parental line after a period of about 20 generations of non-axenic growth. RFLP analysis showed that four of the 24 clones appeared to have undergone a transposition event (Fig. 1B). To verify that these polymorphisms were due to transposition events and not changes in the *Eco*RV restriction sites, another RFLP analysis was performed on the same 24 DNA samples digested with *Csp*45I (data not shown). This analysis showed similar results. The *Eco*RV and *Csp*45I restricted blots show that approximately 15–20 Tdd-4 copies are present in the AX4 genome.

Sequence analysis of Tdd-4

The complete 3839 bp Tdd-4 sequence provided three features indicative of a transposable element: (i) ITRs; (ii) conserved terminal dinucleotides; (iii) a target site duplication. First, the right and left repeats of Tdd-4 ITRs, respectively, have lengths of 146 and 145 bp and differ by only three bases (Fig. 2A). Second, Tdd-4 contains the terminal 5′-TG…CA-3′ dinucleotides found in prokaryotic transposons and integrated LTR retroelement sequences. Third, DNA sequence comparison of the actin 6 promoter region in the G418 resistance gene from p93d1 to that of its parental vector, p88d1, showed that a 5 bp target site duplication was elicited by the Tdd-4 insertion. Seven additional sequences flanking Tdd-4 elements were obtained by inverse PCR of genomic AX4 DNA fragments and one (p435) was derived from a *D.discoideum* genomic library. The nucleotide difference 18 bases from the ends of the right and left ITRs in Figure 2A is also contained in each inverse PCR product. The direct repeat lengths deduced from the Tdd-4 flanking sequences, which are presumed to be target site duplications, are shown in Figure 2B. Five clones, including p93d1, showed a direct repeat of 5 bp and two clones showed a direct repeat of 4 bp. It should be noted that the 5 bp repeat for clone p426 and 4 bp repeat for p427/428 are formed only if a single base change is permitted. The 2 bp direct repeat for clone p435 deviates from the 4–5 bp direct repeat shown for the majority of the Tdd-4 direct repeats. The absence of a direct repeat flanking the Tdd-4 element for clone p432 may be due to a deletion or insertion event of the target DNA during Tdd-4 integration. Anomalies to the target site sequence have also been observed for *in vitro* retroviral DNA integration and include smaller than normal target duplications and imperfect repeats (36) and deletions or insertions of target DNA (37,38).

Additional analysis of the DNA sequence did not reveal an extended open reading frame (ORF) that might encode a protein to mediate transposition. Based on the stop codon profile, either the copy obtained from p93d1 was defective or mRNA splicing may be involved to produce an extended ORF. To explore the first possibility, additional elements were obtained by PCR amplification from genomic AX4 DNA or from a *D.discoideum* genomic library. PCR amplifications were designed such that either the right half (using primer pair OUD6/OUD15) or the left half (using primer pair OUD8/OUD4) of the elements would be amplified. The amplified fragment from the right half corresponded in size to that from the Tdd-4 copy of p93d1. The amplified fragments from the left half yielded two bands, one of which corresponded in size to that from the Tdd-4 copy from p93d1 and another which was ~400 bp smaller. Sequence analysis revealed that one class of elements was identical in sequence to the Tdd-4 copy from p93d1 and a second class contains two major deletions and a single base change that disrupted an *Acc*I restriction site, relative to the p93d1 copy of Tdd-4 (Fig. 3A). The first deletion (nt 334–350) and the second deletion (nt 802–1195) resulted in Tdd-4 elements that were shorter by ~411 bp. The entire length

