Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 6658–6662, July 1992
Biochemistry

Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage

MRIDULA SHARMA, RICHARD L. ELLIS*, AND DEBORAH M. HINTON†

Section on Nucleic Acid Biochemistry, Laboratory of Biochemical Pharmacology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 8, Room 2A-13, Bethesda, MD 20892

Communicated by Herbert Tabor, April 27, 1992 (received for review March 3, 1992)

ABSTRACT **The bacteriophage T4 *segA* gene lies in a genetically unmapped region between the gene *βgt* (β-glucosyltransferase) and *uvsX* (recombination protein) and encodes a protein of 221 amino acids. We have found that the first 100 amino acids of the SegA protein are highly similar to the N termini of four other predicted T4 proteins, also of unknown function. Together these five proteins, SegA–E (similar to *e*ndonucleases of *g*roup I introns), contain regions of similarity to the endonuclease I-*Tev* I, which is encoded by the mobile group I intron of the T4 *td* gene, and to putative endonucleases of group I introns present in the mitochondria of *Neurospora crassa*, *Podospora anserina*, and *Saccharomyces douglasii*. Intron-encoded endonucleases are required for the movement (homing) of the intron DNA into an intronless gene, cutting at or near the site of intron insertion. Our *in vitro* assays indicate that SegA, like I-*Tev* I, is a Mg^{2+}-dependent DNA endonuclease that has preferred sites for cutting. Unlike the I-*Tev* I gene, however, there is no evidence that *segA* (or the other *seg* genes) resides within introns. Thus, it is possible that *segA* encodes an endonuclease that is involved in the movement of the endonuclease-encoding DNA rather than in the homing of an intron.**

Group I introns have been found in the DNA of many nonmetazoan species, including the nuclear DNA of *Tetrahymena*, the mitochondrial genomes of fungi, the genomes of cyanobacteria and bacteriophages, and chloroplast DNA (for reviews see refs. 1–5). Despite their locations in a wide range of organisms, these introns share similar sequence and structural features, as well as a common pathway for splicing. In addition, some group I introns are mobile, transferring at a very high frequency to their respective intronless genes and creating the precise intron–exon junctions. This "homing" movement is distinct from transposition in which an element moves to unrelated insertion sites.

Intron homing was first described for ω, a group I intron in the mitochondrial 21S rRNA gene of *Saccharomyces cerevisiae* (6). Movement of ω is dependent on the endonuclease I-*Sce* I, which is encoded within the ω intron sequence (7, 8). Other group I introns that encode specific endonucleases needed for homing include an intron in the rRNA gene of *Physarum polycephalum* nuclear DNA (9), two introns in the bacteriophage T4 genome (within the *td* and *sunY* genes) (10), a second intron in the *S. cerevisiae* mitochondrial DNA (11, 12), and introns in *Chlamydomonas* chloroplast DNA (13, 14). The *S. cerevisiae*, *Physarum*, and *Chlamydomonas* endonucleases catalyze double-strand (ds) breaks very close to the insertion site (11, 12, 14–17) whereas the phage enzymes cut several base pairs away (18, 19). The cuts are thought to initiate the insertion of the intron into the cleaved site by ds break repair (20).

Although several group I introns are competent both to home and to splice, the sequence elements necessary for splicing appear to be functionally and evolutionarily independent of the DNA that encodes the endonuclease. For example, while the sequences required for splicing of the group I introns present in the T4 *td* and *sunY* genes are highly conserved, the encoded endonucleases vary in both sequence and position (21). Likewise, the homing of the intron in the bacteriophage T4 *td* gene is independent of the sequences and secondary structure required by the intron to splice (18). These findings have supported the idea that the endonuclease-encoding DNAs rather than the introns were the original mobile elements and that introns simply provided locations where insertion of the mobile element could be tolerated by the host genome (3–5, 18). However, such a "mobile endonuclease," which is independent of an intron, has not been found.

Here we report the sequence of a previously unidentified bacteriophage T4 gene, *segA* (*s*imilar to *e*ndonucleases of *g*roup I introns),‡ which is located in an intergenic region of the T4 genome, between the genes *βgt* (β-glucosyltransferase) and *uvsX* (recombination protein) (Fig. 1*A*). We show that *segA* encodes a 25-kDa endodeoxyribonuclease, which, like that encoded by the T4 *td* intron, is Mg^{2+}-dependent and cuts with some site specificity. The sequence of the SegA protein and the sequences of four other open reading frames (ORFs) present on the T4 genome (*segB* through *segE*) share regions of homology with the T4 *td* intron endonuclease and with putative endonucleases of group I introns present in *Neurospora crassa*, *Podospora anserina*, and *Saccharomyces douglasii* mitochondria. Our findings suggest that the *seg* genes may represent the insertion of mobile endonuclease-encoding DNAs into intergenic regions of the T4 genome.

