Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 5577–5581, June 1992
Biochemistry

Intervening sequences in an Archaea DNA polymerase gene

(intron/self-splicing/protein splicing)

FRANCINE B. PERLER*†, DONALD G. COMB*, WILLIAM E. JACK*, LAURIE S. MORAN*, BOQIN QIANG*‡, REBECCA B. KUCERA*, JACK BENNER*, BARTON E. SLATKO*, DONALD O. NWANKWO*, S. KAY HEMPSTEAD*, CLOTILDE K. S. CARLOW*, AND HOLGER JANNASCH§

*New England Biolabs, Inc., 32 Tozer Road, Beverly, MA 01915; and §Woods Hole Oceanographic Institute, Woods Hole, MA 02543

Communicated by Charles C. Richardson, March 18, 1992 (received for review January 29, 1992)

ABSTRACT **The DNA polymerase gene from the Archaea *Thermococcus litoralis* has been cloned and expressed in *Escherichia coli*. It is split by two intervening sequences (IVSs) that form one continuous open reading frame with the three polymerase exons. To our knowledge, neither IVS is similar to previously described introns. However, the deduced amino acid sequences of both IVSs are similar to open reading frames present in mobile group I introns. The second IVS (IVS2) encodes an endonuclease, I-*Tli* I, that cleaves at the exon 2–exon 3 junction after IVS2 has been deleted. IVS2 self-splices in *E. coli* to yield active polymerase, but processing is abolished if the IVS2 reading frame is disrupted. Silent changes in the DNA sequence at the exon 2–IVS2 junction that maintain the original protein sequence do not inhibit splicing. These data suggest that protein rather than mRNA splicing may be responsible for production of the mature polymerase.**

In 1985, a species of extreme thermophile was isolated from a submarine thermal vent near Naples, Italy (1). This organism, *Thermococcus litoralis*, can be cultured at up to 98°C and contains a heat-stable DNA polymerase that we call Vent DNA polymerase (New England Biolabs). This paper describes the cloning, sequencing, and expression of the Vent DNA polymerase gene¶ and the finding of two intervening sequences (IVSs) that make up 55% of the polymerase gene, one of which, IVS2, encodes the I-*Tli* I (I, intron) endonuclease.

To our knowledge, this is the first report of introns in protein coding genes of Archaea or eubacteria, although introns have been found in protein coding genes of eubacteriophage (2). Previously described Archaea or eubacterial introns are mainly pre-tRNA or self-splicing introns in stable RNAs (2–5).

Introns often contain open reading frames (ORFs) that are in-frame with either the 5′ or 3′ exon, but not with both exons. An intron in the *Saccharomyces cerevisiae TFP1* gene forms a single ORF with the surrounding exons; the authors proposed (6, 7) that this intron is spliced at the protein, not the mRNA, level. In the present study, we describe two introns that form a single ORF with the surrounding exons. Furthermore, we present evidence indicating that the Vent DNA polymerase IVSs are removed either by protein splicing or by RNA splicing that requires I-*Tli* I as a maturase.

MATERIALS AND METHODS

Western Blots. Anti-Vent DNA polymerase sera was raised by immunizing mice with purified native *T. litoralis* DNA polymerase (New England Biolabs). Western blot analysis employed protein samples from isopropyl β-D-thiogalactoside-induced cultures of BL21(DE3)plysS (8) containing the indicated expression constructs. Samples were analyzed by SDS/PAGE in 4–20% gels (ISS, Daiichi, Tokyo, Japan) with prestained markers (BRL), transferred to nitrocellulose (9), probed with anti-Vent DNA polymerase sera (10), and detected using alkaline phosphatase-linked anti-mouse secondary antibody as described by the manufacturer (Promega).

Polymerase Assay. Polymerase activity was measured as cpm incorporated into acid-insoluble DNA. Lysates were prepared by resuspending *Escherichia coli* cell pellets in 20 mM Tris·HCl, pH 8/50 mM NaCl/0.1% Triton X-100/bovine serum albumin (0.2 mg/ml), heating to 80°C for 20 min, and pelleting cell debris for 10 min at 12,000 × *g* at room temperature. Assays were incubated at 72°C in 10 mM KCl/10 mM $(NH_4)_2SO_4$/20 mM Tris, pH 8.8/2 mM $MgSO_4$/0.1% Triton X-100/[^{32}P]dCTP (10–20 × 10^6 cpm/ml)/all four dNTPs (each at 33 μM)/activated DNA (0.2 mg/ml) (26).

