PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 9/22, 15/55	A1	(11) International Publication Number: WO 94/04663
		(43) International Publication Date: 3 March 1994 (03.03.94)

(21) International Application Number: PCT/DK93/00264

(22) International Filing Date: 13 August 1993 (13.08.93)

(30) Priority data:
1015/92 14 August 1992 (14.08.92) DK

(71)(72) Applicants and Inventors: DALGAARD, Jacob, Zeuthen [DK/DK]; Pragtstjernevej 24, Emdrup, DK-2400 København NV (DK). GARRETT, Roger, Antony [GB/DK]; Toeltvej 26, DK-3490 Kvistgård (DK). KJEMS, Jørgen [DK/DK]; Solbakken 15, DK-8240 Risskov (DK).

(74) Agent: HOFMAN-BANG & BOUTARD A/S; Adelgade 15, DK-1304 Copenhagen K (DK).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: THERMOSTABLE SEQUENCE-SPECIFIC ENDONUCLEASES

(57) Abstract

Thermostable sequence-specific DNA endonucleases are encoded by archaeal type introns of stable RNA (ribosomal RNA or transfer RNA) or protein genes or are enzymatically active variants thereof in which one or more amino acid residues have been deleted, inserted or substituted by other amino acids. These endonucleases recognize relatively long sequences of about 20 base pairs and are very rare cutters, cleaving with a frequency of about 1:5 000 000. Thus, they are useful as endonuclease tools for gene analysis, such as genome mapping and detection of major rearrangements in large genomes, and for gene manipulation, such as cloning and chromosome targeting.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NE**	Niger
BE	Belgium	**GN**	Guinea	**NL**	Netherlands
BF	Burkina Faso	**GR**	Greece	**NO**	Norway
BG	Bulgaria	**HU**	Hungary	**NZ**	New Zealand
BJ	Benin	**IE**	Ireland	**PL**	Poland
BR	Brazil	**IT**	Italy	**PT**	Portugal
BY	Belarus	**JP**	Japan	**RO**	Romania
CA	Canada	**KP**	Democratic People's Republic of Korea	**RU**	Russian Federation
CF	Central African Republic			**SD**	Sudan
CG	Congo	**KR**	Republic of Korea	**SE**	Sweden
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovak Republic
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LU**	Luxembourg	**TD**	Chad
CS	Czechoslovakia	**LV**	Latvia	**TG**	Togo
CZ	Czech Republic	**MC**	Monaco	**UA**	Ukraine
DE	Germany	**MG**	Madagascar	**US**	United States of America
DK	Denmark	**ML**	Mali	**UZ**	Uzbekistan
ES	Spain	**MN**	Mongolia	**VN**	Viet Nam
FI	Finland				

THERMOSTABLE SEQUENCE-SPECIFIC ENDONUCLEASES

This invention relates to a novel group of thermostable sequence-specific DNA endonucleases isolated from hyper-thermophilic archaea. These endonucleases recognize relatively long sequences of about 20 base pairs and are very rare cutters, cleaving with a frequency of about 1 : 5 000 000.

STATE OF THE ART

Very rare cutters have been described before. A type of DNA endonucleases that are capable of cleaving one specific site in an entire genome has been found encoded by group I introns and very recently in a special kind of protein fusion, which is called an intervening protein sequence. All of these so-called "homing" endonucleases are believed to be involved in genetic mobility of the respective genetic elements which they are encoded by. This mobility is called intron "homing". It was first described for the Ω (Omega) intron of yeast mitochondria, and refers to the genetic element being copied from a specific position in a gene to the same position in another copy of the gene previously lacking the element. This type of genetic mobility takes place when two genomes, one with the genetic element and one without the genetic element are brought together in the same cytoplasm. The mobility event consists of separate steps. The first step is catalyzed by the "homing" endonuclease encoded by the element. The endonuclease is capable of selectively cleaving the gene that lacks the element keeping the rest of the genome unharmed. The second step in the mobility event is the cellular enzyme's repair of the cleaved gene. This repair is a recombination event with the intact copy of the gene. Since the intact copy of the gene contains the element, the repair copies the element to the same position in the new genome. Several "homing" endonucle-

ases have been characterized from group I introns and one from an intervening protein sequence. The site-specificity of these enzymes has only been investigated for a few, but they seem to correlate with the size of the genome they are encoded by. Moreover, the "homing" endonucleases encoded by group I introns in bacteriophages, which have small genomes, are not as sequence-specific as the "homing" endonucleases encoded by group I introns in genomes or mitochondrial genomes of eukaryotes. A few of these enzymes are available commercially at the moment and are used mainly for genome mapping.

Archaea constitutes a separate primary biological kingdom consisting of hyperthermophiles, extreme halophiles and methanogens. So far, archaeal introns have been detected in both hyperthermophilic and extreme halophilic archaea exclusively in stable rRNA and tRNA genes. Only those found in the rRNA genes exhibit ORF's (Kjems & Garrett, Nature, 1985). The archaeal introns differ from group I and other introns in their mode of splicing (Kjems & Garrett, Cell, 1988). The rRNA introns are excised by a protein cleavage enzyme at a structural motif consisting of a "bulge-helix-bulge" which is illustrated in Fig. 1(c). The enzyme cuts in the "bulges" and subsequently the ends of the exons, and the ends of the intron are ligated (Kjems and Garrett, Cell, 1988). This type of intron has, so far, only been detected in archaea.

Archaeal introns are found in two forms. One is expressed in tRNA transcripts generally inserted in the anticodon loop (Kjems et al., 1989: Can. J. Microbiol.), but occasionally in other parts of the tRNA. This intron form which is illustrated in Fig. 1(b) generally consists of a relatively small stem loop structure at the RNA level. This type of intron has also been detected at the rRNA level in the hyperthermophilic archaeon Staphylothermus marinus, again as a single stem loop structure (Kjems & Garrett, 1991:

PNAS). The other intron form which is illustrated in Fig. 1(a) contains this truncated RNA structure, but in addition it contains an ORF which is generally about 200 codons long (Kjems and Garrett, 1985: Nature). Possibly the first form of intron described above (Fig. 1(b)) constitutes a core into which the open reading frame can insert at a DNA or RNA level.

DESCRIPTION OF THE INVENTION

The present invention comprises a thermostable sequence-specific DNA endonuclease which has the amino acid sequence encoded by an archaeal type intron of a stable RNA (ribosomal RNA or transfer RNA) or protein gene or is an enzymatically active variant thereof in which one or more amino acid residues have been deleted, inserted or substituted by other amino acids.

This type of enzymatic activity has not previously been assigned to proteins encoded by archaeal introns, and the thermostable endonucleases of the invention therefore constitute a new group of "homing" DNA endonucleases that is encoded in a new kind of mobile intron. The advantages of the new type to the older type is that they are very stable, due to the high living temperature of their host organisms, and are active over a wide range of temperature. This renders the enzymes suitable tools for molecular biology. A further important advantage is that they are easy to purify when expressed in E. coli, since E. coli's proteins can easily be denatured by heating and removed.

The sequence-specific endonucleases of the invention i.a. have the following practical uses:

1) Genome mapping. Owing to the rarity of their recognition sites, the enzymes are ideal for mapping large genomes, using standard techniques. If no recognition site occurs in a genome - as for example in the E. coli genome, the

recognition site can be inserted artificially. The genome can then be linearized at the single position. Using the fact that the enzyme creates two assymmetric 4 base pair 3'-overhangs, one end of the genome can then be labelled specifically using a hot linker and T4 ligase. Since only one end of the genome is labelled, very large regions can then be mapped by partially digesting with 8 base pair cutters and subsequently separating the fragments by pulse field electrophoresis.

2) Detection of major rearrangements in large genomes. These could be detected by digesting DNA and subjecting it to pulse field gel electrophoresis assuming that at least one cutting site occurs in the large fragments. The enzyme could then be very useful for diagnosing genetic diseases or for genetic fingerprinting.

3) Cloning. The recognition sites can be inserted into vectors for bacteria, archaea and eukaryotes, producing a unique site for linearization of the vector or excision of inserted fragments. This would be very useful when unknown sequences are to be cloned and sequenced, and would make it much easier to work with large vectors such as phage λ and SV40.

