Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 3574–3578, May 1990
Genetics

Characterization of the restriction site of a prokaryotic intron-encoded endonuclease

(T4 bacteriophage/*td* intron open reading frame/hybridization/dideoxynucleotide sequencing/gene conversion)

FREDERICK K. CHU*†‡, GLADYS MALEY*†, JOAN PEDERSEN-LANE*, AI-MIN WANG*, AND FRANK MALEY*†

*Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, NY 12201-0509; and †Department of Biomedical Sciences, School of Public Health, Albany, NY 12237

Communicated by Henry Lardy, February 26, 1990

ABSTRACT **The 1016-base-pair (bp) intron in the T4 bacteriophage thymidylate synthase gene (*td*) contains a 735-bp open reading frame that encodes a protein product with endonucleolytic activity. The endonuclease shows specificity for the intronless form of the *td* gene. Highly purified endonuclease cleaves the DNA of the intronless form of the *td* gene *in vitro* at 24 bp upstream of the exon 1–exon 2 junction, generating a 2-base staggered cut with 3′-hydroxyl overhangs. Although the endonuclease cleaves in exon 1, it requires some exon 2 sequence for recognition. The maximum recognition sequence lies in an 87-bp stretch, from 52 bp upstream to 35 bp downstream of the cleavage site, ending at 11 bp into exon 2. The *td* intron endonuclease appears involved in the conversion of the intronless form of *td* to intron-containing *td* gene in the T-even phages. A role for intron mobility is discussed.**

Introns are ubiquitous in the eukaryote kingdom, but only a few have been found in prokaryotic genes. Apart from the presence of small introns in some tRNA genes in archaebacteria (1), three introns of prokaryotic origin have been found in the T-even phages (2–8), and one such intron has been found in the DNA polymerase gene of the *Bacillus subtilis* phage SPO1 (9). Three genes in the T4 phage—*td* (2–4), *nrdB* (5, 6), and *sunY* (7, 8)—have been shown to be interrupted by a group I intron that contains an open reading frame, in addition to specific sequence determinants characteristic of group I introns (10, 11), described in *Tetrahymena thermophila* and later in the mitochondria of many yeasts and fungi (for review, see ref. 12).

Group I introns, including those in the T4 phage, are able to self-excise *in vitro* concomitant with exon ligation in the presence of a guanosine cofactor and Mg^{2+} (13, 14). Unlike some yeast and fungal introns, the reading frames in these phage introns do not encode mRNA maturases (15–17) that facilitate intron excision at the RNA level. However, similar to other yeast and fungal introns, the phage *td* intron open reading frame (iORF) encodes an endonuclease (18) that has been genetically implicated in promoting intron mobility among the T-even phage *td* genes (19, 20).

The best characterized intron endonuclease to date is that derived from the group I intron (ω^+) of the mitochondrial 21S rRNA gene of *Saccharomyces cerevisiae* (21, 22), where the product of a 235-codon-long iORF, termed ω transposase, was shown to effect a double-strand break in the intronless (ω^-) copy of the same gene (21). This cut site consists of a 4-base staggered cut with 3′-hydroxyl overhangs and occurs at the junction between two exons where the intron is eventually inserted. The recognition sequence for this enzyme is shown to extend over 18 base pairs (bp), centered around the cleavage site. For the T4 phage *td* gene, the protein produced from a 735-bp open reading frame (3) in its 1016-bp intron (2) also possesses double-stranded endonucleolytic activity specific for the intronless copy of the *td* gene (*tdΔI*) (18, 20). With highly purified preparations of the *td* intron endonuclease, we have determined its cleavage site to be upstream of the exon 1–exon 2 junction. In this paper, we show that the endonuclease effects a cleavage that is centered at 24 bp upstream of the exon-fusion site in the *tdΔI* DNA, creating a staggered cut with a 2-base hydroxyl overhang at the 3′ end of each DNA strand; the endonuclease recognition sequence spans an 87-bp region, from 52 bp upstream to 35 bp downstream of the cleavage site in the *tdΔI* gene.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, Enzymes, and Chemicals. *Escherichia coli* strains DH5αF′, JM103, and TB-1 were from Bethesda Research Laboratories. Plasmid pET3c, CE6λ phage, and their respective hosts, HMS174 and ED8739, were supplied by F. W. Studier (Brookhaven National Laboratory, Upton, NY) (23, 24). Recombinant plasmids pK*tdΔI* and pUC*tdΔI*, containing *tdΔI*, have been described (see refs. 25 and 18, respectively). The following enzymes were obtained from United States Biochemical and used according to directions: T4 DNA ligase, T4 polynucleotide kinase, and T7 DNA polymerase. Restriction enzymes were purchased from several suppliers and used according to the suppliers' instructions. [γ-^{32}P]ATP (>5000 Ci/mmol; 1 Ci = 37 GBq) and deoxyadenosine [α-[^{35}S]thio]triphosphate (>4000 Ci/mmol) were from Amersham. The oligodeoxynucleotides were chemically synthesized by using an Applied Biosystems model 381A DNA synthesizer.