Library Construction and Screening. *T. litoralis* DNA partially digested with *Eco*RI was cloned into λgt11 and λZapII (Stratagene) and libraries were screened with anti-Vent DNA polymerase sera (9, 10). *Bam*HI-digested *T. litoralis* DNA was cloned into λDash (Stratagene) and screened by hybridization to the 1.3-kilobase (kb) *Eco*RI insert from λV10-49 (ref. 9).

DNA Sequencing. Both strands of the polymerase coding region (Fig. 1) were sequenced from two independent sources: the three *Eco*RI fragments and subclones of the 12-kb *Bam*HI fragment from λV56-9 (11).

Construction of a Recombinant Expressing Vent DNA Polymerase. Vent DNA polymerase was expressed in *E. coli* by using the T7 expression system of Studier *et al.* (8). The gene lacking IVS1 was created in a multistep ligation involving sequences from (*i*) an *Nde* I site created (12) at the initiation codon [base pair (bp) 291] to the *Pvu* I site (bp 1720), (*ii*) a bridging synthetic double-stranded oligonucleotide with *Pvu* I and *Bsu*36I termini (bp 1721–1772 joined to bp 3387–3412) (5′-CGAAAAGAAAATGCTCGATTATAGGCAAAGGGCTATTAAATTGCTAGCAAACAGCTATTACGGCTATATGGGGTACCC-3′ and 3′-TAGCTTTTCTTTTACGAGCTAATATCCGTTTCCCGATAATTTAACGATCGTTTGTCGATAATGCCGATATACCCCATGGGATT-5′), and (*iii*) a *Bsu*36I (bp 3412) to *Bam*HI (bp 5832) fragment, where numbers in parentheses indicate coordinates. These gene segments were inserted into *Nde* I/*Bam*HI-cut pAII17, a pET11c (8) derivative that had been modified to reduce basal expression from the T7 RNA polymerase promoter by the upstream addition of four copies

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: ORF, open reading frame; IVS, intervening sequence; Polα, polymerase α.

†To whom reprint requests should be addressed.
‡Present address: Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, 5 Dong Dan San Tiao, Beijing, China.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M74198).