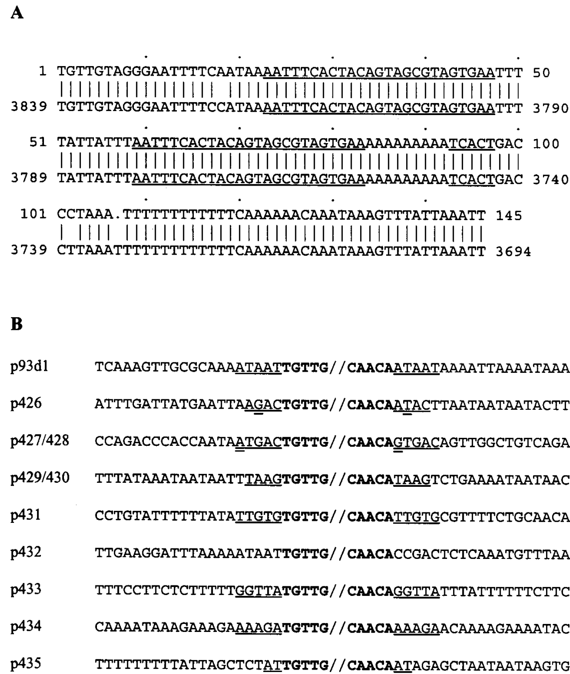

Figure 2. (**A**) Sequence comparison of Tdd-4 ITRs. Alignment of opposite strands from the left ITR (numbered 1→145) and the right ITR (numbered 3839→3694) is shown. The two ITRs differ by three bases. Contained within the ITRs are two perfect 25 bp direct repeats. Also, a five base sequence contains a portion of the same sequence found within these direct repeats. (**B**) Nine individual DNA sequences flanking Tdd-4 elements. Shown are 20 bases flanking each end of the element; the remainder of each flanking sequence from the designated plasmids can be accessed from the GenBank database (accession nos U63593–U63608). The five terminal nucleotides from each end of Tdd-4 are shown in bold and, where appropriate, the direct repeats that flank the Tdd-4 elements are underlined. Double underlines denote a nucleotide deviation in the direct repeats

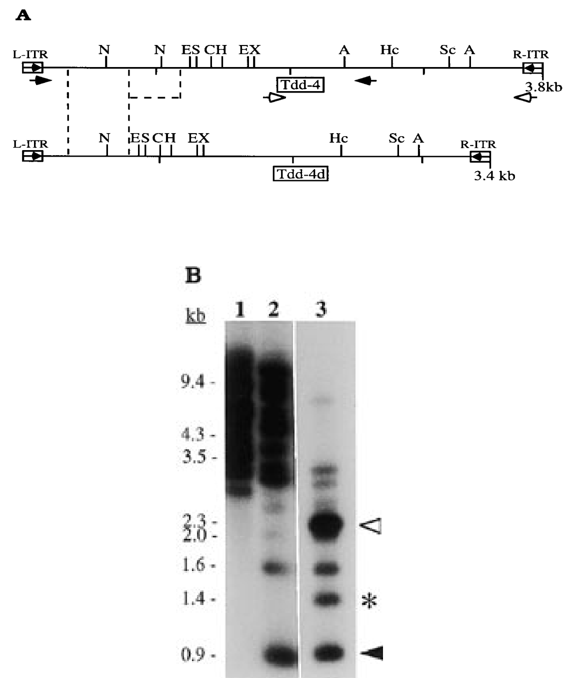

Figure 3. (**A**) Restriction map of Tdd-4 and Tdd-4d. The upper map represents the Tdd-4 element. Open and closed arrowheads represent relative locations of primer pairs used to identify a 411 bp deletion version of Tdd-4, called Tdd-4d. Dotted lines show locations of the 17 and 394 bp sequences which are absent in the Tdd-4d elements. Also, a base pair change in the Tdd-4d elements disrupted an *Acc*I restriction site (ATATAC in Tdd-4d elements compared to GTATAC in Tdd-4 elements). The restriction site symbols are: N, *Nsi*I; E, *Eco*RI; S, *Sac*I; C, *Cla*I; H, *Hin*dIII; X, *Xba*I; A, *Acc*I; Hc, *Hin*cII; Sc, *Sca*I. (**B**) Southern blot to discern variant forms of Tdd-4. AX4 DNA was digested with (1) *Nsi*I, (2) *Acc*I and (3) *Nsi*I+*Acc*I and probed with a 2.15 kb *Nsi*I–*Sca*I Tdd-4-specific probe. Bands present at 0.9 (closed triangle) and 1.4 kb (asterisk) in lane 3 correspond to the calculated sizes for Tdd-4 elements; the 2.3 kb band (open triangle) corresponds to the predicted size for Tdd-4d. The 1.6 kb band (lanes 2 and 3) is not accounted for by the maps of Tdd-4 or Tdd-4d and appears to be an *Acc*I site variant.

of these deleted elements was not sequenced; it is not known whether additional deletions or base changes exist. For clarity, the class of elements containing a DNA sequence identical to the copy from p93d1 was designated Tdd-4. The class of elements that contain the two deletions was termed Tdd-4d.

Since Tdd-4 and Tdd-4d contain differences in *Nsi*I and *Acc*I restriction sites, a Southern blot using AX4 DNA was produced to determine the relative abundance for each class of element (Fig. 3B). Bands present at 0.9 (*Acc*I digest) and 1.4 kb (*Nsi*I+*Acc*I digest) correspond to the calculated size of Tdd-4 elements, 933 and 1357 bp, respectively. A 2.3 kb band matches the predicted 2290 bp *Nsi*I–*Acc*I fragment from Tdd-4d elements. Additional *Nsi*I+*Acc*I digestion products are seen which are not accounted for from the restriction maps of either Tdd-4 or Tdd-4d. These additional fragments appear to result from differences in *Acc*I restriction sites, as determined by their presence in both the *Nsi*I+*Acc*I and *Acc*I digests. Based on hybridization intensity, the 3.8 kb Tdd-4 elements appear to be a less common form in the AX4 genome.

