Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 11669–11673, November 1994
Evolution

An intron in the thymidylate synthase gene of *Bacillus* bacteriophage β22: Evidence for independent evolution of a gene, its group I intron, and the intron open reading frame

(bacteriophage T4/homing endonuclease/helix–turn–helix/self-splicing)

DAVID H. BECHHOFER*, KIM K. HUE*, AND DAVID A. SHUB†

*Department of Biochemistry, Mount Sinai School of Medicine, New York, NY 10029; and †Department of Biological Sciences and Center for Molecular Genetics, University at Albany, State University of New York, Albany, NY 12222

Communicated by John R. Roth, July 25, 1994 (received for review May 2, 1994)

ABSTRACT **The thymidylate synthase gene (*thy*) (EC 2.1.1.45) of *Bacillus subtilis* bacteriophage β22 has a self-splicing, group I intron inserted into a highly conserved region of the coding sequence. The intron is very similar to one that is inserted 21 bp further downstream in the homologous thymidylate synthase gene (*td*) of *Escherichia coli* bacteriophage T4. In contrast, the amino acid sequences of the bacteriophage thymidylate synthases are highly divergent. The β22 intron has a fragmentary open reading frame (ORF) that encodes a putative helix–turn–helix DNA-binding motif, similar to one at the carboxyl terminus of the homing endonuclease (I-TevI) encoded by the T4 *td* intron. The *td* ORF and the *thy* ORF fragments are inserted into different regions of their respective intron structures. These results suggest that the thymidylate synthase genes, their introns, and their respective intron-ORFs all have separate evolutionary histories and that the acquisition of the intron could not have occurred by a simple homing event.**

Several bacteriophage genes contain self-splicing group I introns. The *Escherichia coli* bacteriophage T4 genome includes three such genes: *td* (thymidylate synthase, TS, EC 2.1.1.45), *nrdB* (small subunit of ribonucleotide reductase), and *sunY* (*nrdD*, anaerobic ribonucleotide reductase) (reviewed in ref. 1). Like many other group I introns, the *td* and *sunY* introns contain open reading frames (ORFs) that code for site-specific (homing) DNA endonucleases involved in the transfer of the intron to intronless target sites (2). Although the T4 *nrdB* ORF does not code for a functional protein, it is a truncated version of one found in the *nrdB* intron of a related bacteriophage, RB3, whose product cleaves the intronless version of the T4 *nrdB* gene (3).

In addition to the coliphage introns, an intron was discovered in gene *31* (coding for DNA polymerase) of the *Bacillus subtilis* bacteriophage SPO1 (4). Despite considerable divergence in primary sequence between the SPO1 intron and the three closely related T4 introns, the four bacteriophage examples make up a distinct subgroup (IA2) within the set of all group I introns (5). The SPO1 intron also contains an ORF (4) that encodes a DNA endonuclease (H. Goodrich-Blair and D.A.S., unpublished data).

Speculation on the distribution and origin of introns in bacteria is restricted by the small number of introns that has been discovered so far. Besides the phage introns, these include a few examples of group I introns in tRNA genes and group II introns in protein-coding genes in both cyano- and proteo- ("purple") bacteria (6–11). Each additional example of a bacterial intron should yield new information that can help to clarify the origin and role of these genetic elements.

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.

β22, a virulent *B. subtilis* bacteriophage with a broad host range (12), contains a TS gene that has been cloned (13). In the course of sequencing this cloned gene, we encountered a stop codon that, when compared to the sequence of several previously sequenced TS genes, was located in the middle of the coding sequence. Here we show that the β22 *thy* gene contains a group I intron which is remarkably similar in sequence and structure to the intron in the *td* gene of phage T4.‡

MATERIALS AND METHODS

Cloning and Sequencing of the β22 *thy* Gene. The β22 bacteriophage *thy* gene was subcloned from plasmid pBD214 (13). A *Nsi* I–*Sna*BI fragment from the *thy* region of pBD214 was inserted into pGEM-7Zf(+) (Promega) between the *Nsi* I and *Sma* I sites. In sequencing this fragment we encountered a stop codon in the middle of the putative coding sequence. To determine the DNA sequence further downstream, the *Eco*RI–*Bgl* II fragment from the *thy* region was cloned into pGEM-3Zf(+) between the *Eco*RI and *Bam*HI sites. The remainder of the intron and TS coding sequences were determined from this clone.