MATERIALS AND METHODS

Oligomers and Plasmids. Oligodeoxyribonucleotides were synthesized and purified as described (26). Plasmids pDH428 (27) and pKG1810R3 (28) have been described, and the vectors pTZ18U and pTZ19U were from United States Biochemical. New plasmids were created by published techniques (26). pRP101 and pRP102 were constructed by inserting the 557-bp *Xmn* I fragment from pDH428 (positions 702–1258; Fig. 1*B*) into the *Sma* I site of pTZ18U and pTZ19U, respectively. In pRP101, the region from the *segA*

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: ss, single-strand; ds, double-strand; ORF, open reading frame; aa, amino acid(s).

*Present address: Department of Internal Medicine, Southern Illinois University School of Medicine, Springfield, IL 62794.
†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank database (accession no. M69268).

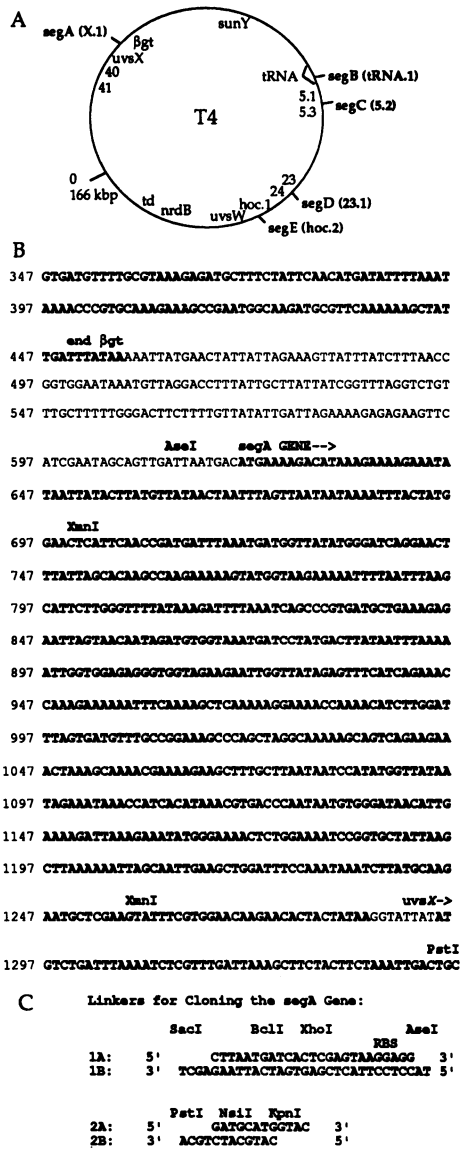

FIG. 1. Relative position of the T4 *segA* gene and its DNA sequence. (*A*) Location of the *seg* genes on the T4 genome (previous designation for each gene is in parentheses): *segA* (X.1) between *βgt* and *uvsX* (this paper), *segB* (tRNA.1) between the tRNA subclusters I and II (22), *segC* (*5.2*) between genes *5.1* and *5.3* (G. Mosig, personal communication), *segD* (*23.1*) between genes *23* and *24* (G. Yasuda, personal communication), and *segE* (*hoc.2*) between *hoc.1* and *uvsW* (23). [The T4 genome is 166 kilobase pairs (kbp). See ref. 24 for map positions of T4 genes.] The relative locations of the three T4 genes known to contain group I introns (*sunY*, *nrdB*, and *td*) are also shown. (*B*) Sequence of the T4 DNA from the end of the *βgt* gene to the beginning of the *uvsX* gene. Numbers are relative to the *Eco*RI site (T4 map unit 24.495) within *βgt* (see Fig. 3 *Upper*). Sequence from position 347 to 1313 was determined on both strands of the DNA. The sequence from 1314 to 1347 is from ref. 25. Restriction sites used for cloning are shown. (*C*) Sequences of the linkers 1A/1B and 2A/2B used to clone *segA* in pMS216. RBS, ribosome binding site.

gene is in the same orientation as the T7 promoter; in pRP102, *segA* is in the opposite orientation. pRP125 was constructed by inserting the 1094-bp *Xmn* I fragment from pDH428 (from 393 bp upstream of the T4 insert in pDH428 to position 701; Fig. 1*B*) into the *Sma* I site of pTZ18U in the same orientation as the T7 promoter.

pMS216, which contains *segA* downstream of a good ribosome binding site and the strong λ promoter P_L, was constructed as follows. A P_L vector (pMS114), having single *Sac* I and *Kpn* I sites just downstream of the promoter, was constructed from the plasmid pDH911 (29) by replacing the T4 DNA with a synthetic multiple cloning site (5'-GATCCAGATCTCCCGGGAGCTCGGTACCATGCATG-3'). The *segA* gene was cloned into pMS114 by a two-step process. The 735-bp *Ase* I–*Pst* I fragment from pDH428 (positions 614–1348; Fig. 1*B*) was first ligated to the synthetic oligomers 1A/1B and 2A/2B (Fig. 1*C*). This created a 5' *Sac* I site and a 3' *Kpn* I site and positioned a good ribosome binding site upstream of *segA*. Insertion of this ligated fragment into a vector derived from pTZ18U resulted in pRE201. pMS216 was constructed by inserting the 772-bp *Sac* I–*Kpn* I fragment from pRE201 between the *Sac* I and *Kpn* I sites of pMS114.