cDNA analysis

Since neither Tdd-4 nor Tdd-4d sequences yielded an extended ORF, this suggested RNA splicing may be involved to produce a functional ORF. Tdd-4 cDNAs were produced by RT–PCR of AX4 RNA. The resulting PCR products were cloned and sequenced. Four clones, designated B1, B2, B3 and B5, were produced. Each clone differed in some characteristic of its DNA sequence. Clones B2 and B3 correspond to the Tdd-4 class of elements, whereas partial sequence and restriction digests revealed that clones B1 and B5 would have originated from transcription of a Tdd-4d element. Compilation of the cDNA sequences for B2 and B3 shows that the Tdd-4 gene would contain six introns (Fig. 4); removal of these six introns produced ORFs of 2118 and 1938 bp, respectively. The variance between B2 and B3 was due to an alternative 5′ splicing of intron 6. Clone B2 used the proximal 5′ splice site (relative to the common 3′ splice site) while clone B3 used the distal 5′ splice site. The splice sites utilized in intron 6 are unusual for mRNA splicing. Introns 1–5 follow the GT-AG rule by possessing a GT terminal

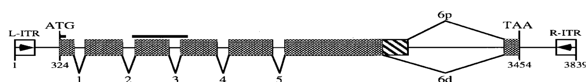

Figure 4. Organization of the Tdd-4 element. Exons are represented by shaded boxes, the alternative splice site by the hatched box and introns by the numbered V-shaped lines. Open boxes represent the left and right ITRs. Nucleotide positions of the putative start (ATG) and stop (TAA) codons are shown. Bars above the figure indicate the known regions that are deleted in Tdd-4d elements.

```
  1 MKQDKINEII NYLKFIEDNN FENTKHKNYK FLHDENRKEK LFRIEKKVMN
 51 CGGDRNISKE VCLEVINDYR MVSLWEEMHK GHIGRDATYG NYKTKYYNMG
101 LYSFVSDAVD TCDICQRNRI KGITKDFAPI VDTEEYSRLV YDLTSIKGEH
151 KEKVTYDDDN EKILTKLDDL IQYDSVQPYD TDVVYIILCI DSFTKFATGR
201 CLTTKRTVPI YNFLALTYFG KPVKVWHCDN GREFKNKVQK EFLKLFPGSK
251 SAHGAPRTPT TQGMVERLNR TIKERISKLK QQDFLDGTSR SLSELLKQAL
301 YDYNNTKTRT IKMKPSQAVG IVPLFINVQS EQDSQSIGVS DVSKEERTAI
351 ILENLTSYQN QWNSKPPKGL KVGDTVLFLE IKKNNKILIL CKIHKVIQED
401 TKQLYKLEFL EDGINSLQKK GLYSGFVGGN KLVLYKQSTV DIFRTSPNIQ
451 KDIDSFTSGL YLNNDGKVIE FLVQFIKNLG KDLSNFNLHQ LVTQNDPFRV
501 LEDALNNPSN IPPIINDNPF QHKLNNEMEQ EEILPFLNNN PTAPNLMNCL
551 RKDMNLKEIV PPIPQVQIIP SYTIPQSGSG IVTTVKRLRG RPHKIPIVNK
601 PPLKSHSKPS KSPLKSPSIS PSNSPFKSPF KSPSKSPFKS PSKSPSISPS
651 NSPFKSPSNS PFKSPFKSPS KCTRGNKARK NSIVCSVYSD MVNQAIEQLR
701 ELIKIN
```

Figure 5. Deduced protein products derived from Tdd-4 cDNAs. The Tdd-4 protein product that utilized the proximal 5′ splice site for intron 6 would contain 706 residues (shown above). Underlined residues represent an HHCC motif (residues 79–115), a DDE motif (residues 142–266) and a SPXX motif (residues 612–671). The Tdd-4 protein product that utilized the distal 5′ splice site for intron 6 would contain 646 residues and would not possess residues 612–671.

dinucleotide at the 5′ splice site and an AG terminal dinucleotide at the 3′ splice site (Table 1). The exact intron–exon junctions for intron 6 were not defined since both the alternate 5′ splice sites as well as the 3′ splice site contain the same 5′-CAAAT-3′ pentanucleotide sequence. The cDNA clones contain a single copy of this pentanucleotide sequence compared to the dual copies present at the intron ends in the genomic DNA sequence.