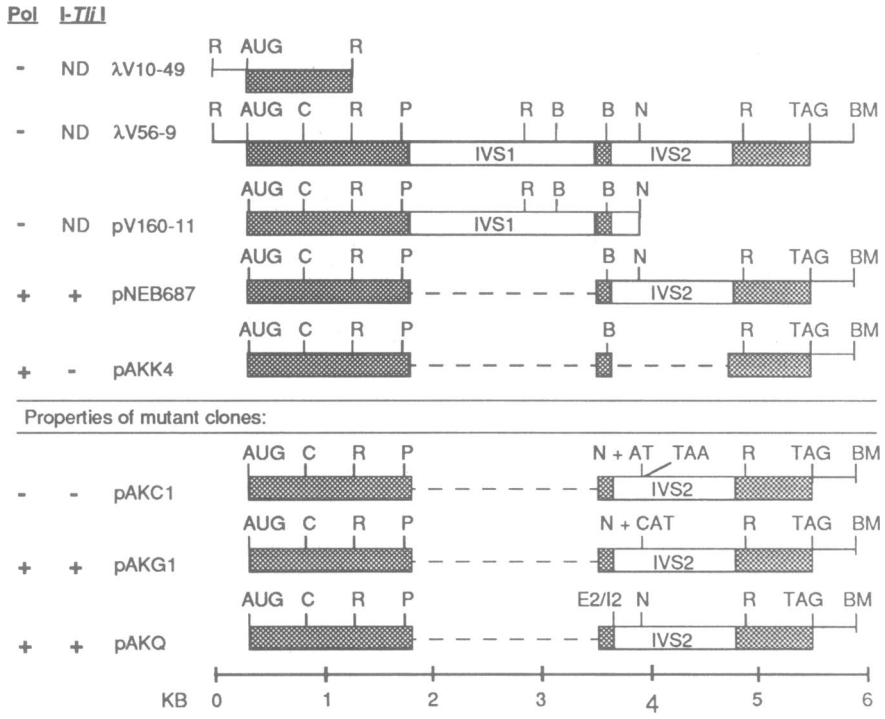


FIG. 1. Maps of Vent DNA polymerase clones. The genomic polymerase gene contains two IVSs (open boxes) and three exons (hatched boxes); dashes represent deleted sequences. Only part of λV56-9 (12 kb) is presented. The polymerase start codon (AUG, bp 291) and the termination codon (TAG, bp 5397) are indicated. IVS2 encodes I-*Tli* I endonuclease. Production in *E. coli* of active Vent DNA polymerase (Pol) or I-*Tli* I is indicated (ND, not determined). R, *Eco*RI; C, *Cla* I; P, *Pvu* I; B, *Bsu*36I; N, *Nde* I; Bm, *Bam*HI. Mutations are as follows (see Tables 1 and 2): N+AT/TAA, AT insertion at the *Nde* I site resulting in a termination codon (TAA); N+CAT, CAT insertion at the *Nde* I site; E2/I2, location of exon 2–IVS2 junction silent mutations. pV160-11 and pAKC1 fail to splice in *E. coli*, whereas pNEB687, pAKG1, and the pAKQ series do splice.

of the rrn_b transcription terminator from pRS415 (13). The resulting construct was named pNEB687 (Fig. 1).

Deletion of IVS2. The polymerase chain reaction (PCR) was used to generate two fragments with termini near the planned deletion. The PCR mixture contained Vent DNA polymerase buffer (New England Biolabs)/bovine serum albumin (0.1 mg/ml)/all four dNTPs (each at 0.2 mM)/pNEB687 (0.9 μg/ml)/primers (0.5 μg/ml)/Vent DNA polymerase (10 units/ml). Amplification was carried out for 15 cycles of 94°C for 0.5 min, 50°C for 0.5 min, and 72°C for 2 min by using a Perkin–Elmer/Cetus thermal cycler. Both PCR fragments incorporated silent changes at bp 3518 (G → A) and 3521 (T → A) to create a *Sca* I site that joined the fragments. The first fragment extended from the *Eco*RI site (bp 1269) to the newly created *Sca* I site (bp 3518). The second fragment included the *Sca* I site to bp 3533, then skipped to bp 4704, and continued to the *Hin*dIII site at bp 5058. The two PCR fragments were ligated at the *Sca* I site and joined to pNEB687 at the *Eco*RI and *Hin*dIII sites. The resulting construct was named pAKK4 (see Fig. 1). The final amino acid sequences across the junction for both pNEB687 and pAKK4 are given in Fig. 2.

Reading Frame Alteration in IVS2. A 2-base insertion within IVS2 was made by cutting pNEB687 at the *Nde* I site (bp 3906), filling-in with the Klenow fragment of DNA polymerase I, and circularizing to create pAKC1 (Table 1). The reading frame was restored by adding a third base pair at the altered *Nde* I site by using a PCR to create pAKG1 (Table 1).

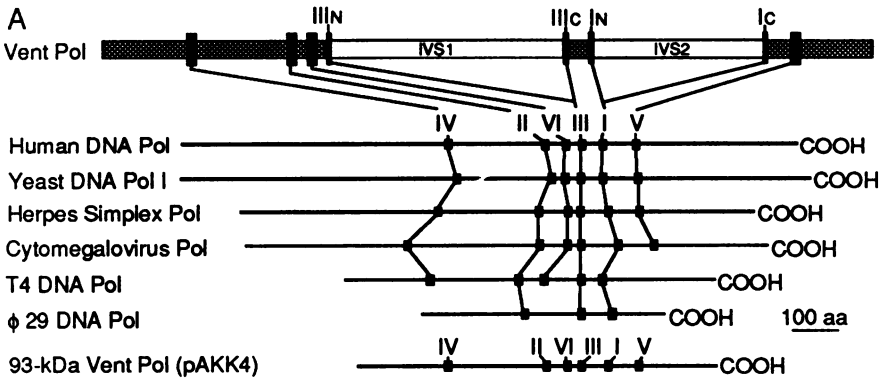
Silent Base Changes Within the Exon 2–IVS2 Boundary. Variants in the exon 2–IVS2 boundary were formed in a three-part ligation. A PCR was performed as above. Fragment 1 was generated with primers 5′-CGGCGCATATGATACTGGACACTGATTAC-3′ (bp 311) plus 5′-pATC(TGA)GCGTA(TGC)AGTACTTTAAAGCCGAACTTTTCC-3′ (bp 3500, reverse); and fragment 2 was generated with 5′-pTC(ACG)GT(GAT)AG(TC)GGAGAAAGTGAGATCATAATAAGG-3′ (bp 3566) plus 5′-GAGACTCGCGGAGAAACTTGGACT-3′ (bp 5414, reverse), where bases in parentheses indicate degenerate mixtures, coordinates of primer 3′ ends are listed, and the p indicates 5′-PO_4. Fragment 1 was digested with *Cla* I (bp 817), and fragment 2 was digested with *Nsi* I (bp 5384). They were then ligated to pNEB687 digested with *Cla* I and *Nsi* I. The exon 2–IVS2 junction was formed by direct ligation of the PCR product blunt ends. The resulting constructs, the pAKQ series, were sequenced (11) to verify alterations (Table 2).