pMS510 is identical to pMS216 except that the CA dinucleotide at positions 630 and 631 has been replaced by TCT (Fig. 1*B*). This +1 frameshift truncates SegA after 4 amino acids (aa). (The construction of pMS510 will be published elsewhere.)

DNA Sequence Analysis. Dideoxy sequencing reactions (30) were performed with previously described protocols (26, 31, 32). The single-strand DNAs (ssDNAs) of pRP101, pRP102, and pRP125 were primed with the universal primer that anneals just downstream of the multiple cloning site. pDH428 plasmid DNA was annealed to oligodeoxynucleotide primers X2 (positions 489–509), X3 (728–709), X4 (999–975), X5 (964–983), X6 (1090–1068), X7 (1349–1329), and X8 (501–476). [Positions as indicated in Fig. 1*B* are relative to the *Eco*RI site within the *βgt* gene, T4 map unit 24.495 (see ref. 24).] The sequence at the beginning of *segA* on T4D$^+$ genomic DNA and RNA was determined by using the oligomer X3.

Searches to find other proteins similar to SegA were performed with the FASTA and TFASTA programs of the Genetics Computer Group sequence-analysis software package, version 7.0 (33) and databases from GenBank (Release 69.0), Genpept (Release 69.0), EMBL (Release 27.0), and the bacteriophage T4 sequence database from E. Kutter (The Evergreen State College, Olympia, WA).

***In Vitro* Transcription/Translation and ssDNA Chromatography.** Proteins expressed by pDH428 or by pKG1810R3 were labeled with [^{35}S]methionine *in vitro* by using the S30 extract, transcription/translation mixture, and protocol from DuPont/NEN Products. A portion of the pDH428 reaction mixture was added to a fraction of proteins isolated from 45 ml of exponentially growing *Escherichia coli* DH1/pDH428 cells (29) that had been partially purified by phosphocellulose chromatography. This mixture of proteins made *in vivo* and *in vitro* was then applied to a 50-μl ssDNA-cellulose column. The column was washed with 0.3 ml of sonication buffer (26), then successively with 75 μl of sonication buffer with 0.2, 0.5, 1.0, and 2.0 M NaCl. The proteins from the *in vitro* reactions and proteins present in the column fractions (aliquots of 15 μl) were separated by SDS/20% PAGE (34).

***In Vitro* Endonuclease Assay.** Protein fractions from N4830 cells (35) containing pMS216 (wild-type *segA* plasmid) or pMS510 (mutant *segA*) were obtained after DEAE and phosphocellulose chromatography. (Details of purification will be published elsewhere.) Protein fractions (10 ng of total protein, ≈2 ng of SegA) were incubated with 0.03 pmol of pDH428 DNA for 30 min at 30°C in 8 μl of 20 mM Tris·HCl, pH 8.0/10% (vol/vol) glycerol/1 mM 2-mercaptoethanol/2 mM ATP/5 mM $MgCl_2$. Reaction mixtures were extracted with phenol and precipitated with ethanol before restriction analysis. Reaction products were separated in a 0.8% agarose gel.

RESULTS

Region Between *βgt* and *uvsX* Encodes a Protein of 221 aa That Has Regions of Similarity to Proteins Encoded by Group

I Introns. Although genetic analyses have failed to locate a gene in the 838 bp between the T4 genes *βgt* and *uvsX* (Fig. 1*A*), this region is expressed as prereplicative, middle RNA along with the downstream genes *uvsX*, *40*, and *41* (26, 36). Our sequence of this region predicts a protein of 221 aa, starting 165 bp downstream of *βgt* and ending just before *uvsX* (Fig. 1*B*; ref. 37). We originally designated this gene X.1 but have changed the name to *segA* based on the analyses detailed below. We have assigned the start of the SegA protein to the first methionine codon, at position 621. However, neither this ATG nor farther downstream ATG codons are preceded by a good ribosome binding site. To eliminate the possibility that the ribosome binding site had suffered a mutation during our cloning of the DNA, we determined the sequence at the beginning of the gene, using both genomic T4 DNA and T4 RNA. This analysis indicated that our sequence at the beginning of *segA* was correct. In addition, it eliminated the possibility that RNA splicing generates a message having a different sequence upstream of the gene.