4) Chromosome targeting. If there is only one recognition site in a genome (which could be artificially inserted), it is straightforward to move genes to that site in the genome. Transformation or transfection with a vector carrying the endonuclease gene would lead to cleavage of the site. The cleavage would then initiate a recombination event. If the DNA sequence which one wishes to target into the genome is flanked by sequences which are identical to the endonuclease recognition site, the whole sequence would be recombined into the genome. By using an appropriate selection marker, this could be developed into a transformation system for archaea and eukaryotes.

Thus, the invention also comprises the use of a DNA endonuclease according to the invention as an endonuclease tool for gene analysis and manipulation.

The family of thermostable sequence-specific endonucleases until now comprises I-Dmo I, I-Por I and I-Por II. I-Dmo I is encoded by the archaeal type intron present in the 23S rRNA gene of the hyperthermophilic archaeon Desulfurococcus mobilis. This intron has the nucleotide sequence stated in SEQ ID No. 1 in this description; and when the 23S rRNA gene is transcribed into pre-23S rRNA, after which the intron is excised, it circularizes by ligation of the ends, and then the enzyme, I-Dmo I, is expressed by translation from nucleotide 59 through nucleotide 622 to nucleotide 21 (stop codon) thus having the amino acid sequence stated in SEQ ID No. 3. I-Dmo I has been proven a sequence-specific endonuclease in the following. However, as also shown in the following, the truncated peptide stated in SEQ ID No. 2 has the same endonuclease activity, and hence it may be concluded that at least all the peptides, in which any number of the last 6 amino acids from the 3'-end of SEQ ID No. 3 has been deleted, have the endonuclease activity of I-Dmo I.
I-Por I and I-Por II are encoded by the two archaeal type introns present in the 23S rRNA gene of the hyperthermophilic archaeon Pyrobaculum organotrophum. I-Por I is encoded by intron 1 having the nucleotide sequence stated in SEQ ID No. 4 and thus has the amino acid sequence stated in SEQ ID No. 5; and I-Por II is encoded by intron 2 having the nucleotide sequence stated in SEQ ID No. 6 and thus has the amino acid sequence stated in SEQ ID No. 7. I-Por I has also been proven a sequence-specific endonuclease in the following, and I-Por II has not yet been fully investigated, but is, due to sequence similarity, believed to be another sequence-specific endonuclease. The cleavage frequency of these enzymes is not known <u>in vitro</u>, but they cleave only once in the respective genomes of the intron$^-$ strains <u>in vivo</u>. Sequences of the genes and enzymes are given below.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1: Putative secondary structures bordering the exon-intron boundaries of (a) the rRNA intron of Dc. mobilis and (b) an intron in $tRNA^{Met}$ of Dc. mobilis, and (c) a scheme of the essential structure involved in endonuclease recognition at the exon-intron boundary (arrows indicate cleavage sites).

Fig. 2: Secondary structure of the circular 23S rRNA intron of Dc. mobilis formed after excision from pre-23S rRNA.

Fig. 3: Comparison of PCR products from the domain IV-V region of the 23S rRNA-encoding genes of (1) Dc. mucosus and (2) Pb. organotrophum run on a 0.75% agarose gel alongside (M) a restriction digest of phage λ DNA with BstEII. Chromosomal DNA was extracted from frozen cell mass of the hyperthermophiles (Kjems and Garrett, 1987) and the selected DNA region was amplified by the PCR method (Sambrook et al., 1989) using two deoxyribonucleotide primers; one (5'-CCTGACTGTTTAATAAAA-3') was complementary to the strand encoding rRNA at a conserved sequence near the 5'-end of domain IV and the other (5'-CCCGTTCCTCTCGTACT-3') was complementary to a conserved sequence within domain VI of 23S rRNA. Primers were annealed at 47 °C and extended at 72 °C. Taq DNA polymerase ("Stratagene") and a thermocycler ("Hybaid") were used.

Fig. 4: DNA sequences of the two inserts (introns) in the 23S rRNA-encoding gene of Pb. organotrophum. Short sequences of the flanking 23S rDNA (exons) are boxed. The identities of the domains where the RNA introns are located are indicated on the right. Amino acid sequences corresponding to the putative open reading frames are given below the nucleotide sequences. (* - stop codon). The DNA sequencing procedure was as follows: PCR products were purified in a low melting agarose gel. Possible recessed ends were filled

in by the Klenow enzyme, and the product was cloned into pUC19 (Yanish-Perron et al., 1985) and mapped with restriction enzymes. Overlapping fragments were inserted into M13mp18 and M13mp19 vectors, and both strands were sequenced using the dideoxynucleotide procedure (Sanger et al., 1977). To avoid errors three different PCR products were sequenced. Sequences are entered in the GenBank/EMBL database with the Accession number M 86622.

Fig. 5: Southern blotting analysis of (A) intron 1 (B) the 23S rRNA-encoding gene + intron 2 and (C) intron 2. 2 µg chromosomal DNA from Pb. organotrophum was digested with each of the following restriction enzymes: 1. SacII, 2. SacI, 3. PstI, 4. BamHI and 5. HindIII and electrophoresed on a 0.75% agarose gel. The gel was stained with ethidium bromide and transferred to a "Hybond N" filter (Amersham) using the capillary transfer method (Sambrook et al., 1989). The positions of the size markers derived from phage λ DNA digested with BstEII are shown on the right. Radioactively labelled probes specific for each of the introns and the 23S rRNA gene were prepared from purified restriction fragments using [α-^{32}P]dATP (3000 mCi mmol^{-1}) and a labelling kit for randomly primed DNA (Boehringer, Mannheim). Probes were denatured at 95 °C for 10 min and absorbed onto a filter. Hybridization was performed under conditions of high stringency.

A scheme of the PCR fragment is given showing the two introns and the locations of two internal SacII sites. The hydridization sites for the restriction fragment probes for introns 1 (A) and 2 (C) and 23S rRNA-encoding gene + intron 2 (B) are also indicated.

Fig. 6: Autoradiograms of sequencing gels demonstrating that intron-excision and exon ligation occur for introns 1 (A to C) and 2 (D to F). The flanking gels show DNA sequences across the 5' exon-intron junction (A, D) and the 3' intron-

exon junction (C, F). The central gels (B, E) show RNA sequences across the exon-exon junctions. They were determined by extension from ^{32}P-5'-end labelled primers hybridized under high stringency conditions close to the 5'-termini of the linear introns using reverse transcriptase (Life Sciences, Florida) and the dideoxynucleotide procedure. Primers were annealed at 50 °C and the reverse transcriptase reaction was performed at 48 °C.

Arrows indicate the intron-exon junctions (DNA) and the exon-exon junctions (RNA) and the bordering sequences are given for each gel. Compressions, indicated by asterisks, occurred in one DNA (A) and in one RNA sequence (E), probably due to the presence of stable secondary structure; the DNA sequence was verified by sequencing the complementary strand.

Fig. 7: Northern analysis of the total RNA extracted from Pb. organotrophum cells to examine the presence of 23S rRNA and RNA introns. The hybridization pattern obtained with ^{32}P-5'-end labelled probes specific for (A) 23S rRNA, (B) intron 1 and (C) intron 2 are shown. The 23S rRNA and intron 1 were probed with restriction fragments, while intron 2 was probed with a deoxyoligonucleotide; their respective hybridization sites are indicated in the scheme. The locations of co-electrophoresed 23S rRNA (2900nt) and a transcript of domain II of E. coli 23S rRNA (720nt) are indicated. 0 denotes the origin of electrophoresis.

Total RNA was extracted from 0.3 g frozen cell mass by a single-step procedure using acidified guanidinium thiocyanate, phenol and chloroform (Chomczynski and Sacchi, 1987) and stored in 0.5% sodium dodecylsulfate (5 mg/ml) at -80 °C. 5 µg total RNA from Pb. organotrophum was fractionated on an agarose gel containing 1% formamide (Sambrook et al., 1989). The gel was stained with ethidium

bromide, and the RNA was transferred to a "Hybond N" filter (Amersham).