***In-Vitro* Transcription and Splicing.** Transcription reactions were performed according to the supplier's instructions (14). T7 RNA polymerase transcripts were uniformly labeled with [α-^{32}P]UTP. Self-splicing of full-length products was carried out at 37°C or 60°C in transcription buffer containing 100 μM GTP.

RESULTS AND DISCUSSION

β22 *thy* Gene Sequence. The DNA sequence of a 1540 bp fragment from plasmid pBD214 (13) containing the β22 *thy* gene is shown in Fig. 1. The location of the *thy* coding sequence was deduced by identifying conserved TS amino acid residues (15). The coding sequence begins with a UUG codon at nucleotide 238 and is preceded by a ribosome binding site (GGAGG) that is located 9 nt upstream of the initiation codon. Upstream of this ribosome binding site is a potentially strong hairpin structure that is followed by a run of U residues, suggesting the presence of an upstream transcriptional terminator sequence.

TS Coding Sequence. An in-frame stop codon is located right after a conserved arginine codon at nucleotide 817. In all of the previously sequenced TS genes, this arginine residue is followed directly by a serine residue. Conserved TS codons in the β22 *thy* sequence resume 392 nt downstream of this

Abbreviations: H-T-H, helix–turn–helix; ORF, open reading frame; TS, thymidylate synthase.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. L31962).

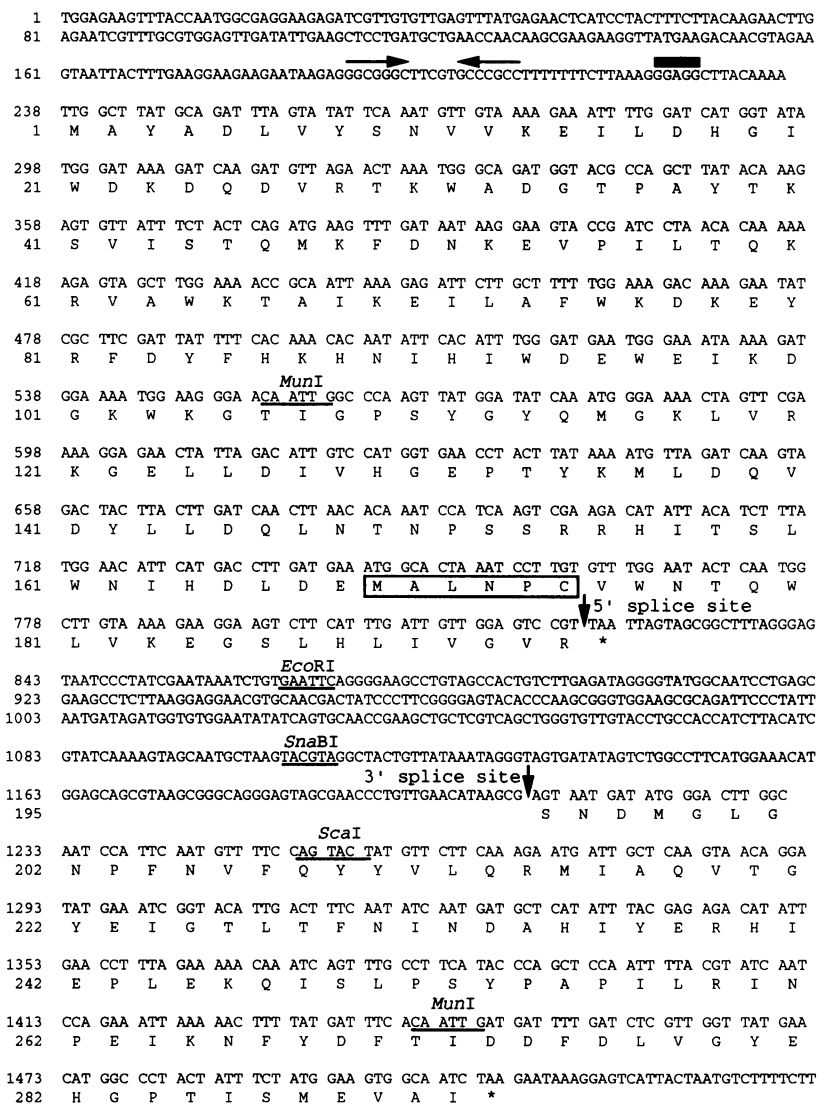

FIG. 1. β22 *thy* gene sequence. Nucleotide sequence of a fragment of β22 bacteriophage DNA containing the *thy* gene. The horizontal, facing arrows indicate a possible strong stem–loop structure followed by a run of Ts that could be a transcription termination signal. The probable *thy* ribosome binding site (GGAGG, horizontal bar) is downstream of this terminator sequence. Locations of the 5′ and 3′ splice sites and relevant restriction sites are indicated. Upper numbers on the left are nucleotide positions and lower numbers in the coding region are amino acid residues. Two in-frame stop codons are indicated by asterisks.