RESULTS

Cloning Vent DNA Polymerase. Vent DNA polymerase is extremely thermostable. It was purified by conventional chromatography to yield a final preparation containing four proteins detectable by Coomassie Blue staining. We identified the 93-kDa (±5%) species as the polymerase by assaying proteins renatured after elution from a SDS/PAGE gel (ref. 15; H. M. Kong, R.B.K., and W.E.J., unpublished data). The eluted polymerase also possesses a 3′ → 5′ exonuclease activity that enhances the fidelity of polymerization (16).

Our cloning strategy was to raise antibodies against Vent DNA polymerase to screen λgt11 and λZapII expression libraries. Fifty-four recombinants were isolated with anti-Vent DNA polymerase sera, but no clone synthesized active thermostable polymerase. One λgt11 recombinant, λV10-49 (Fig. 1), was chosen for further analysis because it produced the largest nonfusion protein (≈38 kDa). To determine whether the 1.3-kb *Eco*RI fragment from λV10-49 encoded the polymerase N terminus, both the cloned DNA fragment and the native protein were sequenced. The amino acid sequence of native Vent DNA polymerase indicated the N terminus was Met-Ile-Leu (17, 18). The deduced amino acid sequence beginning at bp 291 of the 1.3-kb *Eco*RI fragment matched this protein sequence. The initiating methionine was preceded by consensus archaeal promoter elements (19).

Because the insert in λV10-49 was too small to code for the entire gene, larger *T. litoralis* fragments containing the 1.3-kb *Eco*RI fragment were identified by Southern blot analysis (unpublished data). A 12-kb *Bam*HI fragment was identified and cloned to yield phage λV56-9 (Fig. 1). No thermostable DNA polymerase activity could be detected in extracts of λV56-9-infected cells. Therefore, the polymerase gene structure was analyzed by DNA sequencing of a 5837-bp region. The DNA polymerase ORF begins at nucleotide 291 and remains open for 1702 amino acids, coding for a protein



B

Region IV		**V..F..E...L..F.....P....G.NI..FD....L.R**
Human	614	**VEVAATERTLLGFFLAKVHKIDPDIIVGHNIYGFELEVLLQR**
Vent	181	**VDVVSNERE**MIKR**F**VQVV**K**EK**DPDVI**ITYN**GDN**FDLPYLIKR**
Region II		**Y.GG.V..P..G.....V.V.DFNSLYPSII...N.C..T**
Human	844	**YAGGLVLDPKVGFYDKFILLLDFNSLYPSII**QEF**N**ICF**T**T
Vent	387	**YLGGYVKEPEKGLWEN-IIYLDFRSLYPSII**VTH**N**VSPD**T**
Region VI		**L..LL.....R...KK.....P**
Human	916	LPREIRKLVERRKQVKQLMKQQDLNP
Vent	452	IPSILGDLIAMRQDIKKKMK-STIDP
Region III		**D..Q.A.K...NS.YG..G.....A...T..GR..L**
Human	948	**D**IRQKALKLTANSMYGCLGFSYSRFYAKPLAALVTYKGREIL
VentIII	N+C	**D**YRQRAIKLLANSYYGYMGYPKARWYSKECAESVTAWGRHYI
VentIIIN	483	**D**YRQRAIKLLANSilpnewlpiiengeikfvkigefinsyme
VentIIIC	1021	flvgfglllyah**N**SYYGYMGYPKARWYSKECAESVTAWGRHYI
Region I		**...YGDTDS.F.....**
Human	1002	EVIYGDTDSIMINTNSTNLEEVFK
VentI	N+C	KVLYADTDGFYATIPGEKPELIKK
VentIN	1076	KVLYADsvsgeseiiirqngkirf
VentIC	1466	nilvhnTDGFYATIPGEKPELIKK
Region V		**K...KG.D.VR...C**
Human	1076	KQELKGLDIVRRDWC
Vent	1533	RITTRGLEVVRRDWS