Fig. 8: Autoradiograms of RNA sequencing gels demonstrating the circularity of the introns. Sequences were generated using the reverse transcriptase procedure as described in the legend to Fig. 6. The ligated junctions are indicated by arrows, and the bordering sequences corresponding to the rRNA template are given. No termination of reverse transcription was observed at the junctions. RNA was isolated as described in the legend to Fig. 5.

Fig. 9: (A) Schematic secondary structure of the domain IV-V region of the 23S rRNA of Pb. organotrophum. The structure is derived from that of the 23S rRNA of other hyperthermophiles. The cleavage sites of the two newly discovered introns and those previously reported for Dc. mobilis (D.m.) and St. marinus (S.m.) are indicated by arrows. (B) and (C) The nucleotide sequence and putative secondary structures surrounding the exon-intron junctions (including the two boxed areas in (A)) for introns 1 and 2, respectively. Arrows denoting the cleavage sites and the sizes of the omitted intron sequences are given.

Fig. 10: Alignment of the amino acid sequences of the putative intron-encoded proteins from the two archaea showing the similarities between the repeated decapeptide motif. These sequences are also similar to consensus sequences within proteins encoded by eukaryotic group I introns. The GCG package (Devereux et al., 1984) was used for analysing the protein and DNA sequences, and for comparing them to sequences in the EMBL/GenBank databases.

Fig. 11: Comparison of β-sheet distributions of the three putative proteins encoded by the introns from Dc. mobilis (D.m.) and Pb. organotrophum (P.o. 1 and 2) using the

structure prediction algorithm of Garnier et al. (1978). The repeated decapeptides are boxed.

Fig. 12: Products of the Dc. mobilis intron. (A) RNA-products of reading frames. Translational reading frames are represented as solid bars. CJ, cyclization junction. (B) Constructs containing the intron ORF correspond to the circular intron (Dmo-c), the precursor (Dmo-p), and the linear intron (Dmo-1) (constructs 1 and 2). Arrows labelled 1 and 2 represent primers (see Materials and Methods). (C) Protein sequences. Amino acid residues of the intron ORF products of the constructs in B are represented. Asterisks indicate stop codons. (D) Translation products. The in vitro products expressed from pDmo-p (lane 2), pDmo-1 construct 1 (lane 3), and pDmo-c (lane 4) were labelled with [^{35}S]methionine and electrophoresed in an SDS/12% polyacrylamide gel, alongside brome mosaic virus translation products (lane 1).

Fig. 13: Endonucleolytic cleavage. (A) I-Dmo I cleavage activity. The protein products from constructs shown in Fig. 12 were incubated with the intronless substrate (3.4 kb), which had been cleaved from a 7.2 kb M13mp18 vector. Lanes 2-4 correspond to those in Fig. 12D; lane 1 contains unprogrammed extract. Results with both forms of I-DmoII were identical (data not shown). The samples were electrophoresed in a 0.75% agarose gel and stained with ethidium bromide. (B) Thermostability of I-Dmo Ic. Enzyme synthesized from pDmo-c was incubated at the indicated temperatures with endlabelled substrate. Products were separated in a 1% agarose gel, which was fixed in 10% acetic acid and dried before autoradiography and quantitation with a direct radioactivity detector (Betagen, Waltham, MA). The material in the well at 85 °C is denatured vector. (C) Activity plot. Data points were derived from B.

Fig. 14: Cleavage characteristics. (A) Cleavage-site mapping. Cleavage products with I-Dmo Ic were separated in a 6% polyacrylamide/urea gel alongside DNA-sequencing reaction products (A, G, C, and T lanes) (Wenzlau et al. 1989: Cell). The cleaved substrate (+ lanes) was electrophoresed beside uncleaved substrate (- lanes) and was also added to A (A + lane) or T(T + lane) reaction mixtures to determine the precise cleavage site, as indicated by a line on the sequence, adjacent to the intron insertion site (arrowhead). The cleavages indicated on the sequence below were identical with I-Dmo II. (B) Properties of intron-encoded endonucleases (15).

Fig. 15: Phylogenetic distribution of intron-like endonucleases. The tree (Woese et al. 1990:PNAS) shows endonuclease-bearing phyla in the three primary kingdoms, the Bacteria, Archaea, and Eukarya. Information in boxes is arranged according to genome type, endonuclease gene location, and endonuclease type. Genome; Phag, phage; Cell, cellular; Mito, mitochondrial; Nucl, nuclear; Chlo, chloroplast. Location; grpI, with a group I intron; free, freestanding, between genes; arch, within an archeael intron; prot, in a protein fusion. Endonuclease type: bold E, LAGLI-DADG family; open E, other types, which contain at least one different endonuclease family, with the G1Y-Y1G motif (see text). A bold E does not imply that the LAGLI-DADG family is represented exclusively in a particular phylum. The species referred to are as follows: 1, phage T4; 2, phage T4; 3, Mycobacterium tuberculosis (*, endonuclease activity implied but not proven); 4, D. work); 5, Thermococcus literalis; 6, Saccharomyces cerevisiae; 7, S. cerevisiae; 8, S. cerevisiae; 9, S. cerevisiae; 10, Chlamydomonas smithii; 11, Chlamydomonas eugamotes and Chlamydomonas reinhardtii; 12, Physarum polycephalum.

Fig. 16: A graph of the cleavage activity of endonuclease I-Dmo I as a function of temperature.

Fig. 17: A graph of the cleavage activity of endonuclease I-Dmo I as a function of pH.

Fig. 18: A graph of the cleavage activity of endonuclease I-Dmo I in buffers containing varying concentrations of NH_4Cl.

Fig. 19: A graph of the cleavage activity of endonuclease I-Dmo I in buffers containing varying concentrations of $MgCl_2$.

Fig. 20: Expression and purification of I-Dmo I.
A. Expression of I-Dmo I in E. coli BL21(DE3) F' $lacI^q$. Total cell extracts were run on a polyacrylamide/SDS gel.

Lane	1	uninduced
	2	IPTG-induced for ½ h
	3	" 1 h
	4	" 2 h
	5	" 3 h
	6	" 4 h
	7	" 16 h
	8	Molecular marker

B. Precipitation with ammonium sulfate and purification on heparin column.

Lane	1	Molecular marker
	2	Total lysate of induced cells
	3	Debris of lysed cells
	4	45 % to 75 % ammonium sulfate cut
	5	20 % to 45 % ammonium sulfate cut
	6	Run-through from heparin coloumn
	7 - 18	Fractions of heparin coloumn

C. Purification on phenyl-"Sepharose" column.

Lane	1	Molecular marker
	2	Protein eluted from phenyl-"Sepharose"

Fig. 21: Cleavage of a DNA substrate from the intron$^-$ Pb. islandicum by I-Por I.

A: Gene map showing the location of a 23S rDNA fragment on a Cla I fragment (C-C) isolated from M13mp18. Arrows indicate the conserved sequence positions where the rRNA introns of Pyrobaculum organotrophum (Pb. o.) (intron 1) and Desulfurococcus mobilis (Dc. m.) lie.
B: Digestion of the 7600 bp fragment drawn in A by I-Por I and I-Dmo I. The fragments at approximately 5200 (±100) bp correspond to the left hand side of the fragment in A; the fragments corresponding to the right hand side are very weak in the Figure, indicated by small arrows on the left.

Fig. 22: I-Por I cleavage site mapping. 23S rDNA from Pb. islandicum was incubated with (+ lanes) and without (- lanes) I-Por I at 70 °C in 50 mM Hepes-KOH pH 7.0, 100 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, 10 % glycerol, 0.1 % "Triton X-100" for 1 h and electrophoresed on a 5 % polyacrylamide/urea gel together with sequencing reaction mixtures (A, G, C and T). Cleavage sites are indicated by arrows. Arrows far left and right show direction of electrophoreses. Compression of three G's is seen next to the cleavage site.