arginine codon with a serine codon at nucleotide 1212, followed two residues later by an invariant aspartic acid residue. To maintain a conserved TS amino acid sequence, β22 would have to contain a 392-nt intron, with the 5′ splice site located after a U residue at nucleotide 819 and the 3′ splice site located after a G residue at nucleotide 1211. These boundaries are characteristic of group I introns (5).

The amino acid sequence of β22 TS was compared to other TS proteins by the BESTFIT program of the Genetics Computer Group package (Table 1). β22 TS shared between 56% and 58% similarity with the TS genes of eukaryotes, with only slightly greater similarities to the TS genes of bacteria.

Table 1. Comparison of selected TS amino acid sequences

Organism	% similarity to β22 TS sequence	% identity to β22 TS sequence
Escherichia coli phage T4	51.7	30.2
Homo sapiens	55.8	36.7
Leishmania major	56.0	32.0
Saccharomyces cerevisiae	57.6	34.8
Escherichia coli	58.1	35.8
Staphylococcus aureus	59.2	35.0
Bacillus subtilis thyB	60.4	36.5
Bacillus subtilis phage ϕ3T	74.5	60.1

Comparisons to β22 TS amino acid sequence were made by the BESTFIT program. References for all amino acid sequences are given in ref. 15.

Similarity to *B. subtilis thyB* TS was not significantly greater than to other bacterial TSs. The greatest similarity (74.5%) was to TS of the temperate *B. subtilis* phage ϕ3T, whose *thy* gene does not contain an intron (16). Interestingly, phage ϕ3T *thy* resembles *B. subtilis thyA* (a second *B. subtilis* TS gene, whose sequence has not been reported) more closely than it does *thyB*, as determined by their phenotypic properties and nucleic acid hybridization (17). Remarkably, the β22 TS was least similar to the *E. coli* phage T4 TS.

The relationships between the β22 TS and the T4 and ϕ3T TS sequences were highlighted even more dramatically by the analysis of percent identity of amino acid residues. β22 TS amino acid residues shared only 30.2% identity with those of the T4 TS gene, but were 60.1% identical with the ϕ3T gene. In addition, the β22 TS sequence is distinct from other TS genes and more similar to the ϕ3T TS gene in the sequence near the active site. Amino acids 169–174 (boxed in Fig. 1) are conserved residues that constitute the active site, with the cysteine sulfhydryl reacting with substrate dUMP. The amino acid directly adjacent to the carboxyl end of this active site is a histidine in 11 TS sequences and appears to influence TS activity (18). However, a valine is present at this position in both the ϕ3T and β22 sequences. Together these findings imply that the *E. coli* T4 and *B. subtilis* β22 TS genes do not share a recent common ancestor.

Self-Splicing *in Vitro*. A *Mun* I fragment from plasmid pBD214 (Fig. 1) was cloned into the *Eco*RI site of plasmid pGEM-3Zf(+) to give plasmid pTS10. The inserted fragment