We searched GenBank, Genpept, EMBL, and the available bacteriophage T4 sequence (representing about 85% of the phage DNA) for proteins having similarity to SegA. Our analysis revealed four other ORFs present on T4, *segB*, *-C*, *-E*, and *-D*, whose predicted products are 53%, 51%, 47%, and 42% identical, respectively, with SegA between aa 8 and 98 (Fig. 1*A* and Fig. 2). These putative proteins are similar in size to SegA (229, 140, 208, and 205 aa for SegB through SegE, respectively). Like *segA*, these other *seg* genes have been found by DNA sequencing and their biological functions are unknown.

The next best match to SegA after the above four T4 proteins is the ORF residing in Nccob.1, the first intron of the *N. crassa* mitochondrial apocytochrome *b* gene (38). SegA and the Nccob.1-encoded protein have 21% identity over 150 aa, a portion of which is shown in Fig. 2. The Nccob.1 protein belongs to a subset of 12 proteins encoded by group I introns that share homology with each other and with the T4 *td* intron endonuclease, I-*Tev* I (38, 39, 41, 44, 45). Some of the homology found in the Seg proteins overlaps regions of homology among these group I intron proteins (shown in bold in Fig. 2). I-*Tev* I is an endonuclease required for the homing of the *td* intron DNA (10), and from genetic analyses, the fungal Ncnd1.1 and Nccob.1 introns are suspected to home to their respective intronless sites (46, 47). Thus, based on the similarity of SegA and the four other similar T4 ORFs with I-*Tev* I and the fungal intron proteins, we propose that these T4 genes be designated as *segA–E* (*s*imilar to *e*ndonucleases of *g*roup I introns) (Fig. 1*A*).

A 25-kDa Protein, Consistent with the Size Predicted by the *segA* ORF, Is Expressed by pDH428. Plasmid pDH428 contains the region of T4 DNA from the end of *βgt* through gene *41* in the pBR322-based vector pKG1810R3 (Fig. 3 *Upper*; ref. 28). Using a coupled transcription/translation system, we could identify six proteins expressed by this plasmid *in vitro*: the products of the T4 genes *uvsX*, *40*, and *41*, the vector-encoded proteins β-lactamase and galactokinase, and a protein of 25 kDa, consistent with the *segA* ORF (Fig. 3, lane 2). A control reaction using the vector pKG1810R3 (Fig. 3, lane 1) produced only β-lactamase. (Galactokinase is not expressed by pKG1810R3 because no promoter for *galK* is present on the vector.)

Because the predicted amino acid sequence for SegA indicates a highly basic protein, we applied the products of the *in vitro* transcription/translation reaction to a ssDNA-cellulose column (Fig. 3, lanes 3–7). As expected, galactokinase is not retained by this column whereas the recombination protein UvsX, previously shown to bind to ssDNA cellulose (48), is adsorbed. The 25-kDa protein also binds to ssDNA-cellulose, and the bulk of it is eluted with 0.5 M NaCl (Fig. 3, lane 5).

SegA Is a Mg^{2+}-Dependent Endodeoxyribonuclease with Some Site Specificity. To express large amounts of SegA, the *segA* gene with an improved ribosome binding site was cloned downstream of the strong λ promoter P_L. The resulting plasmid, pMS216, was transformed into the *E. coli* strain N4830, which has the temperature-sensitive λ repressor