DETAILED DESCRIPTION OF THE INVENTION

Archaeal type introns located in tRNA- and 23S rRNA-encoding genes have been detected in hyperthermophilic and extreme halophilic archaea. They are excised from rRNA transcripts by a protein enzyme which recognizes a "bulge-helix-bulge" motif (Fig. 1(c)) at the exon-intron junction and cuts in the bulges (Thompson and Daniels, 1990; Kjems and Garrett, 1988; 1991); a similar motif, which may also be cleaved by this enzyme, occurs in the processing stems of the large rRNAs (reviewed by Garrett et al., 1991).

The 23S rRNA intron of the archaeon Desulfuroccus mobilis circularizes after excision from pre-23S rRNA and forms a normal 5'-3'-phosphodiester bond at the ligation junction. This circular intron (622 nt) contains a putative ORF which

crosses the ligation junction and covers most of the intron (Kjems and Garrett, 1985; 1988). The intron generates a complex secondary structure which has now been determined experimentally (Fig. 2). The other characterized archaeal introns from tRNA-encoding genes and other rRNA-encoding genes (Kjems and Garrett, 1991) are small (<110 nt).

Detection of two protein-coding introns in the 23S rRNA-encoding gene of the hyperthermophilic archaeon Pyrobaculum organotrophum

In the following experiments, we used Polymerase Chain Reaction (PCR) to screen hyperthermophilic archaea (Stetter et al., 1990; Kjems et al., 1992) for the presence of larger introns in the DNA region encoding the functionally important domains IV and V of 23S rRNA where most rRNA introns are located in archaea and in eukaryotic nuclei (reviewed by Garrett et al., 1991).

(a) The single 23S rRNA-encoding gene of Pb. organotrophum contains two introns

PCR products were obtained from DNA samples of the hyperthermophiles Archaeoglobus fulgidus, Thermoplasma volcanium and a Pyrodictium-like isolate (AV2) (Stetter et al., 1990) using deoxyoligonucleotide primers bordering the domain IV-V region of the 23S rRNA gene; they were all the same size as the control product from the hyperthermophile Dc. mucosus (Fig. 3) which contains no introns (Kjems and Garrett, 1985). None of these products was subcloned and sequenced. In contrast, the DNA sample of Pb. organotrophum (Huber et al., 1987) yielded a larger PCR product after an increased extension time compatible with the presence of one or more inserts (Fig. 3). The sequence of this DNA product was determined and aligned with 23S rRNA-encoding gene sequences from other hyperthermophiles. This revealed that it contained two additional sequences which could constitute

introns (Fig. 4); both exhibited ORF's which, in contrast to that of Dc. mobilis 23S rRNA (Kjems and Garrett, 1985), do not cross the insert junctions (Fig. 4). A Southern blotting analysis was made to determine the number of 23 rRNA-encoding genes present in the organism. Five restriction enzymes BamH1, HindIII, PstI, SacI and SacII were used in the analysis one of which, SacII, cut twice within the amplified rDNA fragment (Fig. 5). Size markers prepared from a digest of phage λ DNA with BstEII were co-electrophoresed on an agarose gel with the restriction digests. After blotting onto a "Hybond N" filter, ^{32}P-5'-end labelled restriction fragments were used to probe inserts 1 and 2 and the 23S rRNA-encoding gene + insert 2 (Fig. 5).

The hybridization results (Fig. 5) show that for the four restriction enzymes which did not cut within the PCR product (tracks 2 to 5), single bands were observed in each track when the probe for the 23S RNA-encoding gene + insert 2 was hybridized (Fig. 5B). This suggests that the genome of Pb. organotrophum exhibits a single copy of the 23S rRNA-encoding gene which always contains insert 2 implying that there is not a mixed population of intron$^+$ and intron$^-$ strains. The probes for the individual inserts hybridized exclusively to the same bands (Fig. 5A, C) which renders it very likely that both inserts are confined to a single 23S rRNA-encoding gene. These two inferences were reinforced for the SacII digest (track 1) where the DNA was cut between the inserts at the start of insert 2 (see scheme in Fig. 5). Probes specific for the two inserts hybridized to different single bands while the probe for the 23S RNA-encoding gene + insert 2 hybridized to both bands. We concluded, therefore, that both inserts are associated exclusively with a single 23S rRNA-encoding gene.

Next, we investigated whether the inserts were excised during maturation of the rRNA and could, therefore, constitute introns. Deoxyoligonucleotide primers were

hybridized to mature 23S rRNA, 3' to inserts 1 (5'-GAAUCCUGGCCACUGGCGGUACG-3') and 2 (5'-AGUACGAGAGGAACGGG-3'). The sequences at the 3'-boundaries of the inserts were then examined by the dideoxynucleotide sequencing method using reverse transcriptase. The results are shown for both inserts in Fig. 6 alongside the gene sequence showing the corresponding boundaries at the DNA level. Clearly, both inserts are absent from the mature 23S rRNA. Furthermore, the lack of reverse transcriptase termination at the putative exon ligation sites indicates that a normal 5'-3'-phosphodiester bond formed during maturation of the 23S rRNA in an efficient ligation reaction. Therefore, we concluded that both inserts are introns.

(b) Both introns are present in the cell in a circular form

The fate of the excised RNA introns in the cell was examined by a Northern blotting analysis. 5 µg total RNA from Pb. organotrophum was denatured in 17.5% formaldehyde, electrophoresed in an agarose gel containing 2.2 M formamide together with size markers and then blotted onto a "Hybond N" filter. Radioactively end-labelled restriction fragments and a deoxyoligonucleotide (5'-CCAGCTAGATACTCCAAATC-3') were used to probe the 23S rRNA and introns (see scheme in Fig. 7). The results revealed that both free introns were present in the cellular extracts and that they co-migrated with a 720 nt size marker (Fig. 7). The extended band in Fig. 7A probably reflects degradation of the 23S rRNA sample, but we cannot eliminate the possibility that other introns are present. Further analysis of RNA extracts on composite polyacrylamide gels revealed weak bands relative to those of 7S and 5S RNAs suggesting that, in vivo, both introns are of limited stability post-splicing (data not shown).

The possibility that the RNA intron moieties were circular, like that of Dc. mobilis, was investigated by sequencing the

RNA introns using reverse transcriptase primed by deoxyoligonucleotides hybridized close to the 5'-ends of introns 1 (5'-GGAATATATCCACTATACAATTGC-3') and 2 (see above). For both introns the reverse transcript extended from the 5'-end into the 3'-end of the linear intron sequence indicating that circularization had, indeed, ocurred (Fig. 8).

(c) The splicing sites

The locations of introns 1 and 2 in the latest secondary structural model of the domain IV-V region of the 23S rRNA from Pb. organotrophum are shown in Fig. 9 together with the positions of the other known archaeal rRNA introns. Many of the eukaryotic group I and group II introns found in pre-23S-like rRNAs also occur within this functionally important region of the rRNA (reviewed by Garrett et al., 1991).

There is strong evidence that the archaeal cleavage enzyme recognises a "bulge-helix-bulge" motif at the exon-intron junctions (Kjems et al., 1989) which constitutes two three-base bulges on opposite DNA strands separated by 4 bp and bordered by at least one stable helix (Thompson and Daniels, 1990; Kjems and Garrett, 1991). These motifs can be discerned for both introns in the pre-rRNA of Pb. organotrophum (Fig. 9B, C). The presence of the motif requires, as for the other archaeal rRNA introns (Kjems and Garrett, 1991), that a local rearrangement of the exon structures occurs post-splicing in order to generate the mature rRNA structure.

The location of intron 1 is of particular interest because it lies in the boxed stem-loop structure in fig. 9A which constitutes the binding site of the primary binding protein EL2 which has an important role in ribosomal assembly (Egebjerg et al., 1991). Formation of the "bulge-helix-bulge" motif in the pre-23S rRNA (Fig. 9B) will preclude binding of the hyperthermophile L2 protein and, therefore,

of 50S subunit assembly, until the splicing reaction is completed. Intron 2 is located at an RNA site that has been implicated in A-site binding of tRNA; thus, the 50S subunit can neither assemble completely nor function while the two introns are present.