was oriented such that T7 RNA polymerase transcribes the sense strand to give the *thy* RNA sequence. Plasmid pTS10 DNA was digested with *Bam*HI, which cleaves the vector downstream of the inserted *thy* sequence, *Sca* I, which cleaves about 40 nt downstream of the 3′ splice site, or *Sna*BI, which cleaves within the intron sequence (Fig. 1). In addition, plasmid pTS11, which contained the *thy Mun* I fragment in the reverse orientation relative to the T7 RNA polymerase promoter, was linearized with *Bam*HI. Transcription products from these linearized DNAs were separated by electrophoresis on a denaturing 5% polyacrylamide gel (Fig. 2). Transcription of the *Sna*BI-linearized pTS10 DNA and the *Bam*HI-linearized pTS11 DNA each resulted in a single labeled RNA of the predicted size (560 nt and 920 nt, respectively). However, transcription of pTS10 DNA linearized with either *Bam*HI or *Sca* I resulted in four prominent RNAs. Two of the RNAs from the *Bam*HI-digested plasmid appeared to be the same size as two of the RNAs from the *Sca* I-digested plasmid as judged by their migrations on the gel. From the sizes of the predicted splicing products, we could identify the smaller of these two bands as the linear intron (approximately 390 nt). The full-length *Bam*HI-linearized pTS10 RNA comigrated with the *Bam*HI-linearized pTS11 RNA product (920 nt), and, as expected from the DNA sequence, the full-length *Sca* I-linearized pTS10 RNA product was about 210 nt shorter (710 nt). The bands labeled with asterisks in Fig. 2 were the predicted sizes of the ligated exon–exon RNAs derived from the *Bam*HI-linearized pTS10 transcripts and the *Sca* I-linearized pTS10 transcripts (530 nt and 320 nt, respectively). The large comigrating bands are circularized introns. The identity of these RNA species has been confirmed using intron- and exon-specific probes (data not shown). In denaturing 4% polyacrylamide gels (not shown) the circular intron RNA migrates faster than the full-length RNA species rather than slower as in the 5% denaturing gel shown in Fig. 2 (cf. also Fig. 3, which shows a 5% denaturing gel run at a lower voltage). This aberrant migration pattern is characteristic of circular RNA (19).

To follow the *in vitro* splicing reaction, *Bam*HI-linearized pTS10 DNA was transcribed at room temperature for 10 minutes to give a single major transcription product. This RNA was then incubated at 37°C or 60°C, and samples were removed every 2 min after the addition of 100 μM GTP (Fig. 3). The results show a temperature dependence of splicing similar to that demonstrated for other introns (20). At 37°C approximately 5% of the input RNA was spliced after 8 min, whereas at 60°C about 95% of the input RNA was spliced after only 2 min. At 60°C one can see a decrease over time in the amount of the linear intron and a corresponding increase

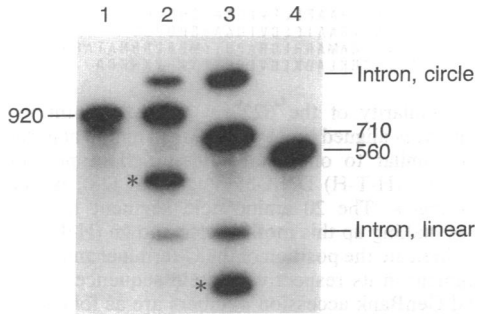

FIG. 2. RNA products from transcription (37°C for 60 min) of *thy* gene DNA. Lanes: 1, pTS11 digested with *Bam*HI; 2, pTS10 digested with *Bam*HI; 3, pTS10 digested with *Sca* I; and 4, pTS10 digested with *Sna*BI. Exon–exon RNAs are marked by asterisks. The positions of intron RNAs are indicated on the right. Predicted sizes (in nt) of the full-length transcription products are shown.

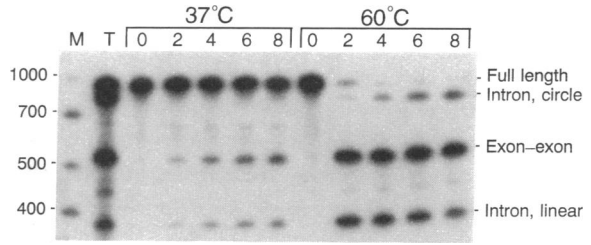

FIG. 3. Self-splicing of the pTS10 *Bam*HI RNA. Transcription of *Bam*HI-linearized pTS10 DNA at room temperature for 10 min gave a single RNA seen in the time 0 lanes. This RNA was incubated at either 37°C or 60°C, and samples were removed at 2-min intervals. The marker lane (M) contained 5′-end-labeled DNA fragments of the indicated sizes in nucleotides (Bio-Ventures Group, Portland, ME). Lane T contained the *in vitro* transcription products of *Bam*HI-digested pTS10 incubated at 37°C for 60 min.

in the amount of the circular form, while the amount of ligated exon–exon RNA remains relatively constant.

Secondary Structure of the Intron. Group I introns share a common structure. A secondary structure model, initially proposed on the basis of a few examples from fungal mitochondria (21, 22), is consistent with the sequences of a large number of subsequently described examples from extremely diverse biological sources (see ref. 5 for a recent compilation). The functional significance of the proposed hydrogen bonding interactions is supported by mutations that inactivate splicing and by second-site suppressors that restore base pairing, as well as by covariation of paired bases in closely related introns (23).