Table 1. Tdd-4 intron splice sites

Intron	5′ Splice site[a]	3′ Splice site[a]	Intron size (bases)
1	TGA-GTAAATAT	AAAAACAG-AAA	60
2	AAG-GTTTGGTA	ATTAATAG-GTA	79
3	AAG-GTATTTAT	TTTTTTAG-ATG	67
4	AGA-GTATGATA	AATTTTAG-TTT	68
5	AAT-GTAAGTAA	TCTAATAG-AAA	76
6d	ATCCAAATCAC[b]		842
		CATCAAATGCA[b]	
6p	TTCCAAATCAC[b]		662

[a]Splice sites are indicated by hyphens. Intron boundaries are right and left of hyphens for the 5′ and 3′ splice sites, respectively.
[b]The exact splice junction for intron 6 was not defined. The underlined five base sequence is present at the distal (d) and proximal (p) 5′ splice sites and the 3′ splice site.

The deletions present in the Tdd-4d class of elements combined with alterations in splicing would yield nonsense mutations to the Tdd-4d transcripts. Although only partial sequences were obtained from clones B1 and B5, these clones were confirmed as cDNA products since B1 used the proximal 5′ splice site for intron 6 and B5 used the distal splice site. Sequence analysis of B1 and restriction analysis of B5 showed that intron 4 was retained in both clones. The two deletions in Tdd-4d elements along with the retention of intron 4 would prevent the production of an extended ORF; the Tdd-4d elements were therefore classified as defective elements.

Once the Tdd-4 ORFs were known, regulatory regions of the DNA sequence could be predicted. The first ATG (nt 324–326) is preceded by three A bases, which is in accordance with other translatable *Dictyostelium* genes (39). Upstream of these A bases, oligo(dT) stretches are present between nt 180 and 311. Oligo(dT) stretches have been shown to lie within promoter regions of *Dictyostelium* translatable genes (39). The opposite end of the coding sequence contains two tandem stop codons at positions 3454–3459. The consensus polyadenylation signal (AATAAA) follows one base downstream from the second stop codon (three additional consensus polyadenylation signal sequences lie within 90 bases downstream of this first sequence).

Open reading frames encoded by Tdd-4

The ORFs derived from the two Tdd-4 cDNAs predict protein products of 706 and 646 residues (Fig. 5). The difference between these proteins result from the alternative splicing of intron 6. Four regions of significance are predicted for the 706 residue protein. These regions are based on homologies to integrase proteins and to other DNA-binding proteins that contain a related amino acid sequence.

HHCC motif. The deduced amino acid sequence for the N-terminal region of the two Tdd-4 encoded proteins suggests that an HHCC motif exists between residues 79 and 115 and contains the sequence $H-X_2-H-X_{29}-C-X_2-C$. The consensus HHCC motif sequence for integrase is $H-X_{(3–7)}-H-X_{(23–32)}-C-X_2-C$ (17). BLAST (GenBank) and BLITZ (SwissProt) searches that utilized this Tdd-4 HHCC region produced highest similarities to the integrase region of LTR retroelement *pol* gene products (Fig. 6A).

DDE motif. A DDE motif appears to be present between residues D142, D229 and E266 in both Tdd-4 encoded proteins. BLITZ searches gave highest scores ($P = 10^{-12}$) to LTR retroelement integrase sequences (Fig. 6B). The spacing between D229 and E266 is similar to the integrase from the *gypsy* retrotransposon in that it contains a D36E spacing rather than the consensus D35E integrase spacing. The assignment of the first D of the Tdd-4 DDE motif to residue 142 is speculative. This D is in agreement with the consensus sequence (11) by having a threonine two residues C-terminal to D142 and D142 lies 26 residues C-terminal to the HHCC motif, which is in line with the 20–29 residue spacing between the HHCC motif and the first D for retroviral integrase (12). The 86 residue spacing between D142 and D229 is greater than the 51–59 spacing of the consensus integrase sequence, but conforms with the D-to-D spacing of transposases that catalyze cut-and-paste transposition. The Tdd-4 DDE motif also contains the two basic residues (R270 and K273) which are in identical