(d) Functions of the putative proteins encoded by the archaeal rRNA introns

Both introns of Pb. organotrophum, and that of Dc. mobilis (Kjems and Garrett, 1985), contain ORFs and circumstantial evidence suggests that they are expressed. First, only one reading frame is possible for each intron and, second, dot matrix analyses of the amino acid sequences (not shown) reveal a decapeptide sequence occurring twice in the ORF of the Dc. mobilis intron and intron 2 of Pb. organotrophum and once in intron 1 of Pb. organotrophum (Fig. 10). Moreover, despite little additional common sequence, the putative proteins resemble each other in amino acid composition, net charge and in their predicted secondary structures. Strikingly similar are the predicted β-sheet distributions in the three putative proteins (Fig. 11). These results (Fig. 11) also provide evidence for a repeated secondary structure within the two halves of each putative protein when they are aligned at the sequence motifs.

Circumstantial evidence suggests that the archaeal introns are mobile. Thus, they are located at different positions in the rRNA genes (Kjems and Garrett, 1991). Moreover, they are absent from Dc. mucosus which is closely related to Dc. mobilis (Kjems and Garrett, 1985) and from Pb. islandicum which is closely related to Pb. organotrophum. Furthermore, the presence of common sequences at the exon-exon (GTA*AG) and intron-intron junctions (GAG*AGGGC) (where asterisks denote the ligation points) shared by the splicing sites of the Dc. mobilis intron and intron 1 of Pb. organotrophum could also reflect common recognition sites. Nevertheless,

the true function(s) of putative proteins encoded by archaeal rRNA introns remains to be established.

(e) Conclusions

(1) Two introns occur within the single 23S rRNA-encoding gene of Pb. organotrophum. Both RNA products circularize after excision from the 23S rRNA and are stable in the cell.

(2) The putative proteins encoded by the two introns exhibit repeated secondary structures in their two halves. Moreover, one half of intron 1 and both halves of intron 2 contain a common decapeptide sequence which is shared by the putative protein encoded by both the archaeal intron of Dc. mobilis and many group I introns. This raises the possibility that the ORF's of the two classes of introns have a common evolutionary origin.

Characterization of Endonuclease Activity of I-Dmo I

The endonucleases encoded by group I introns in both eukaryotes and bacteriophages cleave intronless alleles and thereby confer mobility on the introns, resulting in their idiosyncratic distribution. Similarly, the introns in the above archaeal genera are "optional", with some species such as Dc. mucosus and Pb. islandicum lacking one or more of the introns, suggesting that these introns are also mobile. In the following study we show that the ORF of Dc. mobilis does indeed encode an endonuclease and that the active form of the protein can be expressed from the linear or cyclized intron, but not from the pre-rRNA. Taken together, these findings support the intrusive nature of endonuclease-encoding elements and provoke debate as to whether the ORFs were acquired by trans-kingdom transfer or whether they are primitive, having originated before the divergence of the three primary kingdoms.

a) Materials and Methods

Plasmid Constructs. To construct pDmo-c, a reverse transcript was prepared from Dc. mobilis RNA with primer 1 (5'-CCTACCCCATCGAAC-3'), which anneals to the 5' end of the intron (see Fig. 12B). The cDNA was amplified by PCR with primer 1 and primer 2 (5'-AGGAAACATTAATGCATAATAATGAGA-3'), to generate a fragment of 629 bp that contains the entire ORF with its 3' and traversing the cyclization junction. The fragment was cloned into pUT718, a pUC18 derivative containing a T7 promoter, to generate pDmo-c. To construct pDmo-p, the Nci I-BamHI fragment of pDmo-c was replaced with the Nci I-BamHI fragment of a λ clone containing the entire gene, which restores the 3' end of the intron and a 5' portion of exon II (9). Two different pDmo-1 constructs were made by introducing the indicated mutations (Fig. 12 B - bold type, UAA and CUC) at the first codon in exon II. These changes were designed to create a stop codon at the intron-exon junction (construct 1) or to generate an Ecl136II site for cleavage precisely at the intron-exon boundary, yielding run-off transcripts that corresponded to the 3' end of the intron (construct 2).

In Vitro Synthesis of I-Dmo I. To synthesize I-Dmo I transcripts, linearized plasmid (3 µg) was resuspended in 120 µl of 40 mM Tris.HCl, pH 8.0/6 mM $MgCl_2$/1 mM spermidine/5 mM dithiothreitol/2mM rNTPs and incubated for 2 h at 37 °C with 50 units of T7 RNA polymerase. The transcripts were treated with phenol and chloroform, precipitated with ethanol, dried, and suspended in 20 µl of water. The in vitro translation mixtures contained 35 µl of rabbit reticulocyte lysate (Promega), 5 µl of mRNA, 1 µl of 1 mM amino acid mixture (without methionine), 5 µl of [^{35}S]methionine (1200 Ci/mmol; 1 Ci = 37 GBq) at 10 mCi/ml, and 4 µl of water. After incubation at 30 °C for 60 min., the translation products were separated in an SDS/12% polyacrylamide gel.

Cleavage Assay. Target DNA (1 µg) was incubated with 1 µl of translation mix in a total volume of 30 µl containing 25 mM tris.HOAc (pH 8.65), 10 mM $Mg(OAc)_2$, 10 mM NH_4OAc, and 10% (vol/vol) glycerol at 75 °C for 30 min. The reactions were stopped on ice with 10 µl of dye mix (10% Ficoll/0.01% bromophenyl blue/0.01% toluidine blue/0.3% SDS/50 mM EDTA). The substrate either was unlabelled or was end-labelled with [^{32}P]dNTPs and T4 DNA polymerase. Cleavage products were separated by electrophoresis in agarose gels and visualized as described in figure legends.

b) Experiments

Expression of the Dc. mobilis rRNA Intron ORF. The absence of demonstrable endonuclease activity in crude Dc. mobilis extracts prompted a test of different ORF-encoding templates in an in vitro wheat germ translation system. To optimize the chances of detecting activity, templates corresponding to the intron-exon II reading frame of the pre-rRNA (Dmo-p), the linear intron (Dmo-l), and the circular intron (Dmo-c) were used (Fig. 12A and B). To mimic translation from the linear intron, Dmo-I was prepared in two forms. In construct 1, a stop codon was inserted as the first codon of the downstream exon, to generate a translation product that ended at the 3' end of the intron. In construct 2 an Ecl136II site was created such that cleavage with Ecl136II would result in a T7 transcript with a 3' end identical to that of the linear intron. In vitro translation of the different constructs produced peptides of the appropriate sizes of 25.8, 22.0 and 22.5 kDa for Dmo-p, Dmo-I (both forms), and Dmo-c, respectively (Fig. 12B-D).

The Intron ORF Encodes a "Homing" Endonuclease. The in vitro translation products were tested in cleavage assays with the intron-negative 23S rRNA gene from Dc. mucosus as the DNA substrate. Translation products of the linear (both forms) and circular intron templates showed very high levels of

cleavage activity, whereas the pre-rRNA product was inactive (Fig. 13A). Although cyclization is not necessary for production of the endonuclease, the ability to encode active enzyme is consistent with the abundance of circular form in the cell, as opposed to pre-rRNA and linear intron, which are barely detectable. By convention (Dujon et al. 1989: Gene), the endonuclease is named I-Dmo I, reflecting the genus and species of the host organism, with I-Dmo Ic and I-Dmo II corresponding to products of the circular and linear intron forms, respectively.

Properties of I-Dmo I Endonuclease. Consistent with the hyperthermophile nature of Dc. mobilis, which lives optimally at ~85 °C, I-Dmo Ic is thermostable (Fig. 13B and C). The enzyme is virtually inactive at temperatures up to 37 °C, reaches peak activity in the 65-75 °C range, and loses activity at higher temperatures. It is unclear whether loss of activity above 75 °C reflects denaturation of the enzyme and/or the substrate.