Group I introns have been divided into smaller subgroups according to shared elements of sequence and structure, with the four bacteriophage introns making up subgroup IA2 (5). Not surprisingly, within this subgroup the three introns that reside in the T4 genome are more closely related to each other than they are to the intron in *B. subtilis* bacteriophage SPO1.

Fig. 4 shows the β22 *thy* intron folded according to the group I secondary structure rules. The sequence contains every conserved group I structure element and, like the other phage introns, belongs to subgroup IA2. Surprisingly, the *Bacillus* bacteriophage β22 *thy* intron is more closely related, in both sequence and structure, to the T4 introns than to the one in SPO1.

Base identity at homologous positions of the T4 *td* intron is highlighted in Fig. 4 and extends both to normally conserved sequence elements and to regions that are not highly conserved among distantly related introns. For example, P9 is a stem–loop structure that is universally present in group I introns, typically comprising a 7-bp stem and a 4-base loop with the sequence GNRA, where N is any nucleotide and R is either G or A. There is little conservation of primary sequence in the P9 stem. Additionally, of the 87 group I introns compiled in ref. 5, P7.1 is limited to the 24 members of the IA subgroups and one member of subgroup IB4, among which there is substantial heterogeneity in both secondary structure and primary sequence. P7.1 and P9 are extraordinarily similar in the β22 *thy* and T4 *td* introns. The overall similarity of the β22 *thy* intron to the T4 introns remains high throughout (55% identity to *td* over 245 aligned residues), with similarity to the T4 *nrdB* intron even slightly exceeding that to *td* (not shown).

On the other hand, the intron in the DNA polymerase gene of *B. subtilis* phage SPO1 is more distantly related to the intron in β22. There are large differences in structure: the SPO1 intron has an unusual insertion between P3 and P4 and lacks P9.1 and P9.2. Also, P9 of SPO1 is much longer than P9 of β22 *thy* and is unrelated to it in primary sequence.

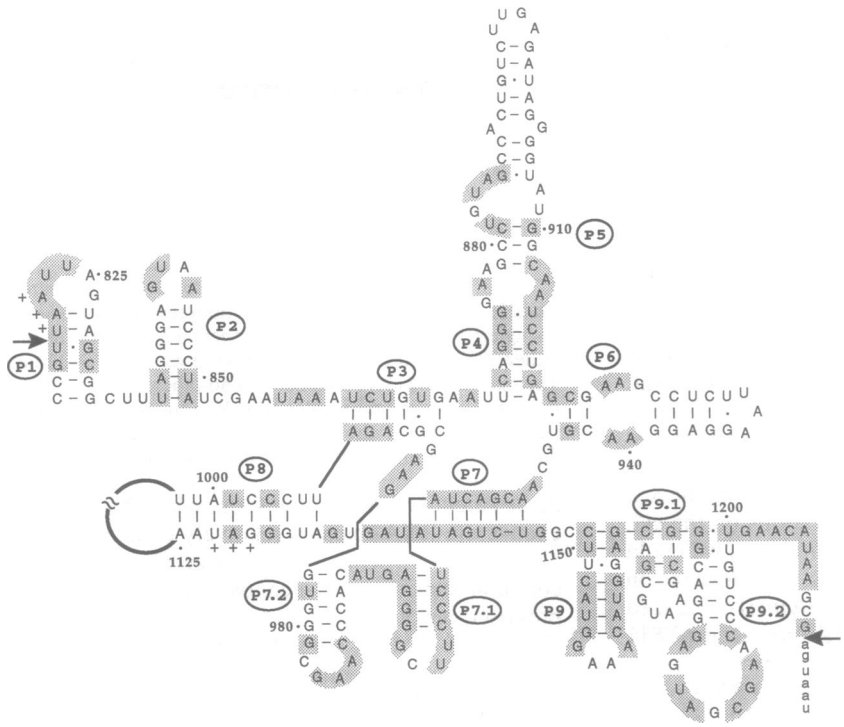

FIG. 4. Secondary structure of the β22 *thy* intron. Conserved base-paired regions are indicated (P). Arrows indicate splice boundaries. Bases identical to those at equivalent positions of the T4 *td* intron are shaded. Stop codons referenced in the text are indicated with +. Numbering is the same as in Fig. 1.

Additionally, only 50% of the SPO1 intron residues that can be aligned are identical to the *thy* intron residues of β22.