FIG. 2. Conserved polymerase (Pol) motifs allow prediction of IVS junctions. (*A*) Schematic arrangement of the six Polα conserved regions (I–VI) (14). Open boxes (IVSs) represent segments that lack similarity to other DNA polymerases. Vent region III is split into regions III_N and III_C by IVS1 and Vent region I is split into regions I_N and I_C by IVS2. (*B*) Selected amino acid sequences. The consensus region, human Polα (ref. 14, Human), and Vent DNA polymerase (Vent) sequences are presented. Periods, nonconserved residues; boldface type, identities to Vent DNA polymerase; dashes, alignment gaps. $VentIII_N$, $VentIII_C$, $VentI_N$, and $VentI_C$ are the genomic sequences before IVS deletion; residues in lowercase letters were deleted. $VentIII_{N+C}$ and $VentI_{N+C}$ are the sequences in pNEB687 and pAKK4 after deletion of IVS1 and IVS2, respectively. Numbers indicate amino acid position.

almost twice as large as the apparent molecular mass of the polymerase.

Comparison with Other DNA Polymerases Predicts Two IVSs. Since the genomic ORF was larger than predicted, the deduced amino acid sequence was compared to other DNA polymerases (Fig. 2*A*). DNA polymerases have a high degree of identity in small interspersed regions that enables classification of DNA polymerases into two families: (*i*) the polymerase I-like class A, whose prototype is *E. coli* DNA polymerase I and includes *Taq* polymerase, and (*ii*) the polymerase α (Polα)-like class B polymerases that are similar to human DNA polymerase α (14, 20). These regions of identity are thought to compose substrate and metal ion binding domains and the active site (14, 20). Polymerases of the same family also have various degrees of similarity throughout the entire sequence.

Vent DNA polymerase is sensitive to aphidicolin, indicating that it is a member of the Polα family. The 3′ → 5′ exonuclease motifs (21) were identified in the following regions: EXOI at Leu-137 (LLAFDIET), EXOII at Val-205 (VIITYNGDNFDLPYLI), and potential EXOIII at Leu-310 (LAQYSMEDARA). Although we were able to identify Polα regions IV, II, VI, and V (14, 20), identification of regions III and I proved more elusive. When these regions were eventually discerned, the motifs were found to be interrupted by sequences that bore no similarity to known polymerases.

Similarity to region III was lost after Asn-494 but resumed at Ser-1033 (Fig. 2*B*). Similarity to human Polα was interrupted a second time after Asp-1081 but resumed at Thr-1472 (Fig. 2*B*, region I). On the basis of the above comparisons and the difference in observed molecular weight vs. deduced

Table 1. Frame-shift mutations at the *Nde* I site

Plasmid	DNA and protein sequences									Act
pNEB687	CTC	ATA	TGC	CCA	AAT	GCA	CCG	TTA	AAG	+
	L	I	C	P	N	A	P	L	K	
pAKC1	CTC	ATa	tAT	GCC	CAA	ATG	CAC	CGT	TAA	−
	L	I	y	a	q	m	h	r	*	
pAKG1	CTC	ATc	atA	TGC	CCA	AAT	GCA	CCG	TTA	+
	L	I	i	C	P	N	A	P	L	

Act, Vent DNA polymerase and I-*Tli* I endonuclease activities. Nucleotide and amino acid changes are in lowercase type. Sequences start at bp 3903. The *Nde* I site is underlined. *, Termination codon; +, activities present; −, no activity.

Table 2. Silent mutations at the exon 2–IVS2 junction

Plasmid	DNA and/or protein sequence									
pNEB687	K	V	L	Y	A	D	S	V	S	G
	AAG	GTT	CTT	TAT	GCG	GAC |	AGT	GTC	TCA	GGA
pAKQ	K	V	L	Y	A	D	S	V	S	G
pAKQ25	AAa	GTa	CTc	TAc	GCt	GAt |	tcg	GTt	agc	GGA
pAKQ45	AAa	GTa	CTa	TAc	GCa	GAt |	tcg	GTg	agt	GGA
pAKQ61	AAa	GTa	CTg	TAc	GCc	GAt |	tcg	GTt	agc	GGA
pAKQ64	AAa	GTa	CTc	TAc	GCt	GAt |	tcg	GTa	agc	GGA
pAKQ68	AAa	GTa	CTa	TAc	GCt	GAt |	tcg	GTg	agt	GGA
pAKQ69	AAa	GTa	CTg	TAc	GCa	GAt |	tca	GTg	agt	GGA
pAKQ71	AAa	GTa	CTg	TAc	GCc	GAt |	tcc	GTg	agc	GGA
pAKQ72	AAa	GTa	CTc	TAc	GCt	GAt |	tcg	GTg	agt	GGA
pAKQ74	AAa	GTa	CTa	TAc	GCc	GAt |	tcg	GTt	agc	GGA

All pAKQ plasmids make active Vent DNA polymerase and I-*Tli* I. Nucleotide changes (lowercase type) do not alter amino acids. The vertical bar denotes the exon 2–IVS2 junction at bp 3533–3534.

amino acid sequence, we predicted that the Vent DNA polymerase gene contains IVSs in the two nonconserved sections.