Preparation of iORF Protein. Induced cells [*E. coli* strain HMS174 harboring the pETdiORF (pETdIrf) plasmid] were processed, as described by West *et al.* (18). The endonuclease released from the inclusion bodies was renatured according to the procedure of P. C. Babbitt (University of California at San Francisco; personal communication). This entailed solubilization of the isolated inclusion bodies in 6 M guanidine hydrochloride, followed by resolution on a Bio-Gel column in 4 M guanidine hydrochloride and refolding by dialysis against 20 mM potassium phosphate, pH 7.5/0.5 M NaCl/20 mM 2-mercaptoethanol. Purification to near homogeneity was obtained by chromatography on a column of phenyl-Sepharose. A detailed description of the purification procedure will be presented elsewhere.

Analysis for Endonuclease Activity. The reaction buffer contained 50 mM Tris·HCl (pH 7.5), 10 mM $MgCl_2$, 100 mM

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: iORF, intron open reading frame (iORF protein and endonuclease are used interchangeably in this work); *tdΔI*, intronless copy of the *td* gene.

‡To whom reprint requests should be addressed at: Wadsworth Center for Laboratories and Research, New York State Department of Health, Empire State Plaza, Albany, NY 12201.

NaCl, and 10 mM dithiothreitol. The reaction was performed in a volume of 10 μl containing 0.1 pmol of DNA and 2 μg of iORF protein at 23°C for 5–10 min, stopped by the addition of 1 μl of 0.4 M Na_2EDTA (pH 7.5) and then chilled on ice. The products were electrophoresed in 0.8% agarose slab gel in TBE buffer (0.1 M Trizma base/0.1 M boric acid/2 mM Na_2EDTA) and visualized by ethidium bromide staining.

Blotting and Hybridization. After electrophoresis, the DNA in the agarose slab gel was denatured in solution A (1.5 M NaCl/0.5 M NaOH) at 23°C for 30 min, neutralized in solution B (0.5 M Tris·HCl, pH 7.2/1.5 M NaCl/0.001 M Na_2EDTA) at 23°C for 30 min, and then transferred onto Hybond-N membrane (Amersham) by capillary blotting at 23°C for 16 hr. The membrane was then rinsed briefly in distilled water and allowed to air dry. The transferred DNA was fixed to the membrane by UV irradiation for 5 min. Hybridization with ^{32}P-labeled oligodeoxynucleotides (1×10^5 cpm/ml of hybridization mixture) was carried out in solution C (0.05 M sodium phosphate, pH 7.7/0.9 M NaCl/0.5 mM Na_2EDTA/5× Denhardt's solution/0.5% SDS) (1× Denhardt's solution is 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin) at 60°C for 2 hr and then cooled to room temperature. The membrane was washed at 23°C, twice with solution D (solution C without Denhardt's solution and SDS) and once with 0.2× solution D, each time for 20 min. The membrane was then air dried and subjected to autoradiography.

DNA Sequencing. The dideoxynucleotide chain-termination method (26) was used for sequencing of double-stranded DNA templates. The reagents for DNA sequencing with Sequenase were purchased from United States Biochemical and used according to the supplier's protocol. Where unlabeled primer was used, deoxyadenosine [α-[^{35}S]thio]triphosphate was employed. Where ^{32}P-labeled (5′) primer was used, deoxyadenosine [α-[^{35}S]thio]triphosphate was replaced by dATP.

Restriction of cDNA Products. To determine the restriction of cDNA products by the endonuclease, half of the sequencing reaction (≈3 μl) was heated at 65°C for 10 min to inactivate the Sequenase and then slow-cooled (≈30 min) to room temperature to allow renaturation of DNA complementary strands. After 10-sec centrifugation in a microcentrifuge to collect moisture, 2 μg of the endonuclease was added, followed by 5-min incubation at 23°C. A one-tenth volume of 0.4 M Na_2EDTA (pH 7.5) was added, and the incubated mixture was heated at 80°C for 3 min and then subjected to electrophoretic analysis in 8% polyacrylamide/6 M urea gel.