The single cleavage site, which was observed exclusively for the intron-negative and not the intron-positive allele (data not shown), was mapped by bidirectional primer extension analysis (Wenzlau et al. 1989: Cell) (Fig. 14A). The cleavage on the two strands generate 4-nt 3' extensions, which overlap the site at which the intron is inserted in the intron-positive allele. Like intron endonucleases in the to other primary kingdoms, the bacteria and eukarya, I-Dmo I appears to have an asymmetric recognition sequence (Fig. 14A) and generates 3'-hydroxyl and 5'-phosphate termini that can be ligated with T4 DNA ligase (data not shown).

c) Conclusion

The protein encoded by the archaeal intron in the 23S rRNA gene of the hyperthermophile Desulfurococcus mobilis is a double-strand DNase that, like group I intron homing

endonucleases, is capable of cleaving an intron-negative allele of the gene. This enzyme, I-Dmo I, is unusual among the intron endonucleases in that it is thermostable and is expressed only from the linear and cyclized RNA intron species and not from the precursor RNA. Expression of several eukaryotic intron endonucleases is dependent upon splicing, whereas this is not the case for the phage enzymes. By analogy to its eukaryotic counterparts, but unlike the bacteriophage enzymes, I-Dmo I makes a staggered double-strand cut at an assymmetric target site that generates 4-nt 3' extensions (Fig. 14B). The cleavage locus proximal to the point at which the intron is inserted also mirrors that of the eukaryotic enzymes (Fig. 14), but contrasts with that of the phage intron endonucleases, which cleave at some distance from the intron insertion site. Additionally, although the archaeal and group I introns have entirely different structural properties and splicing pathways, I-Dmo I shares sequence similarity, in the form of the LAGLI-DADG motif, with group I intron endonucleases of eukaryotes. These observations support the independent evolutionary origin of endonucleases and intron core elements and are consistent with the invasive potential of endonuclease genes.

Although homing of the 23S rRNA intron cannot be demonstrated with the genetic tools currently available for Dc. mobilis, mobility is implied by the existence of intron-positive and intron-negative variants and by cleavage of the intron-negative allele by I-Dmo I encoded by the intron-positive variant. Proof of mobility in an archaeon would be of interest for several reasons. Homing introns are most prevalent in multicopy genomes, such as those of bacteriophages, mitochondria, and chloroplasts, and in repetitive nuclear rRNA genes. Hyperthermophilic rRNA genes are present in only one copy (Garrett et al. 1991); demonstration of homing a single-copy gene. Furthermore, coexistence of intron-positive and intron-negative alleles is implicit in

homing, and mobility would therefore suggest that interspecies genetic exchange takes place in hyperthermophilic archaea.

It has been argued that ORFs were acquired by preexisting introns, based on closely related introns bearing different ORFs (Perlman et al. 1989). For example, the homologous td, nrdB, and sunY introns of phage T4 contain three heterologous ORFs, and the Neurospora ND1 intron contains different ORFs at different locations depending on species. The discovery of an archaeal endonuclease in an intron that differs in structure and splicing pathway from the group I introns that frequent the other two kingdoms supports the argument that the endonuclease ORFs and the introns that house them arose separately (Bell-Pedersen et al. 1990).

The independent origin of the endonuclease ORF from that of the intron core is further supported by the existence of LAGLI-DADG-containing endonucleases in both freestanding form between genes (Nakagawa et al. 1991) and as protein fusions in an archaeon (Perler et al. 1992), a yeast, and a mycobacterium (Fig. 15). Another family of proteins, containing the GIY-YIG motif, is encoded by fungal mitochondrial introns and also occurs in the form of both freestanding and intron-encoded (Michel et al. 1986) endonucleases in phage T4. These observations underscore the phylogenetically widespread distribution of different families of endonucleases and also support their intron-independent ancestry, since the endonuclease genes do not occur in exclusive association with introns but also at other genetic loci.

The endonucleases have the ability to facilitate recombination by virtue of their double-strand cleavage activity. This property can facilitate the entry of DNA, including that of the endonuclease gene itself, into a genome. Whether or not the foreign DNA could be tolerated within a gene

would depend upon an innate splicing activity at either RNA or protein levels [the latter has been recently demonstrated for endonucleases in several different systems] or upon the fortuitous colonization of a site that confers this ability. Introns of different kinds provide such hospitable landing sites and, if invaded by endonuclease ORFs, can be converted into mobile elements. Mobility of the composite intron is then driven by the intron-encoded endonuclease, while the intron core splicing structure maintains gene function in subsequent transfer events (Bell-Pedersen et al. 1990). In this situation, as with endonuclease genes that are inserted directly into protein coding sequences and are spliced at the protein level, the endonuclease ORF can be viewed as the primary mobile element, with all of the invasive, propagative, and selfish properties of classical transposons.

The similarity of the endonucleases across phylogenetic boundaries, and particularly in late-arising lineages (Figs. 14B and 15), raises the possibility of horizontal transfer. Direct lateral transfer between the thermophilic archaea and mesophilic, late-branching eukarya would be impeded in either direction by genetic barriers in addition to functional constraints, since the endonucleases are inactive at the ambient temperature of their trans-kingdom counterparts. However, multistep lateral transfer involving organisms adapted to temperatures intermediate between those of the meso- and thermophiles is feasible. Nevertheless, until these endonuclease families are demonstrated to be absent from more primitive lineages, it remains possible that the endonucleases are ancient, having existed in the common ancestor that predated the divergence of the three major kingdoms.

Description and Properties of I-Dmo I

I-Dmo I has been shown to be active as a sequence-specific endonuclease when translated in vitro in rabbit reticulate

lysates as well as when expressed in vivo in E. coli. In the case of the in vitro translation, it has been shown above that I-Dmo I is active as expressed from both the circular and the truncated linear mRNA.

The cleavage pattern of the homing site by the enzyme is given below. The enzyme generates 3' extensions. The products of the cleavage reaction possess 3'-hydroxyls and 5'-phosphates. The products can be religated using T4 DNA ligase. The methods used to determine these data are given in the section "Characterization of Endonuclease Activities" above. The degree of sequence specificity of the enzyme is not known but the minimal recognition sequence has been mapped by the procedure described in Wenzlau, J.M., et al., 1989. A clone containing the homing site was sequenced in both directions using sequenase ("Stratagene"). To the sequencing reactions were after the extension reaction added 6 µl of cleavage mix, and the mixture was incubated 15 min at 65 °C. The reactions were stopped by adding 8 µl of formamide loading buffer. Cleavage products were separated on a 6 % acrylamide-urea gel alongside untreated sequencing reactions. The minimal recognition sequence gives an indication of the sequences necessary for binding of, and cleavage by, the enzyme. It should be noted that a larger region may be involved in that the enzyme probably allows some degree of sequence variation. The homing site and cleavage pattern of I-Dmo I is as follows:

5'-CCAAATGCCTTGCCGGGTAA|GTTCCGGCGCGCATGAA-3'
3'-GGTTTACGGAACGGCC|CATTCAAGGCCGCGCGTACTT-5'

where the staggered vertical line indicates the bonds cleaved, and the mapped minimal recognition site is underlined.

The enzymatic activity of I-Dmo I which is expressed in vitro in rabbit reticulocyte lysate has been characterized by using different buffer conditions. This data is only

indicative of optimal conditions for the enzyme in vitro and might also be different for the enzyme expressed in E. coli. The characterization was performed using enzyme synthesized in vitro in rabbit reticulocyte lysate. The method of synthesis is given in the section "Characterization of Endonuclease Activities" above.

The assays were performed on substrate endlabelled with ^{32}P using T4 DNA polymerase. The substrate used, the treatment of the samples after the assays and the method used for quantification of the cleavage are given in material and methods in appendix 4, here is also the characterization of the enzymes activity as a function of temperature. These data only give an indication of the conditions to use for optimal activity of I-Dmo I in vitro.

Activity as a function of temperature

The results of the assay for temperature dependence are given in the graph of Fig. 16.

Temperature Optimal activity was observed in the temperature interval 65 - 75 °C.

Activity as a function of pH.

Assay:
20 ng labelled substrate
50 mM Tris-HCl of varying pH
10 mM $MgCl_2$
25 mM NaCl
0.1 µl translation mixture
Total volume 20 µl.

Assayed at 65 °C for 10 min. The pH values of the buffers are given at 20 °C. The results are given in the graph of Fig. 17.

pH Optimal activity was observed in the range pH 7.8 - 8.4 using Tris-HCl buffer. The pH was measured at 20 °C, but assays were performed at 65 °C.