In contrast to P9 and P7.1, whose primary sequences are not conserved, residues 1134–1145 (the P8/P7 junction and the 3′ portion of P7) are highly conserved among all the group I introns (5). It is, therefore, not surprising that the β22 sequence should be identical at 10, 11, and 12 of the 12 positions with the *sunY*, *td*, and *nrdB* introns of T4, respectively. However, the first position of this sequence is almost universally an A, with a G in *td* and *nrdB* as the only exceptions in the compilation of Michel and Westhof (5). The second position is somewhat more variable, with A represented in a majority of cases and U as the major variant. Only *nrdB* shares both of these variations, as well as identity at all the other positions, with β22.

A Truncated ORF Within the Intron. One feature that is common to all of the other bacteriophage introns, but is lacking in β22, is a protein coding sequence inserted into one of the terminal loops. The longest ORF in the β22 intron (residues 935–1126) is only 64 codons, and it lacks a start codon in the context of a recognizable ribosome binding site. However, one indication that this may be a remnant of a functional ORF is the location of its UAG stop codon within the 3′ portion of the P8 helix (Fig. 4). We have previously noted that, although almost all of their coding regions are in terminal loops peripheral to the intron core, the 3′ ends of all the phage intron ORFs coincide with base-paired parts of the structure that are required for splicing (4, 24–26). In particular, the SPO1 intron ORF also ends in the 3′ portion of P8 (4).

A search of protein sequence databases with BLASTP (27) through the network service at the National Center for Biotechnology Information revealed similarities between the C-terminal portion of this ORF and several proteins, a surprising number of which (8 of 13) are from proteobacteria or their viruses (Fig. 5). In four of the eight bacterial examples, the similarities are also at their extreme C termini, suggesting that the sequence similarities might reflect a functional relationship. Interestingly, one of these is I-TevI, the *td* intron-encoded endonuclease of bacteriophage T4. All of these proteins bind DNA and, with the exception of I-TevI, are transcriptional regulators. The existence of an H-T-H DNA-binding motif has been noted in some of them (28). A log probability score was calculated for the β22 sequence by using a recently presented H-T-H motif model (28). The near perfect resulting score of 8.5×10^{-8} (0.0 is perfect) was statistically significant compared to 101 randomly shuffled controls ($P < 0.01$).

The existence of a vestigial ORF within an intron is not novel. An example exists in the bacteriophage T4 *nrdB* intron, whose ORF has undergone a deletion of 491 bp, either in the wild or very early in its propagation in the laboratory (3). We propose that a similar deletion in the loop of P8 of the β22 intron has left behind the coding sequence for the C-terminal portion of a functional protein.

The presence of a putative H-T-H domain at the extreme C terminus of I-TevI is unexpected, as (*i*) this protein has been characterized as binding to its DNA target primarily in

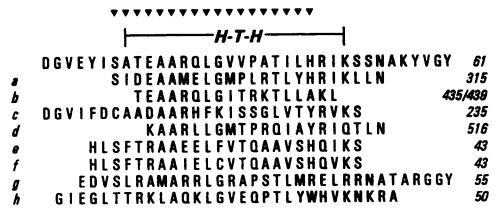

FIG. 5. Similarity of the β22 *thy* ORF fragment to bacterial protein sequences aligned by BLASTP. An 18 amino acid sequence in β22 that is similar to one that is identified in proteins with a helix–turn–helix (H-T-H) DNA-binding motif (28), is indicated by inverted triangles. The 20 amino acid segment that is typically identified as making up this motif is also shown (H-T-H). Numbers on the right indicate the position of the C-terminal amino acid of each aligned segment in its respective protein sequence. Proteins, their source, and GenBank accession numbers are as follows: (*a*) WtsA, *Erwinia stewartii* (L06093). (*b*) HydG, *Escherichia coli* (M28369) and *Salmonella typhimurium* (M64988), whose sequences are identical in this region. (*c*) I-TevI, bacteriophage T4 (M12742). (*d*) NifA, *Azotobacter vinelandii* (Y00554). (*e*) GcvA, *Escherichia coli* (X73413). (*f*) AmpR, *Pseudomonas aeruginosa* (X54719). (*g*) IS1086, *Alcaligenes eutrophus* (X58441). (*h*) TetR, Tn10, *Haemophilus parainfluenzae* (M15539).

the minor groove (29) and (*ii*) a truncated version of I-TevI that lacks approximately the distal third of the protein is fully active *in vitro* (30). Either the H-T-H domain of I-TevI plays a minor role in binding its substrate or I-TevI has an alternative, uncharacterized DNA target.