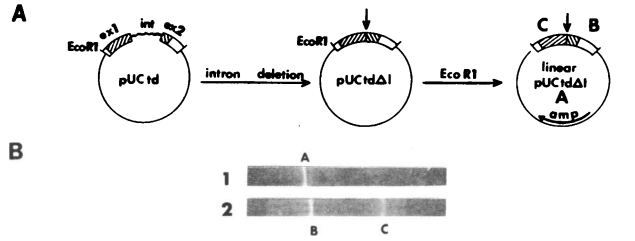
RESULTS

Localization of the Restriction Site of *td* Intron-Encoded Endonuclease. The 735-bp iORF of the T4 phage *td* gene has been shown (3) to encode a protein endonuclease that effects a specific double-strand break at or near the exon junction in *td*ΔI (18, 20). To define the limits of the break, we have synthesized several oligodeoxynucleotides and labeled their 5′ ends with ^{32}P. The radioactive oligodeoxynucleotides, complementary in sequence to regions progressively upstream and downstream of the exon 1–exon 2 junction (see Table 1 for their size, designation, and location), were used as hybridization probes for the endonuclease-generated products of *td*ΔI DNA. If the endonuclease cuts the *Eco*RI-linearized intronless pUC*td*ΔI DNA [4.26-kilobase-pair (kbp) fragment A] at the junction of exon 1 and exon 2, the resulting products, consisting of a 3.5-kbp exon 2-containing fragment B and a 0.76-kbp exon 1-containing fragment C (Fig. 1 *A* and *B*), should not hybridize to the exon 1–exon 2 junction sequence represented by probe 1. To our surprise, fragment B, and not C, hybridized strongly to probe 1 (Fig. 1*C*), indicating the presence of some exon 1 3′-end sequence in the larger exon 2-containing fragment. This presence was confirmed by the observed hybridization of fragment B to probe 5, representing the last 21 bases in exon 1 and corroborated by hybridization data showing that the 3′ end of exon 1 (at least 21 bp in length) is absent in the smaller exon 1-containing fragment C. On the other hand, fragment B contains, in addition to the 3′ end of exon 1, the entire exon 2, as determined by its hybridization to probes 6, 7, and other

Table 1. Synthetic oligodeoxynucleotides

Probe designation	Exon location	Size	Nucleotide position*
1	Exon 1–2†	18	541–558
2‡	Exon 1	18	453–470
3	Exon 1	13	490–502
4	Exon 1	21	516–536
5	Exon 1	21	529–549
6	Exon 2	13	555–567
7	Exon 2	16	591–606

*Nucleotides are numbered from the initiation codon (ATG) in *td*ΔI, which is 858 bp in length (2, 25).
†Exon 1 and 2 designation pertain to the exon topography in the intron-containing *td* gene, in which exon 1 runs from nucleotide 1 through 549, exon 2 runs from nucleotide 550 through 858, and the exons are separated by a 1016-bp intron (2, 27).
‡This oligodeoxynucleotide contains the codon (mRNA) sequence, in contrast to the others, which all contain anticodon (anti-mRNA) sequences.



C

Oligodeoxynucleotide probe	EcoRI–linearized pUC18tdΔI(A) 4.26 kbp	td intron → endonuclease	Ex2–containing fragment(B) 3.5 kbp	+	Ex1–containing fragment(C) 0.76 kbp
1*	+†		+		–
2	+		–		+
3	+		–		+
5	+		+		–
6	+		+		–
7	+		+		–

FIG. 1. Mapping of the cleavage site: hybridization analysis. (*A*) Schematic diagram showing the generation of linear double-stranded DNA substrate (pUC*td*ΔI) for the *td* intron-encoded endonuclease. Short arrow at exon junction indicates the anticipated endonucleolytic site; ex, exon; int, intron; amp, ampicillin resistance gene. (*B*) Products of cleavage by the endonuclease. Lane 1 shows the *Eco*RI-linearized pUC*td*ΔI DNA substrate (fragment A), and lane 2 shows the resulting products of the endonuclease (fragments B and C). (*C*) Oligonucleotide hybridization analysis of the cleavage site. Six sets of substrate and products shown in *B* were resolved by electrophoresis in an 0.8% agarose horizontal slab gel. They were then blotted onto Hybond-N membrane (Amersham), and each subset was subjected to hybridization analysis by using a specific oligodeoxynucleotide labeled with ^{32}P in its 5′ end as probe. The size of *td*ΔI DNA fragments was estimated from a marker lane containing *Hin*dIII-restricted λ DNA fragments. Ex1 and Ex2, exon 1 and 2, respectively.
*See Table 1 for description of oligodeoxynucleotide probes.
†+, Hybridizing; –, nonhybridizing.

exon 2 probes not shown. Analysis of the 3′ limit of fragment C and 5′ limit of fragment B by hybridization to exon 1 probes (probes 2 and 3 in Table 1) corresponding to nucleotide sequences further upstream of the exon junction suggests the site of endonucleolytic cleavage to fall between bp 502 (present in fragment C) and 529 (present in fragment B), which specify codons 168–176 in the *td*ΔI gene. Another recombinant plasmid pK*td*ΔI (25), constructed by cloning the *td*ΔI gene fragment into pKC30 (28), exhibits identical probe-specific hybridization characteristics in its products after treatment with the *td* intron endonuclease (data not shown).