Activity in buffers containing different anions.

Assay: 1) 20 ng labelled substrate
50 mM Tris-Acetate pH 8.7
10 mM NH_4(Acetate)
10 mM Mg(Acetate)$_2$
0.1 µl translation mixture
Total volume 20 µl.

2) 20 ng labelled substrate
50 mM Tris-HCl pH 8.7
10 mM NH_4Cl
10 mM $MgCl_2$
0.1 µl translation mixture
Total volume 20 µl.

3) 20 ng labelled substrate
50 mM Tris-H_2SO_4 pH 8.7
10 mM $(NH_4)_2SO_4$
10 mM $MgSO_4$
0.1 µl translation mixture
Total volume 20 µl.

4) 20 ng labelled substrate
50 mM Tris-H_3PO_4 pH 8.7
10 mM NH_4Cl
10 mM $MgCl_2$
0.1 µl translation mixture
Total volume 20 µl.

Assayed at 65 °C for 10 min.

Anions — Optimal activity was higher in acetate than in sulfate and chloride; inhibition was caused by phosphate.

Activity in buffers containing different cations.

Different monovalent metal ions:

Assay:
20 ng labelled substrate
50 mM Tris-HCl pH 8.7
10 mM $MgCl_2$
25 mM chloride salts of different monovalent metal ions and ammonium.
0.1 µl translation mixture
Total volume 20 µl.

Assayed at 65 °C for 10 min.

Cations — Optimal activity was observed as follows $Cs+>NH_4+>K+>Na+>Li+$.

Activity in buffers containing varying concentrations of NH_4Cl.

Assay:
20 ng labelled substrate
50 mM Tris HCl pH 8.7
10 mM $MgCl_2$
Varying concentrations of NH_4Cl.
0.1 µl translation mixture cont. I-Dmo I
Total volume 20 µl.

Assayed at 65 °C for 10 min. The results are given in the graph of Fig. 18.

Ammonium — When varying the concentration of NH_4Cl, optimal activity was observed in the concentration range 1-50 mM.

Activity in buffers containing varying concentrations of $MgCl_2$.

Assay:
20 ng labelled taget
50 mM Tris-HCl pH 8.7
Varying concentration of $MgCl_2$
25 mM NaCl
0.1 µl translation mixture cont. I-Dmo I
Total volume 20 µl.

Assayed at 65 °C for 10 min. The results are given in the graph of Fig. 19.

Magnesium: The enzymatic activity is magnesium dependent. EDTA inhibits activity. Optimal activity was observed in the concentration range 4-25 mM $MgCl_2$.

The expression and purification of the active form of I-Dmo I from E. coli is described in the following example.

EXAMPLE

Expression and purification of I-Dmo I

Using this method in the order of 100 000 units of the enzyme can be purified from 20 g of induced cells of E. coli expressing I-Dmo I. One unit is defined as the enzyme necessary to cleave 1 µg of a pUC vector containing the homing sequence in one hour at 65 °C using the following conditions: 50 mM Tris-Acetate pH 8.0, 10 mM Magnesium Acetate, 10 mM Ammonium Acetate.

Expression:

The expression system developed by W.F. Studier and B.A. Moffatt (1986) was used for the expression of the enzyme,

with the exception that a modified strain of BL21(DE3) was used. The plasmid is described in the section "Characterization of Endonuclease Activities" above.

Strain used:

Bacterial strain:
BL21(DE3) F' lacI^q; hsd S gal (Λ cIts857 ind1 Sam7 nin5 lacUV5-T7 gene1) [F' pro AB+ lacI^q lacZ ΔM15 Tn10(tet^r)] pUT19-Dmo-c.

Media: LB media containing 100 µg ampicilin/liter.

Cultures were grown at 32 °C. A 20 ml culture was started from a glycerol stain, and grown overnight. A 1:100 dilution was made to a final volume of 750 ml. The culture was grown until OD_{600} reached 0.5, and then induced by adding IPTG to a final concentration of 0.4 mM. Induction was performed overnight (16 hours) (Fig. 20A). Cells were spun down and stored as a paste at -80 °C for subsequent purification of the enzyme. Typically 20 g of cell paste was obtained from an induction.

Purification:

A) Sonication and DNAse I treatment

The cell paste was suspended in 50 ml lysis buffer and incubated 30 min at 0 °C, and was subsequent sonicated until low viscosity was obtained. Manganese chloride and magnesium chloride were added to final concentrations of respectively 1 mM and 10 mM. 1 mg DNAse I (BRL) was added, and the suspension was incubated at 0 °C for 1 hour. In order to remove cell debris the lysate was spun at 10 000 rpm in a Sorwall centrifuge (4 °C) (Fig. 20B 2-3).

Lysis buffer: 50 mM Tris-HCl pH 7.5, 20 mM NaCl, 1 mM DTT,0.8 mg/ml lysozyme.

B) Ammonium sulfate precipitation

Ammonium sulfate was added to the supernatant to a saturation of 20%. Ammonium sulfate precipitation was performed on ice for 30 min. The precipitation mixture was spun for 20 min at 12 000 rpm in a Sorwall centrifuge (4 °C). The volume of the supernatant was measured, and ammonium sulfate was added to a final saturation of 45%. Incubation on ice and subsequent centrifugation was repeated (Fig. 20B 4-5). The pellet was stored overnight at 4 °C.

C) Heparin column purification

A 5 ml prepackaged Heparin Hiload (Pharmacia) column was used for the chromatography which was performed on FPLC equipment from Pharmacia.

The pellet was dissolved in 15 ml Buffer A, and dialyzed 2 hours against 2 liter Buffer A. The sample was applied on the heparin column using a flow of 0.5 ml/min. The column was washed with 10 ml 25 ml Buffer A, and bound enzyme was eluted using a gradient from 0 to 100% Buffer B over 50 ml. The enzyme eluted at a salt concentration ranging from 700 to 1200 mM NaCl (Fig. 20B 6-18). The fractions containing activity were pooled. The enzyme at this stage was pure enough for most applications and was stored the following way:

The pooled fractions were concentrated using centricon 10 spin columns (Amicon) and transferred to Buffer C using same method. To the concentrated enzyme solution was added an equal volume of glycerol, and the preparation was stored at -20 °C.

If pure enzyme is needed, the protein can be purified further by means of a phenyl-"Sepharose" column.

Buffer A:	50	mM Tris HCl pH 8.0
	500	mM NaCl
	1	mM DTT
Buffer B:	50	mM Tris HCl pH 8.0
	4	M NaCl
	1	mM DTT
Buffer C:	50	mM Tris Acetate pH 7.0
	10	mM NH_4 Acetate
	1	mM DTT
	0.1	% Tween 20

D) Phenyl-"Sepharose" column purification

A 1 ml phenyl-"Sepharose" column for FPLC from Pharmacia was used. The phenyl-"Sepharose" was first equilibrated with Buffer B, then with Buffer A. The pooled fractions from the heparin column were applied without prior treatment. The column was washed in 5 ml Buffer A, and subsequently in 5 ml Buffer B. The bound protein was eluted in Buffer C and stored at -20 °C (Fig. 20C). If large amounts of enzyme was purified, it was necessary to add 0.1% "Tween 20" to Buffer C in order to elute the enzyme.

Buffer A:	50	mM Tris HCl pH 8.0
	500	mM NaCl
	1	mM DTT
Buffer B:	50	mM Tris HCl pH 8.0
	1	mM DTT
Buffer C:	50	mM Tris Acetate pH 7.0
	1	mM DTT

50 % Ethylene glycol

The same procedure can be used to express and purify the other endonucleases comprised by the invention.