***In Vitro* Deletion of IVS1.** When the entire Vent DNA polymerase gene was cloned in plasmids, recombinants grew poorly, plasmids rearranged, and no active polymerase could be detected. Expression of active polymerase in *E. coli* was achieved by placing the gene under a tightly controlled promoter and removing IVS1. We first reduced basal expression in pET11c (8) by inserting transcription terminators upstream of the T7 RNA polymerase promoter. Next, site-directed mutagenesis (12) was employed to create an *Nde* I site overlapping the Vent DNA polymerase initiation codon to allow linkage of the N terminus of the polymerase gene (bp 291–1274) to the T7 promoter.

When the N terminus clone was extended to the *Nde* I site, bp 3906 (pV160-11, Fig. 1), only an ≈135-kDa protein was detected in Western blots (Fig. 3). Since the 135-kDa protein represents the entire coding capacity of the 3.6-kb insert, in the absence of the remainder of the polymerase gene, IVS1 was not spliced in *E. coli*.

IVS1 splits region III and thus was deleted by overlapping $VentIII_N$ and $VentIII_C$ at the duplicated asparagine and serine residues (Fig. 2*B*, $VentIII_{N+C}$) to create a consensus region III in pNEB687. pNEB687 (Fig. 1) produces active thermostable polymerase with fidelity and kinetic parameters indistinguishable from native Vent DNA polymerase (ref. 16 and H. M. Kong *et al.*, unpublished data). It should be noted that because some amino acids in region III are not conserved, alternate combinations between $VentIII_N$ and $VentIII_C$ would also fit the region III consensus.

Identification of IVS2. Even after deletion of IVS1, the remaining coding capacity of the gene is 132 kDa, whereas the observed molecular masses of both the native and recombinant enzymes are 93 kDa (Fig. 3). We suspected that there was a second self-splicing IVS interrupting region I because region I is absent and amino acids 1082–1471 lack similarity to DNA polymerases. This sequence was deleted by joining Asp-1081 to Thr-1472 (Fig. 2*B*). The resultant clone pAKK4 (Fig. 1) produced active Vent DNA polymerase.

Identification of I-*Tli* I. A thermostable endonuclease was detected during purification of recombinant Vent DNA polymerase. The 42-kDa endonuclease I-*Tli* I (F.B.P., D.G.C., and J.B., unpublished data) does not react with anti-Vent DNA polymerase sera and is absent in purified polymerase preparations (Fig. 3). The I-*Tli* I N terminus was sequenced (17, 18) to yield 30 amino acids that are identical to the polymerase ORF beginning at Ser-1082, the first residue of IVS2. The size of the endonuclease approximates the coding capacity of IVS2. The mature endonuclease must be a processed protein since *E. coli* initiates translation with *N*-formylmethionine. I-*Tli* I is similar (29% identity) to the HO endonuclease from *S. cerevisiae* (22, 23) and to group I intron ORFs, especially at the conserved dodecapeptide motifs beginning at Trp-1227 and Ala-1325 (2, 4). Furthermore, like intron endonucleases, deletion of IVS2 creates an I-*Tli* I site at the exon 2–exon 3 junction (CTTTATGCGGA-CAC'TGACGGCTTT; the exact size of the recognition site has not been determined).

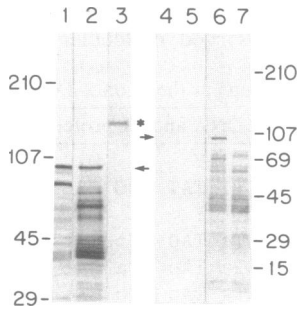


FIG. 3. Western blot analysis. Native and recombinant Vent DNA polymerases comigrate in SDS/PAGE gels. Splicing is inhibited in the absence of the C-terminal half of the gene (pV160-11) or in the presence of a frame-shift mutation in IVS2 (pAKC1). *T. litoralis* cell lysates or cell lysates prepared from indicated induced *E. coli* recombinants were probed by Western blot analysis using anti-Vent DNA polymerase sera. Lanes: 1, *T. litoralis*; 2, pNEB687; 3, pV160-11; 4, pAII17; 5, purified I-*Tli* I; 6, pNEB687; 7, pAKC1. Arrow, Vent DNA polymerase; asterisk, 135-kDa pV160-11 product.