The *td* Intron Endonuclease Generates a 2-Base Staggered Cut with 3′-Hydroxyl Overhangs Centered at 24 bp Upstream of the Exon Junction. To determine the precise point of cleavage by the *td* intron endonuclease, we synthesized 5′-end-labeled double-stranded substrates *in vitro* by priming the double-stranded DNA from recombinant plasmid pUC*td*ΔI with 5′-end-labeled oligodeoxynucleotide (probes 2 and 7 in Table 1). The cDNA was synthesized in the presence of the 4 dNTPs (mixture of dATP, dCTP, dGTP, and dTTP) with and without the individual dideoxynucleotide chain terminators. Probe 2, which contains exon 1 sequence located at 32 bases upstream of bp 502, uses the coding strand as template and is elongated in the exon 2 direction, whereas probe 7 containing exon 2 sequence at 62 bases downstream of bp 529 copies the noncoding strand towards the exon 1 region. In each case, the double-stranded DNA product, consisting of an unlabeled template strand and a ^{32}P-labeled cDNA strand produced in the absence of dideoxynucleotide chain terminators, was used as substrate for the endonuclease. The products were denatured and electrophoresed in an 8% polyacrylamide/6 M urea gel alongside the respective sequence ladder (the four dideoxynucleotide-containing cDNA reactions), which was used also as a size marker. Fig. 2*A* shows the endonucleolytic cleavage site on the noncoding (mRNA-like) DNA strand to be right of bp 526, yielding a 74-base product; Fig. 2*B* reveals this site to be left of bp 525 on the coding (mRNA-complementary) strand, giving an 82-base product. Taken together, the sequencing results allow us to conclude that the endonuclease cleaves each strand at a specific position, two bases apart, to produce a 3′ overhang (Fig. 2*C*). The cut also results in hydroxyl groups at the 3′ protruding ends and phosphoryl groups at the 5′ recessed ends, as deduced from the observed ligation of the cut ends by T4 DNA ligase (data not shown).

Two additional lines of experimental evidence confirm the location of the endonucleolytic cut in the *td*ΔI gene. (*i*) By using pUC*td*ΔI DNA precut by the endonuclease as template and primed with the same exon-1 and exon-2 probes, runoff synthesis resulted, in each case, in a cDNA fragment that was 2 nucleotides shorter than that in the previous experiment: 72 nucleotides long instead of 74 when primed with probe 2 on the coding strand (Fig. 3*A*); and 80 nucleotides instead of 82 when primed with probe 7 on the noncoding strand (Fig. 3*B*). This result is to be expected when a 2-base 3′ overhang exists at the cleavage site. (*ii*) A 21-mer oligodeoxynucleotide (Table 1 probe 4) that spans the nucleotide sequence centered at the observed endonucleolytic cut in exon 1 failed to hybridize to pUC*td*ΔI DNA pretreated with the endonuclease (Table 2). In this instance, the truncation in the *td*ΔI sequence greatly destabilized its interaction with probe 4, leading to loss of the probe during stringent washing. The data in Table 2 also show that the double-stranded pUC*td* DNA, which contains the 1016-bp intron juxtaposed between exon 1 and exon 2, and the single-stranded M13*td*ΔI DNA, containing the noncoding *td*ΔI sequence, are not substrates for the endonuclease. These results reemphasize both the sequence and double-strand specificities of the enzyme.

Determination of the Recognition Sequence of the *td* Intron Endonuclease. Although the pUC*td* DNA contains the endonucleolytic cleavage sequence, this DNA is not a substrate for the endonuclease (Table 2). It follows, therefore, that the presence of exon 2 sequence immediately after exon 1 is required for recognition by the endonuclease. To determine