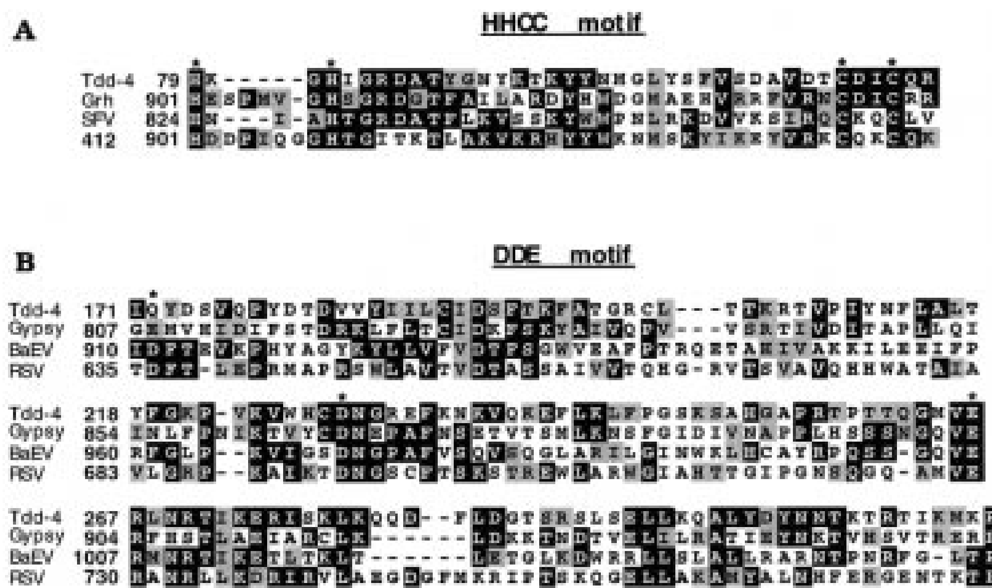

Figure 6. Comparison of the Tdd-4 HHCC and DDE motifs to integrase proteins. The three highest scoring integrase sequences, derived from BLAST or BLITZ searches, were aligned with Tdd-4 motifs. Alignments were performed using the Clustal W program and 50% or greater regions of similarity (shaded) or identity (black) were indicated using BOXSHADE. Asterisks indicate conserved residues for the respective motif (asterisk placement of the first D of the DDE motif corresponds to the conserved D residue for BaEV or RSV integrase). Numbers designate residue positions in the Tdd-4 encoded protein sequence or the integrase region from the respective *pol* polyprotein. (**A**) Members used for the HHCC alignment together with their accession no./SwissProt ID name are as follows: *grasshopper* (Grh) retrotransposon from *Magnaporthe grisea* (M77661); simian foamy virus (POL_SFV1); 412 retrotransposon from *Drosophila melanogaster* (POL4_DROME). (**B**) Members used for the DDE alignment together with their SwissProt ID name are as follows: *gypsy* retrotransposon from *D.melanogaster* (POLY_DROME); baboon endogenous virus (POL_BAEVM); Rous sarcoma virus (POL_RSVP).

positions with respect to the conserved glutamate (E266) when compared to an alignment of retroviral integrases and prokaryotic transposases (40).

Proline-rich region. C-terminal to the DDE motif, a proline-rich region has been identified within non-lentiviral integrase sequences (41). Clusters of proline residues are also present (residues 508–602) in both Tdd-4 encoded proteins. In particular, a PPIP sequence is present in both Tdd-4 encoded proteins (residues 561–564) as well as the proline-rich C-terminal region of HTLV-II integrase. The PPIP sequence resembles the PXXP consensus motif that serves as a ligand-binding site to SH3 domains (42).

SPXX motif. The binding preference of SPXX repeats is reported to be in the minor groove of A/T-rich DNA (44–46). These repeats are so-named since a serine or threonine residue occupies the first position, a proline residue occupies the second position and the third and fourth positions may contain any (X) residue (43). The portion of the 706 residue Tdd-4 protein that likely equates to the C-terminal DNA-binding region of integrase is the 15 SPXX repeats between residues 612 and 671. This region also corresponds to the alternate splice site; the proximal 5′ splice site produces the SPXX motif whereas the distal 5′ splice site eliminates the motif.

DISCUSSION

Tdd-4 is the sixth transposable element to be isolated from *Dictyostelium* and the only element thus far derived from this organism to contain structural, coding and mobility features of a DNA element. Tdd-4 is transpositionally active, as evidenced by its insertion into a target plasmid and its mobility displayed in the RFLP analysis. To account for the loss of hybridizing fragments in the RFLP patterns of Figure 1, Tdd-4 is presumed to have transposed by excision of an element from the donor site. The cut-and-paste mode of transposition utilized by the P element of *Drosophila* (47), Tc1 of *Caenorhabditis elegans* (48) and the bacterial transposon Tn*7* (49) suggest that the preferred pathway for repair of the excised element from the donor site would be accomplished by homologous recombination. Use of a homologous sequence would regenerate the transposable element at the donor site and give the appearance of a replicative transposition event. The Tdd-4 RFLP analysis implies that transposition can be correlated with disappearance of an element from a previously existing location, thus the DNA repair following Tdd-4 excision does not appear to use a homologous sequence for gap repair. However, this interpretation selects for only those transposition events in which loss of an element has occurred. Considering that a deleted version of Tdd-4 was identified in this study (the Tdd-4d class of elements) and since deletion versions of intact elements are believed to arise via an aberrant gap repair, some degree of recombination is likely involved during the Tdd-4 gap repair process.