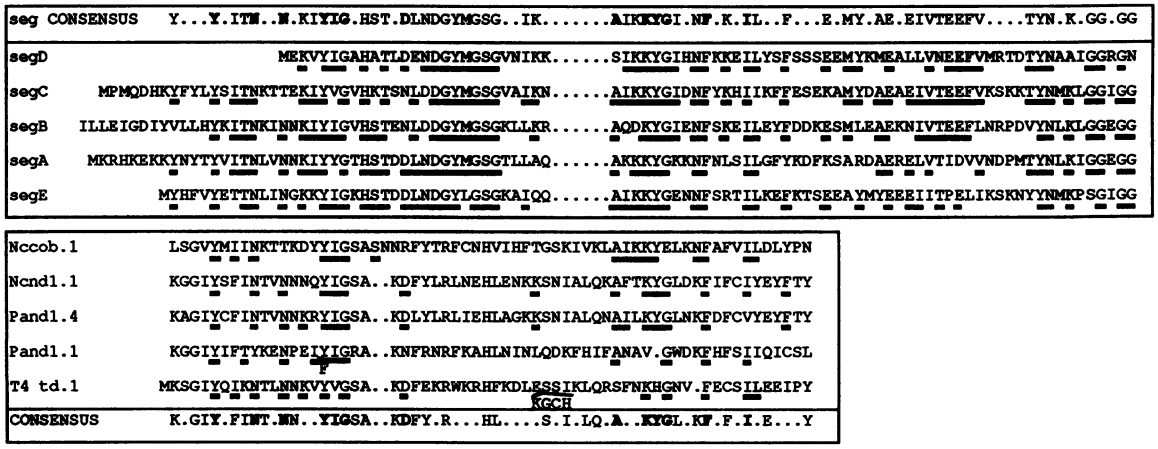

FIG. 2. Amino acid alignment of the T4 Seg proteins and comparison to proteins encoded by some group I introns present in *N. crassa*, *P. anserina*, and T4. The N-terminal portions of the Seg protein sequences are given (see Fig. 1 legend for refs.), starting with the first methionine (for SegA, -C, -D, and -E) or the first amino acid (for SegB). Sequences of proteins encoded by the following group I introns are shown: Nccob.1, first intron of the mitochondrial apocytochrome *b* gene of *N. crassa* (38), sequence shown starts at aa 100; Ncnd1.1, first intron of the mitochondrial *nd1* gene of *N. crassa* (39), sequence shown starts at aa 85; Pand1.1 and Pand1.4, first and fourth introns, respectively, of the mitochondrial *nd1* gene of *P. anserina* (40, 41), sequences shown start at aa 63 and 66, respectively; T4 td.1, I-*Tev* I endonuclease, encoded by the intron in the *td* gene of phage T4 (42, 43), sequence shown starts at aa 1. An amino acid is listed in the consensus sequence for the Seg proteins (shown at the top) or for the group I intron proteins (shown at the bottom) if that amino acid is found in three of the five proteins in the boxed group. Identical amino acids in the two consensus sequences are shown in bold. An amino acid is underlined if it matches the Seg consensus sequence. Statistical analysis for pairwise comparisons of SegA aa 8–98 with the corresponding region of the other proteins (using the ALIGN program of the Protein Identification Resource of the National Biomedical Research Foundation, Georgetown University Medical Center, Washington) gave the following standard deviations of real score above the best score from 200 randomizations: SegA vs. SegB, 15.8; vs. SegC, 15.2; vs. SegD, 7.5; vs. SegE, 14.9; vs. Nccob.1, 4.9; vs. I-*Tev* I, 2.6. SegB and SegE gave higher values, 4.2 and 5.1, respectively, in a similar comparison with I-*Tev* I.

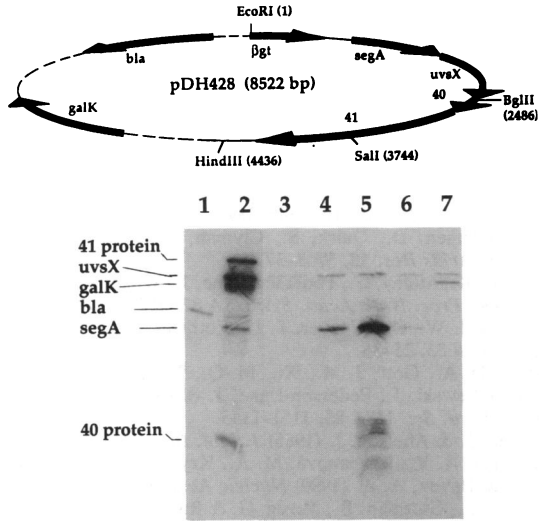

FIG. 3. *In vitro* expression of a 25-kDa DNA-binding protein by pDH428. Reactions and the ssDNA-cellulose column chromatography were performed as described in *Materials and Methods*. (*Upper*) Diagram of pDH428 showing the locations of genes and restriction sites discussed in the text. T4 DNA insert is represented by the unbroken line, the vector DNA by the dashed line. (*Lower*) Autoradiogram of labeled transcription/translation products separated by SDS/20% PAGE. Lanes: 1, proteins expressed by pKG1810R3 (vector); 2, proteins expressed by pDH428; 3–7, fractions from the ssDNA-cellulose column after application of the proteins expressed by pDH428 and washing with buffer plus 0.05 M (lane 7), 0.2 M (lane 6), 0.5 M (lane 5), 1.0 M (lane 4), and 2.0 M (lane 3) NaCl. Proteins expressed by pDH428 [41 protein, UvsX, galactokinase (GalK), β-lactamase (Bla), 40 protein, and SegA] were identified by size and were confirmed by using plasmids or restriction fragments that contained only portions of the T4 DNA present in pDH428 (ref. 28 and D.M.H., unpublished data). Low molecular weight bands in lane 5 may represent degradation products due to proteolysis.

cI857 (35). Two and one-half hours after heat induction of P_L, a high level of the 25-kDa SegA protein is expressed by cells containing pMS216 (Fig. 4, lane 1). As expected, no such protein is induced in cells containing the vector (lanes 5–8).