Origin of TS Introns in Bacteriophage. Very little is known about bacteriophage β22, and any proposed relationship to the T-even phages would be entirely speculative. Because β22 infects Gram-positive *Bacillus* species, and T4 and its relatives infect Gram-negative enteric bacteria, there would seem to be scant opportunity for genetic recombination in a common host. Indeed, the genetic isolation of β22 and T4 can be inferred from the evolutionary distance of their TS genes (Table 1). While this distance may be attributed, in part, to unusually rapid evolution of viral genes, the β22 amino acid sequence is more divergent from T4 than even the eukaryotic proteins.

In this context, the remarkable similarity of the β22 and T4 introns is strong evidence in favor of relatively recent acquisition. The mechanism for the spread of group I introns presumably involves their homing endonucleases, site-specific endonucleases that recognize and cleave the DNA sequence interrupted by the intron (31). Because its recognition sequence is split, an intron-containing gene is not sensitive to its own endonuclease, serving instead as a template for the repair of cleaved DNA. The result is unidirectional gene conversion to the intron-bearing allele, with coconversion of flanking markers (32). This phenomenon, called intron homing, is also thought to result in occasional cleavage at ectopic sites and, if there is sufficient similarity of flanking sequences to those in the exons, to lead to transposition of the intron during repair.

Thus, once established in an intron, an ORF encoding a homing endonuclease would tend to maintain its specificity for sequences surrounding the insertion site. The only way for a gene to permanently rid itself of such an intron in a single step would be through inexact deletion, as exact deletion simply recreates the target for future homing events. It is probably for this reason that these introns are found in strongly conserved positions of highly conserved genes (33).

It is, therefore, not surprising that genes for TS, one of the most conserved enzymes (15), should be hosts to an intron with a homing endonuclease. Nor should it be surprising that the intron is inserted in the coding region for one of the most highly conserved domains within this protein (15). It is not at all obvious, however, why the location of the β22 intron within this conserved domain is shifted exactly seven codons upstream from the location in T4 (Fig. 6). Moreover, the location of the β22 *thy* intron ORF is also shifted to P8, compared with P6 in T4 *td* (34). These differences imply that the intron and the ORF are separately mobile (35–37) and make it difficult to explain the movement of the intron as a simple homing event.

The T4 homing endonuclease, I-TevI, has several properties that distinguish it from most others. Whereas other enzymes have very specific target sequences and cleave very close to the intron insertion site, I-TevI cleaves the two strands of its target sequence 23 and 25 nt from the intron insertion site and is very tolerant of sequence changes in its recognition sequence (29, 32, 38). The close proximity of the cleavage site of the T4 enzyme and the insertion site of the β22 intron is striking (Fig. 6). Together with its unusual ability to cleave variations of its homing sequence, this suggests that the T4 ORF may be adapting to a relatively recent change in intron position. It appears that the group I introns, the ORFs that render them mobile, and the genes they inhabit have different evolutionary histories. TS introns of bacteriophages may provide a useful experimental system for reconstructing their mutual interactions.

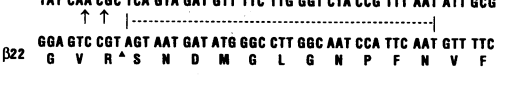

FIG. 6. Insertion sites of the group I introns in the TS genes of bacteriophages T4 and β22. Aligned amino acid sequences and their DNA codons are shown. Filled triangles indicate intron insertion sites. The segment of the T4 DNA that interacts with I-TevI is indicated by a dashed line (29). Arrows show the cleavage sites on the two strands of T4 DNA.

We thank Chip Lawrence for communicating results prior to publication and for analyzing the β22 sequence with an H-T-H model, Carole Keith for help with the figures, and Marlene Belfort, Steve LaBonne, and Robert Osuna for comments on the manuscript. This research was supported by National Institutes of Health Grants GM48804 (D.H.B.) and GM37746 (D.A.S.).