The Tdd-4 ITR is notable by containing a nucleotide similarity to a retroelement LTR and by possessing two direct repeats within each Tdd-4 ITR. Integrated LTR retroelement DNA sequences possess a short ITR sequence ranging from the conserved terminal dinucleotides up to an imperfect repeat of 16 bp in

Moloney murine leukemia virus (50). Interestingly, the terminal nucleotides (5′-TGTTGTA…TACAACA-3′) from Tdd-4 are identical to the terminal 7 nt of the 8 bp ITR from the Ty3 retrotransposon of *Saccharomyces cerevisiae* (51). A second feature is the two 25 bp direct repeats contained within the Tdd-4 ITR sequence. The ends of a few other transposons contain repeated sequences that serve as transposase binding sites (11). In particular, the *C.elegans* Tc3 transposon contains two repeated transposase binding sites within its 462 bp ITRs, with 29 of the 36 bases being identical in these two binding sites (52). Analogously, the two 25 bp direct repeats in the Tdd-4 ITRs likely act as the binding sites for the Tdd-4 transposase.

Moving inward from the ITRs, the coding region of Tdd-4 is produced by removal of six introns. Introns 1–5 of the Tdd-4 transcripts conform to the GT-AG rule for splicing of pre-mRNAs. Curiously, Tdd-4 intron 6 deviates from this rule by containing a 5′-CAAAT-3′ sequence at both alternative 5′ splice sites and the 3′ splice site. This repeat prevents the exact intron 6 splice junctions being defined from the available data. Reports have demonstrated a minor class of pre-mRNA introns, called AT-AC introns, which contain an AT dinucleotide at the 5′-end and an AC dinucleotide at the 3′-end (53,54). A requirement of pre-mRNA splicing involves a non-Watson–Crick interaction between the first and last nucleotides of an intron, forming a G-G pair for GT-AG introns and an A-C pair for AT-AC introns (55). In addition, it was proposed that a non-Watson–Crick A–A interaction is possible for an intron containing terminal A nucleotides (56). In support of this proposal, G→A mutation of the terminal nucleotides from a GT-AG intron allows splicing to occur with the doubly mutant AT-AA intron ends (57,58). Combining information of the AT-AC introns with the terminal dinucleotides from the mutant AT-AA introns may explain the splicing of intron 6 in Tdd-4 transcripts. Splitting the CAAAT sequence into constituent segments of CAA and AT would allow an AT-AA intron to be formed from the CAAAT sequence. This proposed AT-AA intron of intron 6 shows a better match to a consensus sequence from AT-AC introns compared to GT-AG introns. Five of the eight nucleotides from the proposed intron 6 distal splice site (/ATcaCCTc) or the proximal splice site (/ATcaCCaT) agree with the consensus 5′ splice site (/ATATCCTT) for AT-AC introns (56,59). Two of the three nucleotides from the proposed intron 6 3′ splice site (CAa/) agree with the consensus 3′ splice site (CAC/) for AT-AC introns. Additional experimentation will be necessary to confirm the proposed intron 6 splice sites.

The Tdd-4 class of elements can produce, by alternative splicing, two ORFs of 2118 and 1938 bp. For other transposons that produce two proteins from a single gene (60,61), the full-length protein functions as the transposase and the truncated protein plays a role in the regulation of transposition. By analogy, the 2118 bp ORF and the 1938 bp ORF presumably encode transposase and inhibitor proteins, respectively. The difference between these proteins results from the presence or absence of the SPXX motif, which emphasizes the significance of this motif. Binding of the SPKK tetrapeptide to DNA has been shown to affect base pair hydrogen bonding and base stacking interactions leading to a destabilization of the DNA duplex structure (46). Relatedly, DNA-binding proteins that induce structural changes to DNA have been shown to affect the efficiency and selection sites of retroviral integration (62–65). The SPXX motif of the Tdd-4 transposase may create a similar structural deformation to the DNA double helix and obviate the need of host proteins to perform this function.

In summary, Tdd-4 is a unique DNA transposon by possessing amino acid and nucleotide similarities to LTR retroelements. Since an HHCC motif has previously not been identified in a transposase, combined with the high degree of similarity to the integrase DDE motif plus a proline-rich C-terminal region similar to non-lentiviral integrases and the 7 bp terminal nucleotide identity to the Ty3 retrotransposon, together these features suggest an evolutionary relationship between Tdd-4 and an LTR retroelement lineage.