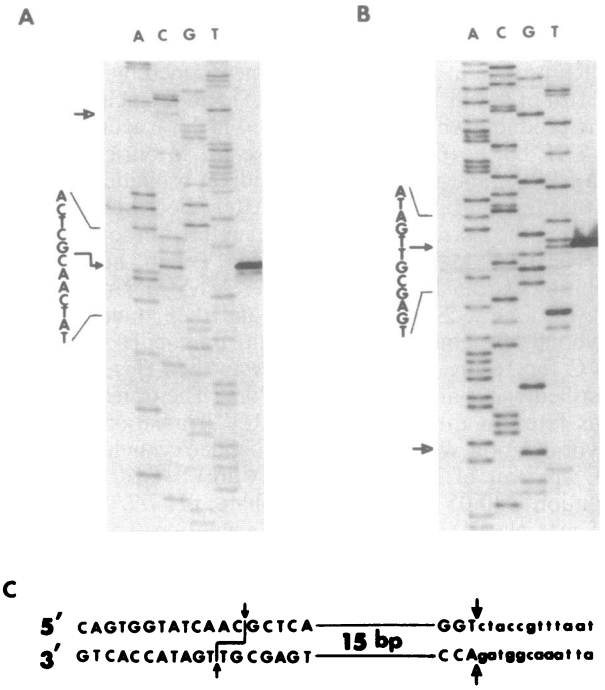


FIG. 2. Structural analysis of the cleavage site. Double-stranded DNA from plasmid pUC*td*ΔI was used as template for cDNA synthesis, according to the Sequenase protocol (United States Biochemical). Oligodeoxynucleotides labeled in their 5′ ends with ^{32}P by T4 polynucleotide kinase were used as primers. Probe 2 (containing exon 1 sequence) was used to copy the coding strand in the exon 1-to-exon 2 direction (*A*), and probe 7 (containing exon 2 sequence) was used to copy the noncoding strand in the opposite direction (*B*). Besides the reactions containing the individual dideoxynucleotide chain terminators (A, C, G, and T lanes serving as a sequence ladder), duplicate dNTP reactions without terminators (unmarked lanes) were prepared at the same time. One dNTP reaction (right lanes) was treated with the *td* intron-encoded endonuclease before electrophoresis in 8% polyacrylamide/6 M urea sequencing gel. (*C*) Summary of the sequencing data shown in *A* and *B*. Exon 1 sequence is in uppercase letters, and exon 2 sequence is in lowercase letters. The site of cleavage on each strand (▼) and the exon 1–exon 2 junction (▲) are indicated.

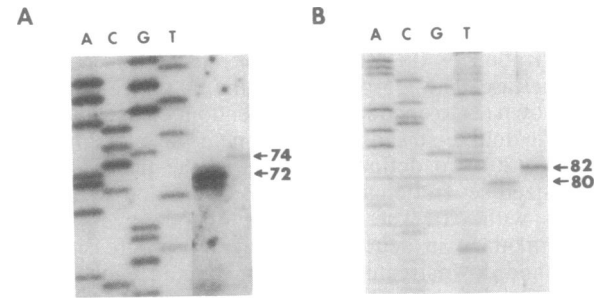


FIG. 3. Sequenase-directed runoff products of endonuclease-cleaved pUC*td*ΔI template DNA. Plasmid pUC*td*ΔI was first treated with the *td* intron-encoded endonuclease, deproteinized with phenol, and then used as template in cDNA synthesis in the absence of dideoxynucleotide chain terminators: probe 2-primed cDNA products (*A*) and probe 7-primed cDNA products (*B*). Lane 5 (from left), runoff cDNA from endonuclease-cleaved template; lane 6, endonuclease-cleaved cDNA synthesized from uncleaved template. The respective sequence ladders are marked A, C, G, and T. Major cDNA products are indicated by arrows, and their sizes (in nucleotides) were computed from the respective sequence ladders.

Table 2. Hybridization analysis of *td* and *td*ΔI DNA before and after treatment with *td* intron endonuclease

DNA	Oligodeoxynucleotide probe* 2	4	5
pUC*td*	+†	+	+
pUC*td* + endonuclease	+	+	+
pUC*td*ΔI	+	+	+
pUC*td*ΔI + endonuclease‡	+	−	+
M13*td*ΔI + endonuclease§	−	+	+

*See Table 1 for description of oligodeoxynucleotide probes.
†+, Hybridizing; −, nonhybridizing.
‡Endonucleolytic cut converts the circular pUC*td*ΔI DNA to a linear form.
§Single-stranded M13 recombinant phage DNA containing the *td*ΔI sequence.

the recognition sequence lying upstream and downstream of the cleavage site, an extension of the experiments described in Fig. 2 was executed. The cDNA products in the dideoxynucleotide-containing lanes, which served as a sequence ladder in Fig. 2, actually contain a nested set of truncated molecules differing in length by one nucleotide. Thus, the double-stranded hybrid molecules, each made up of a ^{32}P-labeled cDNA of different length base-paired to its unlabeled template strand, could be used as substrates for the endonuclease. By following the number of bases that must be converted to a double-stranded form before the endonuclease will cut, an upper limit for the recognition sequence can be determined. The boundary of recognition can be identified by the marked reduction or even disappearance of cDNA bands beyond a specific point when the endonuclease-treated cDNA products are resolved on a sequencing gel. The result of this experiment is depicted in Fig. 4, which shows the sequence profile both downstream (*A*) and upstream (*B*) of the endonucleolytic cut. Overexposure of the x-ray film to the radioactive sequencing gel was necessary to reveal the cDNA bands proximal to the boundaries of the recognition sequence. Based on the respective sequence profiles, we conclude that the maximum endonuclease recognition sequence is 87 bp long, running from 52 bp upstream of the cleavage site in exon 1 to 35 bp downstream of it (*C*), which includes the first 11 bp of exon 2.