Requirement for Intact IVS2 ORF—Possible Splicing Mechanisms. One of the most unusual properties of these IVSs is that they form a single reading frame with the polymerase. The yeast *TFP1* gene is split by a 50-kDa spacer peptide that is spliced at the protein, not the mRNA, level (7). Is Vent DNA polymerase the second example of protein splicing? We have been unable to detect Vent DNA polymerase mRNA in *T. litoralis* extracts by Northern blot analysis using the N-terminal 1.3-kb *Eco*RI fragment as a probe (unpublished data), perhaps because of low mRNA abundance, the absence of a unique-sized transcript, or degradation. Preliminary analysis using a cDNA PCR with RNA from *T. litoralis* or *E. coli* has also been inconclusive. Accordingly, we created mRNA variants to probe the essential features of the splicing event.

To determine whether the IVS2 reading frame is required for splicing, we created two frame-shift mutations by first disrupting the ORF and then reestablishing it (Table 1 and Fig. 1). When a 2-bp frame-shift insertion was introduced in the *Nde* I site (pAKC1), DNA polymerase and I-*Tli* I activity were abolished, the 132- and 93-kDa polymerase products were no longer synthesized, and only premature termination products <78 kDa were observed (Fig. 3, lane 7).

Two causes for the loss of polymerase activity can be envisioned. Secondary structure required for RNA splicing may have been disrupted by the insertion, or the ORF is required for producing mature polymerase. We reasoned that insertion of a third nucleotide at the same position would restore the ORF but would be unlikely to restore RNA structure disrupted by the original mutation. The resulting recombinant pAKG1 (Table 1 and Fig. 1) produced both active polymerase and endonuclease, thus establishing the requirement for the IVS2 ORF in the maturation process.

Silent Mutations at the Exon 2–IVS2 Junction. Although the IVS2 ORF is required for processing, it is still possible that I-*Tli* I acts as a maturase for RNA splicing (refs. 2 and 5; see *Discussion*). If RNA splicing is involved, one would predict that RNA structure at the splice junctions would be particularly crucial to the splicing events. Accordingly, we introduced a number of silent mutations in the vicinity of the exon 2–IVS2 splice junction without altering the protein sequence (Table 2 and Fig. 1). Nine plasmids of the pAKQ family were sequenced; 13 positions were changed, including 6 positions where several changes were introduced. These mutations included 4 bp surrounding the predicted splice junction. Despite these alterations, the observation that levels of endonuclease and polymerase were comparable to the parent plasmid casts doubt on the involvement of RNA splicing in IVS2 processing.

DISCUSSION

The only copy of an essential gene, Vent DNA polymerase, contains two IVSs that interrupt conserved DNA polymerase motifs. Moreover, both IVSs form one continuous ORF with the three polymerase exons.

IVS2 encodes an endonuclease, I-*Tli* I, that has features in common with group I intron endonucleases—namely, the conserved dodecapeptide motifs and recognition of the homing site (2, 4). Group I intron endonucleases are thought to

mediate intron mobility by site-specific recombination initiated by endonuclease cleavage at the "homing site" in genes that lack the intron (2). The question of Vent IVS mobility has not yet been addressed. However, although IVS2 encodes a group I intron-like endonuclease, it lacks the secondary structures and/or sequences that define previously described group I, group II, and Archaea pre-tRNA introns (2–5). Neither IVS is similar to the intron in the phage SPO1 DNA polymerase gene (2).

We have been unable to gather any data that would support an RNA splicing mechanism in production of Vent DNA polymerase in *E. coli*. No spliced mRNA has been detected in cDNA PCR experiments (unpublished data), although this negative result could be due to technical limitations or to immediate degradation after or coincident with translation. We have, however, been able to demonstrate a requirement for in-frame expression within IVS2. Recently, some intron endonucleases have also been shown to act as maturases to facilitate mRNA splicing (2, 4, 5). If this is the case, then the pAKC1 frameshift mutation should be complemented by I-*Tli* I produced in trans.