ACKNOWLEDGEMENTS

I thank J. Hughes for sharing data of the plasmid maintenance experiments. Additional thanks to G. Podgorski for use of laboratory facilities and T. Doak for helpful discussions regarding protein sequence analysis. This work was supported by grants to D. Welker from the Utah State University Biotechnology Center and NSF grant DMB8915591.

REFERENCES

1 Finnegan,D.J. (1992) *Curr. Opin. Genet. Dev.*, **2**, 861–867.
2 LaFemina,R.L., Callahan,P.L. and Cordingley,M.G. (1991) *J. Virol.*, **65**, 5624–5630.
3 Leavitt,A.D., Rose,R.B. and Varmus,H.E (1992) *J. Virol.*, **66**, 2359–2368.
4 Vink,C., van Gent,D.C., Elgersma,Y. and Plasterk,R.H.A. (1991) *J. Virol.*, **65**, 4636–4644.
5 Gary,P.A., Biery,M.C., Bainton,R.J. and Craig,N.L. (1996) *J. Mol. Biol.*, **257**, 301–316.
6 Rowland,S.J., Sherratt,D.J., Stark,W.M. and Boocock,M.R. (1995) *EMBO J.*, **14**, 196–205.
7 Polard,P., Prere,M.F., Fayet,O. and Chandler,M. (1992) *EMBO J.*, **11**, 5079–5090.
8 Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 942–946.
9 Fayet,O., Ramond,P., Polard,P., Prere,M.F. and Chandler,M. (1990) *Mol. Microbiol.*, **4**, 1771–1777.
10 Kulkosky,J., Jones,K.S., Katz,R.A, Mack,J.P.G. and Skalka,A.M. (1992) *Mol. Cell. Biol.*, **12**, 2331–2338.
11 Polard,P. and Chandler,M. (1995) *Mol. Microbiol.*, **15**, 13–23.
12 Engelman,A. and Craigie,R. (1992) *J. Virol.*, **66**, 6361–6369.
13 van Gent,D.C., Oude Groenger,A.A.M. and Plasterk,R.H.A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 9598–9602.
14 Kirchner,J. and Sandmeyer,S.B. (1996) *J. Virol.*, **70**, 4737–4747.
15 van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1994) *Cell*, **79**, 293–301.
16 Bolland,S. and Kleckner,N. (1996) *Cell*, **84**, 223–233.
17 Khan,E., Mack,J.P.G., Katz,R.A, Kulkosky,J. and Skalka,A.M. (1991) *Nucleic Acids Res.*, **19**, 851–860.
18 Capy,P., Vitalis,R., Langin,T., Higuet,D. and Bazin,C. (1996) *J. Mol. Evol.*, **42**, 359–368.
19 Rice,P., Craigie,R. and Davies,D.R. (1996) *Curr. Opin. Struct. Biol.*, **6**, 76–83.
20 Puras Lutzke,R.A., Vink,C. and Plasterk,R.H.A. (1994) *Nucleic Acids Res.*, **22**, 4125–4131.
21 Vink,C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1993) *Nucleic Acids Res.*, **21**, 1419–1425.
22 Woerner,A.M. and Marcus-Sekura,C.J. (1993) *Nucleic Acids Res.*, **21**, 3507–3511.
23 Leng,P., Klatte,D.H., Schumann,G., Boeke,J.D. and Steck,T.L. (1998) *Nucleic Acids Res.*, **26**, 2008–2015.
24 Rosen,E., Sivertsen,A. and Firtel,R.A. (1983) *Cell*, **35**, 243–251.
25 Cappello,J., Handelsman,K. and Lodish,H.F. (1985) *Cell*, **43**, 105–115.
26 Poole,S.J. and Firtel,R.A. (1984) *Mol. Cell. Biol.*, **4**, 671–680.
27 Winckler,T., Tschepke,C., de Hostos,E.L., Jendretzke,A. and Dingermann,T. (1998) *Mol. Gen. Genet.*, **257**, 655–661.
28 Marschalek,R., Hofmann,J., Schumann,G., Gosseringer,R. and Dingermann,T. (1992) *Mol. Cell. Biol.*, **12**, 229–239.