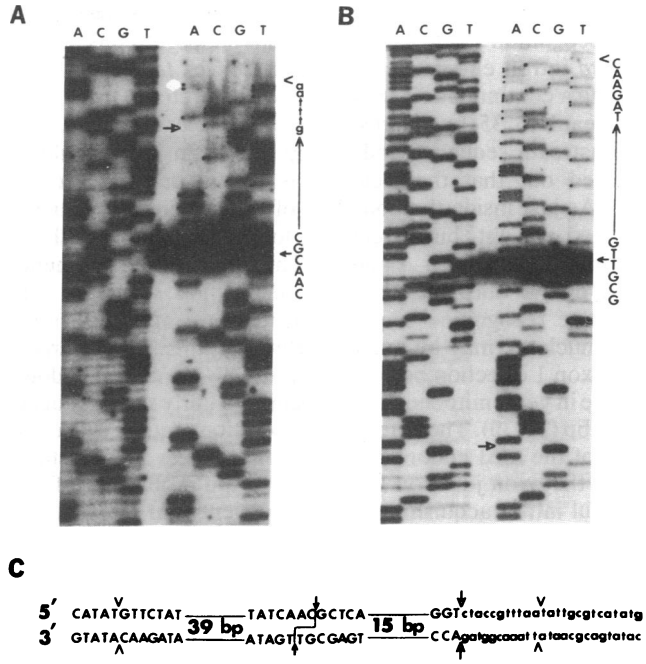


FIG. 4. Delineation of the recognition sequence for endonucleolytic cleavage. An extension of the experiment described in Fig. 2 was done. The cDNA products in the sequence ladder reactions containing dideoxynucleotide chain terminators in *A* and *B* (left four lanes) were used as substrates for the *td* intron-encoded endonuclease (right four lanes). (*A*) Probe 2-primed cDNA products. (*B*) Probe 7-primed cDNA products. (*C*) Maximum recognition sequence on each side of endonucleolytic cleavage site. Exon 1 sequence is in uppercase letters, and exon 2 sequence is in lowercase letters. Also indicated are the cleavage site on each strand (▼), the exon junction or potential intron-insertion point (▼), and the boundaries of the recognition sequence (v), beyond which the cDNA products are cleaved efficiently by the endonuclease—resulting in total or almost total disappearance of bands.

DISCUSSION

Group I introns, characterized by their ability to undergo self-splicing *in vitro* (for review, see ref. 29), are found in such diverse genetic sources as the *Tetrahymena* nucleus, yeast and fungal mitochondria, the *Physarum* nucleus, some chloroplasts, and the T-even bacteriophages. Although some group I introns, such as that in the *Tetrahymena* nuclear large rDNA, are relatively small and devoid of open reading frames, others contain extensive open reading frames that have become the subject of intense study. These group I iORFs can be generally divided into two categories: (*i*) those encoding maturases that facilitate splicing of the intron in which they are encoded at the RNA level and (*ii*) those that encode endonucleases that may be involved in unidirectional gene transfer at the DNA level, resulting in intron acquisition by homologous intronless genes. Some examples in the latter category include the 21S rDNA (21, 22) and *cox1* genes (30, 31) in yeast mitochondria, *Physarum* nuclear extrachromosomal rDNA (32), and T4 bacteriophage *td* (18, 20) and *sunY* (19) genes. Analysis of the substrate specificity of two yeast mitochondrial intron-encoded endonucleases *in vitro* has revealed that they produce 4-base-long 3′-hydroxyl overhangs, either at the intron-insertion site for the 21S rRNA gene (22) or 1 bp downstream of that site for the *cox1 a14α* gene (30, 31).

To study the structure and function of the product encoded by the 735-bp *td* iORF, the iORF was cloned into a high-expression vector, and the resulting partially purified protein was shown to possess endonucleolytic activity specific for the *td*ΔI gene (18). The data of Quirk *et al.* (19) obtained from genetic analyses showed that deleting a 21-bp stretch from 11 bp upstream to 10 bp downstream of the exon junction abolished endonuclease-facilitated *td* intron mobility. Our present work, which made use of a highly purified preparation of the cloned *td* intron endonuclease, has allowed us to determine unequivocally from a direct analysis of the endonucleolytic products that the cleavage site is centered at 24 bp upstream of the exon junction (Fig. 2) and outside of the 21-bp region. The cleavage is such that it forms a 2-base staggered cut, with 3′-hydroxyl protruding ends and 5′-phosphoryl recessed ends (Fig. 3 and F.K.C., unpublished results). That the *td* intron endonuclease cuts a site fairly removed from the exon junction and generates a 2-base 3′-hydroxyl overhang contrasts with the yeast mitochondrial intron endonucleases, which cut at or very close to the exon junction and produce 4-base overhangs (22, 30, 31). Regardless of the source, all intron-encoded endonucleases characterized thus far belong to a subclass of type 2 endonucleases that cleave nonsymmetrical sequences (33) and exhibit a very high degree of specificity.