The strongest evidence against RNA splicing is the lack of an effect of multiple silent mutations at the exon 2–IVS2 junction, particularly since 4 bases adjoining the splice junction were altered. The exon–intron junctions in previously described group I, group II, and Archaea pre-tRNA introns form secondary and tertiary structural elements required for splicing. The silent mutations are likely to disrupt these structural elements but had no effect on splicing. This result, coupled with the lack of similarity to known intron structures, indicates that if RNA splicing occurs, it must differ significantly from previously identified systems. It is also possible, but seems unlikely, that the silent mutations occur in noncritical regions or that I-*Tli* I, acting as a maturase, is insensitive to the silent mutations.

We have less data concerning the splicing mechanism of IVS1, although the assumption is that it is similar to IVS2. IVS1 fails to splice in *E. coli* in the absence of amino acids 1208–1702 (pV160-11) and may require the complete polymerase or a maturase to splice.

Protein splicing is becoming more widely recognized and has been observed in formation of concanavalin A and the yeast *TFP1* gene product (7, 24). A gene product corresponding to the IVS is produced by both *TFP1* and Vent DNA polymerase (I-*Tli* I). We note that the two Vent DNA polymerase IVSs and the *TFP1* IVS are bordered by residues (e.g., threonine, serine, and cysteine) that are active as nucleophiles in known proteases. These same amino acids can also mediate peptide cleavage when prohistidine decarboxylase is activated by peptide cleavage (25). Thus serine, cysteine, or threonine may be critical in the peptide cleavage and rejoining reactions that generate the spliced protein and free IVS.

We thank Drs. F. Michel and J. Szostak for help in attempting to find catalytic core structures, Dr. H. Paulus for suggestion of possible protein splicing mechanisms, and Drs. P. Riggs, C. Guan, C. Noren, M. Southworth, and I. Schildkraut for help and discussions.

1. Belkin, S. & Jannasch, H. W. (1985) *Arch. Microbiol.* **141,** 181–186.
2. Belfort, M. (1990) *Annu. Rev. Genet.* **24,** 363–385.
3. Daniels, C. J., Gupta, R. & Doolittle, W. F. (1985) *J. Biol. Chem.* **260,** 3132–3134.
4. Cech, T. R. (1990) *Annu. Rev. Biochem.* **59,** 543–568.
5. Delahodde, A., Goguel, B., Becam, A. M., Creusot, F., Perea, J., Banroques, J. & Jacq, C. (1989) *Cell* **56,** 431–441.
6. Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K. & Anraku, Y. (1990) *J. Biol. Chem.* **265,** 6726–6733.
7. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M. & Stevens, T. H. (1990) *Science* **250,** 651–657.
8. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185,** 60–89.
9. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
10. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1194–1198.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
12. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987) *Methods Enzymol.* **154,** 367–382.
13. Simons, R. W., Houman, F. & Kleckner, N. (1987) *Gene* **53,** 85–96.
14. Wang, T. S.-F., Wong, S. W. & Korn, D. (1989) *FASEB J.* **3,** 14–21.
15. Hager, D. A. & Burgess, R. R. (1980) *Anal. Biochem.* **109,** 76–86.
16. Mattila, P., Korpela, J., Tenkanen, T. & Pitkanen, K. (1991) *Nucleic Acids Res.* **19,** 4967–4973.
17. Looney, M. C., Moran, L. S., Jack, W. E., Feehery, G. R., Benner, J. S., Slatko, B. E. & Wilson, G. G. (1989) *Gene* **80,** 193–208.
18. Matsudaira, P. (1987) *J. Biol. Chem.* **262,** 10035–10038.
19. Reiter, W., Hudepohl, U. & Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9509–9513.
20. Leavitt, M. C. & Ito, J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4465–4469.
21. Bernad, A., Blanco, L., Lazaro, J. M., Martin, G. & Salas, M. (1989) *Cell* **59,** 219–228.
22. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
23. Russell, D. W., Jensen, R., Zoller, M. J., Burker, J., Errede, B., Smith, M. & Herskowitz, I. (1986) *Mol. Cell. Biol.* **6,** 4281–4294.
24. Carrington, D. M., Auffret, A. & Hanke, D. E. (1985) *Nature (London)* **313,** 64–67.
25. van Poelje, P. D., Kamath, A. V. & Snell, E. E. (1990) *Biochemistry* **29,** 10413–10418.
26. Aposhian, H. V. & Kornberg, A. (1962) *J. Biol. Chem.* **237,** 519–525.