29 Nellen,W., Silan,C. and Firtel,R.A. (1984) *Mol. Cell. Biol*., **4**, 2890–2898.
30 Podgorski,G.J. and Deering,R.A. (1980) *Mutat. Res*., **74**, 459–468.
31 Welker,D.L., Hirth,K.P. and Williams,K.L. (1985) *Mol. Cell. Biol*., **5**, 273–280.
32 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning*: *A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
33 Ochman,H., Gerber,A.S. and Hartl,D.L. (1988) *Genetics*, **120**, 621–623.
34 Hughes,J.E., Podgorski,G.J. and Welker,D.L. (1992) *Plasmid*, **28**, 46–60.
35 Nellen,W., Silan,C., Saur,U. and Firtel,R.A. (1986) *EMBO J*., **5**, 3367–3372.
36 Katz,R.A., Merkel,G., Kulkosky,J., Leis,J. and Skalka,A.M. (1990) *Cell*, **63**, 87–95.
37 Bushman,F.D., Fujiwara,T. and Craigie,R. (1990) *Science*, **249**, 1555–1558.
38 Craigie,R., Fujiwara,T. and Bushman,F. (1990) *Cell*, **62**, 829–837.
39 Kimmel,A. and Firtel,R.A. (1983) *Nucleic Acids Res*., **11**, 541–552.
40 Jenkins,T.M., Esposito,D., Engelman,A. and Craigie,R. (1997) *EMBO J*., **16**, 6849–6859.
41 Cannon,P.M., Byles,E.D., Kingsman,S.M. and Kingsman,A.J. (1996) *J. Virol*., **70**, 651–657.
42 Cohen,G.B., Ren,R. and Baltimore,D. (1995) *Cell*, **80**, 237–248.
43 Suzuki,M. (1989) *J. Mol. Biol*., **207**, 61–84.
44 Bailly,F., Bailly,C., Waring,M.J. and Henichart,J.P. (1992) *Biochem. Biophys. Res. Commun*., **184**, 930–937.
45 Churchill,M.E.A. and Suzuki,M. (1989) *EMBO J*., **8**, 4189–4195.
46 Takeuchi,H. and Sasamori,J. (1995) *Biopolymers*, **35**, 359–367.
47 Engels,W.R. (1996) In Saedler,H. and Gierl,A. (eds), *Current Topics in Microbiology and Immunology: Transposable Elements*. Springer-Verlag, Berlin, Germany, Vol. 204, pp. 103–123.
48 Plasterk,R.H.A (1996) In Saedler,H. and Gierl,A. (eds), *Current Topics in Microbiology and Immunology: Transposable Elements*. Springer-Verlag, Berlin, Germany, Vol. 204, pp. 125–143.
49 Craig,N.L. (1996) In Saedler,H. and Gierl,A. (eds), *Current Topics in Microbiology and Immunology: Transposable Elements*. Springer-Verlag, Berlin, Germany, Vol. 204, pp. 27–48.
50 Kulkosky,J. and Skalka,A.M. (1990) *J. Acquir. Immune Defic. Syndr*., **3**, 839–851.
51 Sandmeyer,S.B. and Menees,T.M. (1996) In Krausslich,H.G. (ed.), *Current Topics in Microbiology and Immunology: Morphogenesis and Maturation of Retroviruses*. Springer-Verlag, Berlin, Germany, Vol. 214, pp. 261–296.
52 Colloms,S.D., van Luenen,H.G.A.M. and Plasterk,R.H.A. (1994) *Nucleic Acids Res*., **22**, 5548–5554.
53 Hall,S.L. and Padgett,R.A. (1996) *Science*, **271**, 1716–1718.
54 Tarn,W.Y. and Steitz,J.A. (1996) *Cell*, **84**, 801–811.
55 Umen,J.G. and Guthrie,C. (1995) *RNA*, **1**, 869–885.
56 Mount,S.M. (1996) *Science*, **271**, 1690–1692.
57 Carothers,A.M., Urlaub,G., Grunberger,D. and Chasin,L.A. (1993) *Mol. Cell. Biol*., **13**, 5085–5098.
58 Chanfreau,G., Legrain,P., Dujon,B. and Jacquier,A. (1994) *Nucleic Acids Res*., **22**, 1981–1987.
59 Hall,S.L. and Padgett,R.A. (1994) *J. Mol. Biol*., **23**, 357–365.
60 de la Cruz,N.B., Weinreich,M.D., Wiegand,T.W, Krebs,M.P. and Reznikoff,W.S. (1993) *J. Bacteriol*., **175**, 6932–6938.
61 Misra,S. and Rio,D.C. (1990) *Cell*, **62**, 269–284.
62 Bor,Y.C., Bushman,F.D. and Orgel,L.E. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 10334–10338.
63 Muller,H.P. and Varmus,H.E. (1994) *EMBO J*., **13**, 4704–4714.
64 Pruss,D., Bushman,F.D. and Wolffe,A.P. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 5913–5917.
65 Pruss,D., Reeves,R. Bushman,F.D. and Wolffe,A.P. (1994) *J. Biol. Chem*., **269**, 25031–25041.

NOTE ADDED IN PROOF

While this manuscript was being processed for publication, a minireview was published [Wu,Q. and Krainer,A.R. (1999) *Mol. Cell. Biol*., **19**, 3225–3236] that has compiled three naturally occurring AT-AA introns.