To assay the *in vitro* activity of SegA protein with DNA, we partially purified SegA expressed by N4830/pMS216 cells (M.S. and D.M.H., unpublished data). As a control for these assays, we used a similarly induced and purified fraction from N4830 cells containing pMS510, a plasmid that is identical to pMS216 except that a short DNA substitution truncates the

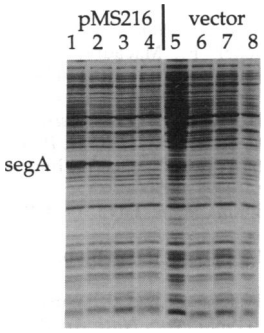

FIG. 4. pMS216 expresses high levels of SegA *in vivo*. N4830 cells (cI857; ref. 35) containing either pMS216 (wild-type *segA* plasmid, lanes 1–4) or pMS114 (vector, lanes 5–8) were grown in Luria broth plus ampicillin (25 μg/ml) at 29°C to mid-logarithmic phase. Proteins of crude extracts (29) were separated by SDS/20% PAGE using samples taken immediately before (lanes 4 and 8), 30 min after (lanes 3 and 7), 1 hr after (lanes 2 and 6), or 2.5 hr after (lanes 1 and 5) cells were shifted to 42°C.

segA gene after 4 aa (see *Materials and Methods*). In the presence of Mg^{2+}, both supercoiled plasmid (Fig. 5) and linear wild-type T4 DNA (M.S. and D.M.H., unpublished data) are cut by fractions containing the SegA protein. Incubation of pDH428, the 8522-bp plasmid that contains the T4 sequence from the end of *βgt* through gene *41* (Fig. 3 *Upper*), with SegA produces discrete linear products, migrating as ds linear plus smaller DNAs (Fig. 5, lanes 9 and 13). This reaction requires the addition of Mg^{2+} (compare lanes 6 and 9) and is stimulated by the addition of ATP (compare lanes 4 and 9). Using a reaction in which the linearized plasmid species was the main product (a reaction performed with a less active fraction of SegA; lane 13), we analyzed this product by digestion with restriction enzymes. Digestion with *Sal* I, which has a single restriction site in pDH428 at position 3744 (Fig. 3 *Upper*), produces DNAs of 6.7 and 1.8 kbp (Fig. 5, lane 12). Digestion with *Bgl* II, which has one site in the plasmid at position 2486 (Fig. 3 *Upper*), produces a DNA of 0.6 kbp and a species slightly smaller than ds pDH428 (Fig. 5, lane 10). This analysis maps a SegA endonuclease site in pDH428 to near position 1900, within the *uvsX* gene. From primer extension analyses (M.S. and D.M.H., unpublished data), the sequence surrounding this site has been determined:

```
      ↓↓
5′-AAACACAAGAAATGTTTAGT-3′
3′-TTTGTGTTCTTTACAAATCA-5′
         ↑↑
```

However, it should be noted that SegA cuts pDH428 at several other sites (Fig. 5, lane 9; M.S. and D.M.H., unpublished data), suggesting that there is a hierarchy of preferred sites rather than the recognition of a specific sequence.

DISCUSSION

Although the proteins encoded by group I introns vary in function and sequence, they have been classified into families based on the presence of conserved amino acid motifs. One such family includes I-*Tev* I, the endonuclease encoded by the T4 *td* intron and required for homing of that intron into intronless phage (10, 42, 43). Also included in this family are several ORFs found in group I introns in the mitochondria of *N. crassa*, *P. anserina*, and *S. douglasii* (44, 45). Our sequence analyses suggest that five other T4 ORFs, *segA–E*, located within intergenic regions of the phage genome belong to the I-*Tev* I family. Our finding is strengthened by *in vitro*

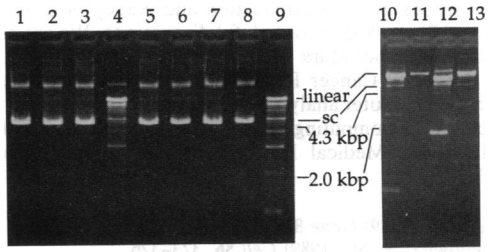

FIG. 5. SegA has Mg^{2+}-dependent endodeoxyribonuclease activity. Supercoiled pDH428 DNA was incubated with protein fractions from N4830/pMS216 cells (wild-type *segA* plasmid; lanes 2, 4, 6, 9, 10, 12, and 13), with equivalent fractions from N4830/pMS510 cells (mutant *segA* plasmid in which the protein is truncated after 4 aa; lanes 1, 3, 5, and 8) or without protein (lane 7). Lanes 1 and 2, reaction without $MgCl_2$ and ATP; lanes 3 and 4, reaction without ATP; lanes 5 and 6, reaction without $MgCl_2$; lane 7, reaction without protein; lanes 8, 9, and 13, complete reaction; lanes 10 and 12, digestion of products from lane 13 with *Bgl* II and *Sal* I, respectively. Positions of molecular size markers and supercoiled (sc) and linear pDH428 are shown. Lane 11 shows the position of linear pDH428 obtained after digestion of the plasmid with *Bgl* II.

assays indicating that SegA is a Mg^{2+}-dependent DNA endonuclease that cuts DNA with some site specificity and, thus, is functionally similar to I-*Tev* I. Similarly, previous workers have found that the site-specific endonuclease encoded by the mitochondrial *ENS2* gene of *S. cerevisiae* and *S. uvarum*, a gene that is not located within an intron, contains two amino acid motifs common to another family of group I intron-encoded endonucleases (49–51).