1. Shub, D. A., Coetzee, T., Hall, D. H. & Belfort, M. (1994) in *Molecular Biology of Bacteriophage T4*, ed. Karam, J. D. (Am. Soc. Microbiol., Washington, DC), pp. 186–192.
2. Quirk, S. M., Bell-Pedersen, D. & Belfort, M. (1989) *Cell* **56**, 455–465.
3. Eddy, S. R. & Gold, L. (1991) *Genes Dev.* **5**, 1032–1041.
4. Goodrich-Blair, H., Scarlato, V., Gott, J. M., Xu, M.-Q. & Shub, D. A. (1990) *Cell* **63**, 417–424.
5. Michel, F. & Westhof, E. (1990) *J. Mol. Biol.* **216**, 585–610.
6. Kuhsel, M. G., Strickland, R. & Palmer, J. D. (1990) *Science* **250**, 1570–1573.
7. Xu, M.-Q., Kathe, S. D., Goodrich-Blair, H., Nierzwicki-Bauer, S. A. & Shub, D. A. (1990) *Science* **250**, 1566–1570.
8. Reinhold-Hurek, B. & Shub, D. A. (1992) *Nature (London)* **357**, 173–176.
9. Ferat, J.-L. & Michel, F. (1993) *Nature (London)* **364**, 358–361.
10. Ferat, J.-L., LeGouar, M. & Michel, F. (1994) *C.R. Acad. Sci. Ser. 3* **317**, 141–148.
11. Biniszkiewicz, D., Cesnaviciene, E. & Shub, D. A. (1994) *EMBO J.* **13**, 4629–4635.
12. Hemphill, H. E. & Whiteley, H. R. (1975) *Bacteriol. Rev.* **39**, 257–315.
13. Gryczan, T. J. & Dubnau, D. (1982) *Gene* **20**, 459–469.
14. Promega Corp. (1991) *Promega Protocols and Applications Guide* (Promega, Madison, WI), 2nd Ed.
15. Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Montfort, W. R., Maley, G. F., Maley, F. & Stroud, R. M. (1990) *Proteins* **8**, 315–333.
16. Kenny, E., Atkinson, T. & Hartley, B. S. (1985) *Gene* **34**, 335–342.
17. Stroynowski, I. T. (1981) *J. Bacteriol.* **148**, 91–108.
18. Maley, G. F. & Maley, F. (1989) *Adv. Enzyme Regul.* **29**, 181–187.
19. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. & Cech, T. R. (1982) *Cell* **31**, 147.
20. Zaug, A. J., McEvoy, M. M. & Cech, T. R. (1993) *Biochemistry* **32**, 7946–7953.
21. Michel, F., Jacquier, A. & Dujon, B. (1982) *Biochimie* **64**, 867–881.
22. Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. (1982) *Nature (London)* **300**, 719–724.
23. Burke, J. M. (1988) *Gene* **73**, 273–294.
24. Shub, D. A., Xu, M.-Q., Gott, J. M., Zeeh, A. & Wilson, L. D. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**, 193–200.
25. Gott, J. M., Zeeh, A., Bell-Pederson, D., Ehrenman, K., Belfort, M. & Shub, D. A. (1988) *Genes Dev.* **2**, 1791–1799.
26. Michel, F., Jaeger, L., Westhof, E., Kuras, R., Tihy, F., Xu, M.-Q. & Shub, D. A. (1992) *Genes Dev.* **6**, 1373–1385.
27. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
28. Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F. & Wootton, J. C. (1993) *Science* **262**, 208–214.
29. Bryk, M., Quirk, S. M., Mueller, J. E., Liozos, N., Lawrence, C. & Belfort, M. (1993) *EMBO J.* **12**, 2141–2149.
30. Chu, F. K., Maley, F., Wang, A.-M., Pedersen-Lane, J. & Maley, G. (1991) *Nucleic Acids Res.* **19**, 6863–6869.
31. Dujon, B. (1989) *Gene* **82**, 91–114.
32. Bell-Pedersen, D., Quirk, S., Clyman, J. & Belfort, M. (1990) *Nucleic Acids Res.* **18**, 3763–3770.
33. Shub, D. A. & Goodrich-Blair, H. (1992) *Cell* **71**, 183–186.
34. 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.
35. Mota, E. M. & Collins, R. A. (1988) *Nature (London)* **332**, 654–656.
36. Perlman, P. S. & Butow, R. A. (1989) *Science* **246**, 1106–1109.
37. Dalgaard, J. Z., Garrett, R. A. & Belfort, M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5414–5417.
38. Chu, F. K., Maley, G., Pedersen-Lane, J., Wang, A.-M. & Maley, F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3574–3578.