Although the intron-containing *td* gene possesses the endonuclease cleavage sequence, it nevertheless is not a substrate (Table 2), indicating that some exon 2 sequence may be

required by the endonuclease for recognition. Delineation of the recognition sequence (Fig. 4) lends support to this notion. The *td* intron endonuclease readily cleaves *td*ΔI DNA containing a continuous 87-bp region, which stretches from 52 bp upstream to 35 bp downstream of the cleavage site. The downstream sequence ends 11 bp into exon 2. It should be pointed out that the cleavage is not all-or-none because cDNAs just inside the boundary of the recognition sequence showed a slight, but detectable, loss when treated with the endonuclease (Fig. 4 *A* and *B*). Therefore, our experimental data define only the optimum recognition sequence. The absolute recognition sequence in the case of the *td* intron endonuclease may be much shorter than this, particularly in the exon 1 direction because the *cox1 a14α*-encoded endonuclease in yeast mitochondria recognizes a fairly short sequence of 18 bp (22, 30). The extensive recognition sequence on either side of the *td*ΔI endonucleolytic cleavage site some distance from the exon junction may improve the probability of successful intron acquisition by the recipient *td*ΔI DNA, a phenomenon shown by genetic experiments (19).

What function does the *td* intron-encoded endonuclease serve? Intron mobility in the T-even bacteriophages was demonstrated for the *td* and *sunY* introns by a genetic approach that showed this process to absolutely depend on the integrity of the corresponding iORF (19). We have shown that the *td* iORF encodes a protein endonuclease specific for the double-stranded *td*ΔI DNA (18), which was subsequently confirmed by others (20). Because the intron transfer appears always to be in the direction of acquisition, it probably occurs by a mechanism similar to that postulated for the site-specific transfer of the 21S rDNA intron in yeast mitochondria (22), where Colleaux *et al.* invoked the double-strand break–repair model of Szostak *et al.* (34). However, they also suggested that the endonuclease-generated 3′ overhang can invade the corresponding intron-containing gene and act directly as a primer to elongate a copy of the intron sequence, using the appropriate intron-containing strand as template (22). Such a duplicative recombination mechanism will result in the coconversion of flanking exon sequences.

If the *td* endonuclease facilitates intron mobility among T-even phages, what does intron spreading accomplish? Speculations abound for lack of definitive experimental data in this aspect. We prefer to view intron acquisition by the phage *td* gene and phage intron mobility, in general, as a means to expand and/or diversify the coding capacity of a given gene in an organism, such as bacteriophage (27), which contains a relatively limited amount of genetic material compared to the eukaryotes. Thus, precise intron acquisition by the *td*ΔI gene increases the number of potential reading frames within this gene from one to four (*td*, exon 1, exon 2, and iORF), thereby expanding its coding capacity. In this context, we had shown that the exon 1 peptide is expressed early in a temporal fashion from the intron-containing *td* gene both *in vivo* (35) and in an *in vitro*-coupled transcription–translation system (36). Finally, genetic evidence from studying T4 phage deletion mutants has implied that phage introns could provide target sequences for recombination during exon shuffling, resulting in new combinations of exons (37).

Note. Shortly after submission of this work, Perlman and Butow (38), citing a personal communication of M. Belfort, indicated that the T-even intron-encoded endonucleases cleave at a distance (up to 25 bp) from the intron-insertion site.

We thank Dr. Angelo Lebo and Ms. Santos Seitz for synthesizing the oligodeoxynucleotides. The careful preparation of the manuscript by Ms. Judy Valentino and Ms. Susanne Koban is greatly appreciated. This work was supported, in part, by grants from the National Cancer Institute (CA44355) and the National Science Foundation (DMB86-16273).