The *td* intron-encoded protein, I-*Tev* I, is required for the homing of the *td* intron DNA but appears to have no role in RNA splicing (52, 53). Furthermore, mutations that abolish splicing are still competent to home as long as the intron-encoded endonuclease retains activity (18), and phylogenetic analyses suggest that the group I introns themselves evolved independently from the proteins they sometimes encode (refs. 54 and 55, reviewed in refs. 1–5). These findings have led to the speculation that it was the endonucleases rather than the introns which were the original mobile elements (3–5, 18). The intron simply provided a location where insertion of the endonuclease would not harm the host. By this model, movement of the endonuclease into intergenic regions could also occur if the expression of the surrounding genes was not adversely affected. Features of the T4 *seg* family and specifically the *segA* gene are consistent with this postulated intergenic insertion. All the *seg* genes appear to be in intergenic regions of the T4 genome, and despite extensive analyses of the RNA from the end of *βgt* through *41* both *in vivo* and *in vitro* (refs. 26 and 36; D.M.H., unpublished data), we have no evidence that *segA* is within an intron. As with I-*Tev* I, the T4 *td* intron endonuclease, no biological function for the phage has been ascribed to any of the *seg* genes. In particular, we have been unable to find a defect in a T4 *segA* amber mutant constructed by *in vitro* mutagenesis (M.S. and D.M.H., unpublished data), and T4 deletion mutants that remove *segB* (as well as the surrounding tRNA genes) have been isolated and are viable under normal growth conditions (56). Finally, we have recently found that phage T2, which is closely related to T4, lacks DNA from inside the *βgt* gene up to the beginning of *uvsX*, a region that includes *segA* (M.S. and D.M.H., unpublished data). Together, these observations are consistent with the idea that the *seg* genes arose from the movement of endonuclease-encoding DNA, rather than an intron, into the T4 genome. However, it is not clear whether the *seg* genes are mobile themselves or whether they simply have a common origin with the endonucleases present in some mobile group I introns.

We thank Gisela Mosig and Glenn Yasuda for communicating their DNA sequences prior to publication and Nancy Nossal, Anthony Furano, and Tania Baker for helpful discussions. We are grateful to Gary Smythers and Mark Gunnell (Biomedical Supercomputing Center, Frederick Cancer Research and Development Center) for help with the computer analyses and to the National Cancer Institute for allocation of computing time and staff support. R.L.E. was a Howard Hughes Medical Institute–National Institutes of Health Research Scholar.

1. Dujon, B. (1989) *Gene* **82**, 91–114.
2. Lambowitz, A. M. (1989) *Cell* **56**, 323–326.
3. Perlman, P. S. & Butow, R. A. (1989) *Science* **246**, 1106–1109.
4. Belfort, M. (1990) *Annu. Rev. Genet.* **24**, 363–385.
5. Belfort, M. (1991) *Cell* **64**, 9–11.
6. Dujon, B. (1980) *Cell* **20**, 185–197.
7. Jacquier, A. & Dujon, B. (1985) *Cell* **41**, 383–394.
8. Macreadie, I. G., Scott, R. M., Zinn, A. R. & Butow, R. A. (1985) *Cell* **41**, 395–402.
9. Muscarella, D. E. & Vogt, V. M. (1989) *Cell* **56**, 443–454.
10. Quirk, S. M., Bell-Pedersen, D. & Belfort, M. (1989) *Cell* **56**, 455–465.
11. Delahodde, A., Goguel, V., Becam, A. M., Creusot, F., Perea, J., Banroques, J. & Jacq, C. (1989) *Cell* **56**, 431–441.
12. Wenzlau, J. M., Saldanha, R. J., Butow, R. A. & Perlman, P. S. (1989) *Cell* **56**, 421–430.
13. Durrenberger, F. & Rochaix, J.-D. (1991) *EMBO J.* **10**, 3495–3501.
14. Marshall, P. & Lemieux, C. (1991) *Gene* **104**, 241–245.
15. Colleaux, L., D'Auriol, L., Galibert, F. & Dujon, B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6022–6026.
16. Wernette, C. M., Saldahna, R., Perlman, P. S. & Butow, R. A. (1990) *J. Biol. Chem.* **265**, 18976–18982.
17. Muscarella, D. E., Ellison, E. L., Ruoff, B. M. & Vogt, V. M. (1990) *Mol. Cell. Biol.* **10**, 3386–3396.
18. Bell-Pedersen, D., Quirk, S., Clyman, J. & Belfort, M. (1990) *Nucleic Acids Res.* **18**, 3763–3770.
19. Chu, F. K., Maley, G., Pedersen-Lane, J., Wang, A.-M. & Maley, F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3574–3578.
20. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. (1983) *Cell* **33**, 25–35.
21. Shub, D. A., Gott, J. M., Xu, M.-Q., Lang, B. F., Michel, F., Tomaschewski, J., Pedersen-Lane, J. & Belfort, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1151–1155.
22. Broida, J. & Abelson, J. (1985) *J. Mol. Biol.* **185**, 545–563.
23. Kaliman, A. V., Khasanova, M. A., Kryukov, V. M., Tanyashin, V. I. & Bayev, A. A. (1990) *Nucleic Acids Res.* **18**, 4277.
24. Kutter, E., Guttman, B., Mosig, G. & Ruger, W. (1990) in *Genetic Maps*, ed. O'Brien, S. J. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 1.24–1.51.
25. Fujisawa, H., Yonesaki, T. & Minagawa, T. (1985) *Nucleic Acids Res.* **13**, 7473–7481.
26. Hinton, D. M. (1991) *J. Biol. Chem.* **266**, 18034–18044.
27. Hinton, D. M., Silver, L. L. & Nossal, N. G. (1985) *J. Biol. Chem.* **260**, 12851–12857.
28. Hinton, D. M. & Nossal, N. G. (1986) *J. Biol. Chem.* **261**, 5663–5673.
29. Hinton, D. M. & Nossal, N. G. (1985) *J. Biol. Chem.* **260**, 12858–12865.
30. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
31. Hinton, D. M. (1989) *J. Biol. Chem.* **264**, 14440–14446.
32. Slatko, B. E., Heinrich, P., Nixon, B. T. & Eckert, R. L. (1991) in *Current Protocols in Molecular Biology*, eds. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (Wiley, New York), pp. 7.3.1–7.3.9.
33. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
34. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
35. Adhya, S. & Gottesman, M. (1982) *Cell* **29**, 939–944.
36. Hinton, D. M. (1989) *J. Biol. Chem.* **264**, 14432–14439.
37. Ellis, R. L. & Hinton, D. M. (1989) *J. Cell. Biochem.* **13D**, 123 (abstr.).
38. Collins, R. A., Reynolds, C. A. & Olive, J. (1988) *Nucleic Acids Res.* **16**, 1125–1134.
39. Burger, G. & Werner, S. (1985) *J. Mol. Biol.* **186**, 231–242.
40. Cummings, D. J., MacNeil, I. A., Domenico, J. & Matsuura, E. T. (1985) *J. Mol. Biol.* **185**, 659–680.
41. Cummings, D. J., Domenico, J. M. & Michel, F. (1988) *Curr. Genet.* **14**, 253–264.
42. Chu, F. K., Maley, G. F., Maley, F. & Belfort, M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3049–3053.
43. Chu, F. K., Maley, G. F., West, D. K., Belfort, M. & Maley, F. (1986) *Cell* **45**, 157–166.
44. Michel, F. & Dujon, B. (1986) *Cell* **46**, 323.
45. Tian, G.-L., Michel, F., Macadre, C., Slonimski, P. P. & Lazowska, J. (1991) *J. Mol. Biol.* **218**, 747–760.
46. Mannella, C. A. & Lambowitz, A. M. (1979) *Genetics* **93**, 645–654.
47. Infanger, A. & Bertrand, H. (1986) *Curr. Genet.* **10**, 607–617.
48. Formosa, T. & Alberts, B. M. (1986) *J. Biol. Chem.* **261**, 6107–6118.
49. Shibata, T., Watabe, H., Kaneko, T., Iino, T. & Ando, T. (1984) *J. Biol. Chem.* **259**, 10499–10506.
50. Séraphin, B., Simon, M. & Faye, G. (1987) *J. Biol. Chem.* **262**, 10146–10153.
51. Nakagawa, K., Morishima, N. & Shibata, T. (1991) *J. Biol. Chem.* **266**, 1977–1984.
52. Ehrenman, K., Pedersen-Lane, J., West, D., Herman, R., Maley, F. & Belfort, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5875–5879.
53. Hall, D. H., Povinelli, C. M., Ehrenman, K., Pedersen-Lane, J., Chu, F. & Belfort, M. (1987) *Cell* **48**, 63–71.
54. Mota, E. M. & Collins, R. A. (1988) *Nature (London)* **332**, 654–656.
55. Colleaux, L., Michel-Wolwertz, M.-R., Matagne, R. F. & Dujon, B. (1990) *Mol. Gen. Genet.* **223**, 288–296.
56. Wilson, J. H., Kim, J. S. & Abelson, J. N. (1972) *J. Mol. Biol.* **71**, 547–556.