1. Kaine, B. P., Gupta, R. & Woese, C. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3309–3312.
2. Chu, F. K., Maley, G. F., Maley, F. & Belfort, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3049–3053.
3. Chu, F. K., Maley, G. F., West, D. K., Belfort, M. & Maley, F. (1986) *Cell* **45,** 157–166.
4. Chu, F. K., Maley, F., Martinez, J. & Maley, G. F. (1987) *J. Bacteriol.* **169,** 4368–4375.
5. Sjöberg, B.-M., Hahne, S., Mathews, C. Z., Mathews, C. K., Rand, K. N. & Gait, M. J. (1986) *EMBO J.* **5,** 2031–2036.
6. Gott, J. M., Shub, D. A. & Belfort, M. (1986) *Cell* **47,** 81–87.
7. Tomaschewski, J. & Ruger, W. (1987) *Nucleic Acids Res.* **15,** 3632–3633.
8. Shub, D. A., Xu, M.-Q., Gott, J. M., Zeeh, A. & Wilson, L. D. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 193–200.
9. Goodrich, H. A., Gott, J. M., Xu, M.-Q., Scarlato, V. & Shub, D. A. (1989) in *Molecular Biology of RNA*, UCLA Symposia on Molecular and Cellular Biology, New Series, ed. Cech, T. (Liss, New York), Vol. 94, pp. 59–66.
10. Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. (1982) *Nature (London)* **300,** 719–724.
11. Michel, F., Jacquier, A. & Dujon, B. (1982) *Biochimie* **64,** 867–881.
12. Been, M. D., Barfod, E. T., Burke, J. M., Price, J. V., Tanner, N. K., Zaug, A. J. & Cech, T. R. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 147–157.
13. Cech, T. R., Zaug, A. J. & Grabowski, P. J. (1981) *Cell* **27,** 487–496.
14. Chu, F. K., Maley, G. F. & Maley, F. (1987) *Biochemistry* **26,** 3050–3057.
15. Lazowska, J., Jacq, C. & Slonimski, P. P. (1980) *Cell* **22,** 333–348.
16. De La Salle, H., Jacq, C. & Slonimski, P. P. (1982) *Cell* **28,** 721–732.
17. Carignani, G., Groudinsky, O., Frezza, D., Schiavon, E., Bergantino, E. & Slonimski, P. P. (1983) *Cell* **35,** 733–742.
18. West, D. K., Changchien, L.-M., Maley, G. F. & Maley, F. (1989) *J. Biol. Chem.* **264,** 10343–10346.
19. Quirk, S. M., Bell-Pedersen, D. & Belfort, M. (1989) *Cell* **56,** 455–465.
20. Bell-Pedersen, D., Quirk, S. M., Aubrey, M. & Belfort, M. (1989) *Gene* **82,** 119–126.
21. Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. & Dujon, B. (1986) *Cell* **44,** 512–533.
22. Colleaux, L., d'Auriol, L., Galibert, F. & Dujon, B. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6022–6026.
23. Studier, F. W. & Moffat, B. A. (1986) *J. Mol. Biol.* **189,** 113–130.
24. Rosenberg, A. H., Lade, B. N., Chui, D.-S., Lin, S.-W., Dunn, J. J. & Studier, F. W. (1987) *Gene* **56,** 125–135.
25. West, D. K., Belfort, M., Maley, G. F. & Maley, F. (1986) *J. Biol. Chem.* **261,** 13446–13450.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
27. Chu, F. K., Maley, G. F. & Maley, F. (1988) *FASEB J.* **2,** 216–223.
28. Rosenberg, M., Ho., Y.-S. & Shatzman, A. (1983) *Methods Enzymol.* **101,** 123–138.
29. Cech, T. R. & Bass, B. L. (1986) *Annu. Rev. Biochem.* **55,** 599–629.
30. Wenzlau, J. M., Saldanha, R. J., Butow, R. A. & Perlman, P. S. (1989) *Cell* **56,** 421–430.
31. Delahodde, A., Goguel, V., Becam, A. M., Creusot, F., Perea, J., Banroques, J. & Jacq, C. (1989) *Cell* **56,** 431–441.
32. Muscarella, D. E. & Vogt, V. M. (1989) *Cell* **56,** 443–454.
33. Modrich, P. & Roberts, R. J. (1982) in *Nucleases*, eds. Linn, S. M. & Roberts, R. J. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 109–154.
34. Szostak, J. W., Orr-Weaver, T. L. & Rothstein, R. J. (1983) *Cell* **33,** 25–35.
35. Belfort, M., Pedersen-Lane, J., Ehrenman, K., Chu, F. K., Maley, G. F., Maley, F., McPheeters, D. S. & Gold, L. (1986) *Gene* **41,** 93–102.
36. Chu, F. K., Maley, G. F., Belfort, M. & Maley, F. (1985) *J. Biol. Chem.* **260,** 10680–10688.
37. Hall, D. H., Liu, Y. & Shub, D. A. (1989) *Nature (London)* **340,** 574–576.
38. Perlman, P. S. & Butow, R. A. (1989) *Science* **246,** 1106–1109.