Characterization of Endonuclease Activity and Specificity of I-Por I

In order to test for endonuclease activity of the protein encoded by intron 1 of Pyrobaculum organotrophum 23S rRNA, the ORF was expressed in E. coli using the system applied to I-Dmo I in the above Example. A 23S rDNA substrate was prepared from the closely related organism Pyrobaculum islandicum (Huber et al., 1987). A fragment of Pb. islandicum 23S rRNA covering most of domains IV and V was cloned into M13mp18. The 23S rDNA fregment was excised together with flanking regions of the vector end isolated as a 7.5 kb linear fragment as illustrated in Fig. 21A. This was subjected to digestion with both I-Dmo I and with the putative endonuclease I-Por I. The results shown in Fig. ?B demonstrate that I-Por I cuts about 100 nucleotides upstream from the I-Dmo I site. It yielded fragments of about 5450 bp and 2450 bp (weak fragment) (Fig. 21B) corresponding to the location of intron 1 in Pb. organotrophum.

The actual site of cleavage was determined using the same procedures as stated for I-Dmo I above, and the cleavage site mapping is illustrated in Fig. 22. The cutting site corresponds to a 4 nucleotide 3'-overhang as for I-Dmo I or the site of ligation of the intron in the 23S rRNA of Pb. organotrophum. The minimal recognition sequence has also been mapped by the procedure described in Wenzlau, J.M., et al., 1989, as for I-Dmo I above. The homing site and cleavage pattern of I-Por I is as follows:

35

5'-CCCCGCGAGCCCGTAAGGGTGTGTACGGGGGCTGAA-3'
3'-GGGGCGCTCGGGCATTCCCACACATGCCCCCGACTT-5'

where the staggered vertical line indicates the bonds cleaved, and the mapped minimal recognition site is underlined.

REFERENCES

Bell-Pedersen, D., Quirk, S., Clyman, J. and Belfort, M.: Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. Nucleic Acids Res. **18** (1990), 3763-3770.

Chomczynski, P. and Sacchi, N.: Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162** (1987), 156-159.

Devereux, H., Haeberli, P. and Smithies O.: A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12** (1984), 387-395.

Dujon, B., Belfort, M., Butow, R.A., Jacq, C., Lemieux, C., Perlman, P.S., and Vogt, V.M. (1989), Gene **82**, 115-118.

Garnier, J., Osguthorpe, D.J. and Robson, B.: Analysis of accuracy and implications of simple methods for predicting secondary structure of globular proteins. J. Mol. Biol. **120** (1978), 97-120.

Garrett, R.A., Dalgaard, J., Larsen, J., Kjems, J. and Mankin, A.S.: Archaeal rRNA operons. Trends Biochem. Sci. **16** (1991), 22-26.

Huber, R., Kristjansson, J.K. and Stetter, K.O.: Pyrobaculum gen.nov., a new genus of neutrophilic rod-shaped archaebacteria from continental solfataras growing optimally at 100 °C. Arch. Microbiol. **140** (1987), 95-101.

Kjems, J. and Garrett, R.A.: An intron in the 23S ribosomal RNA gene of the archaebacterium Desulfuroccocus mobilis. Nature **318** (1985), 675-677.

Kjems, J. and Garrett, R.A.: Novel expression of the rRNA genes in the extreme thermophile and archaebacterium Desulfurococcus mobilis. EMBO J. 6 (1987), 3521-3530.

Kjems, J. and Garrett, R.A.: Novel splicing mechanism for the ribosomal RNA intron in the archaebacterium Desulfurococcus mobilis. Cell 54 (1988), 693-703.

Kjems, J., Jensen, J., Olesen, T. and Garrett, R.A.: Comparison of transfer RNA and ribosomal RNA intron splicing in the extreme thermophile and archaebacterium Desulfurococcus mobilis. Can. J. Microbiol. 35 (1989), 210-214.

Kjems, J. and Garrett, R.A.: Ribosomal RNA introns in archaea and evidence for RNA conformational changes associated with splicing. Proc. Natl. Acad. Sci. USA 88 (1991), 439-443.

Michel, F., and Dujon, B., (1986), Cell 46, 323-***.

Nakagawa, K., Morishima, N., and Shibata, T. (1991), J. Biol. Chem. 266, 1977-1984.

Perler, F.B., Comb, D.G., Jack, W.E., Moran, L.S., Ouiang, B., Kuccra, R.B., Benner, J., Slatko, B.E., Nwankwo, D.O., Hempstead, S.K., Carlow, C.K., and Jannasch, H., (1992), Proc. Natl. Acad. Sci. USA 89, 5577-5581.

Perlman, P.S. and Butow, R.A.: Mobile introns and intron-encoded proteins. Science 246 (1989), 1106-1109.

Sambrook, J., Fritsch, E.F. and Maniatis, T.: Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor 1989.

Stetter, K.O., Fiala, G., Huber, G., Huber, R. and Segerer, A.: Hyperthermophilic organisms. FEMS Microbiol. Rev. 75 (1990), 117-124.

Studier, F.W. and B.A. Moffatt (1986), J. Mol. Biol. 189, 113.

Thompson, L.D., Brandon, L.D., Nieuwlandt, D.T., and Daniels, C.J. (1989), Can. J. Microbiol. 35, 36-42.

Thompson, L.D. and Daniels, C.J.: Recognition of exon-intron boundaries by the Halobacterium volcanii tRNA intron endonuclease. J. Biol. Chem. 265 (1990), 18104-18111.

Wenzlau, J.M., Saldanha, R.J., Butow, R.A., and Perlman, P.S. (1989), Cell 56, 421-430.

Woese, C.R., Kandler, O., and Wheelis, M.L., (1990), Proc. Natl. Acad. Sci. USA 87, 4576-4579.

Yanish-Perron, C., Vieira, J. and Messing, J.: Improved M13 phage cloning vectors and host strains:

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Jacob Zeuthen Dalgaard
(B) STREET: Pragtstjernevej 24, Emdrup
(C) CITY: DK-2400 Copenhagen NV
(E) COUNTRY: Denmark
(F) POSTAL CODE (ZIP): DK-2400

(A) NAME: Roger Antony Garrett
(B) STREET: Toeltvej 26
(C) CITY: DK-3490 Kvistgaard
(E) COUNTRY: Denmark
(F) POSTAL CODE (ZIP): DK-3490
(G) TELEPHONE: +45 42192428

(A) NAME: Joergen Kjems
(B) STREET: Solbakken 15
(C) CITY: DK-8240 Risskov
(E) COUNTRY: Denmark
(F) POSTAL CODE (ZIP): DK-8240

(ii) TITLE OF INVENTION: Thermostable Sequence-specific DNA Endonucleases

(iii) NUMBER OF SEQUENCES: 7

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: DK 1015/92
(B) FILING DATE: 14-AUG-1992

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 622 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Desulfurococcus mobilis

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 1..622
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /function= "Codes for endonuclease I-Dmo I"
/evidence= EXPERIMENTAL
/standard_name= "Dmo 23S rRNA intron 1"
/label= Dmo_intron_1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 59..622
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /codon_start= 59
/function= "genome cleavage"
/product= "DNA endonuclease"
/evidence= EXPERIMENTAL
/standard_name= "Endonuclease from 23S rRNA intron 1 of Dc. mobilis"
/label= I-Dmo_I
/note= "After excision, the 23S rRNA intron circularizes, and the coding sequence runs from nucleotide 59 to nucleotide 21 (stop codon)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
AGGGCAGGGG GTTACACTTA AAAATTAGCA TTATACTTAG TTCGATGGGG TAGGATTA        58

ATG CAT AAT AAT GAG AAT GTT AGT GGA ATA TCT GCT TAC CTG CTT GGA       106
Met His Asn Asn Glu Asn Val Ser Gly Ile Ser Ala Tyr Leu Leu Gly
  1               5                  10                  15

TTG ATA ATA GGT GAT GGA GGA CTT TAC AAG TTA AAA TAT AAG GGT AAC       154
Leu Ile Ile Gly Asp Gly Gly Leu Tyr Lys Leu Lys Tyr Lys Gly Asn
             20                  25                  30

AGA AGC GAA TAT AGG GTT GTA ATA ACG CAA AAG TCT GAA AAC TTA ATT       202
Arg Ser Glu Tyr Arg Val Val Ile Thr Gln Lys Ser Glu Asn Leu Ile
        35                  40                  45

AAA CAA CAC ATA GCA CCA TTA ATG CAA TTT CTC ATA GAT GAA CTA AAT       250
Lys Gln His Ile Ala Pro Leu Met Gln Phe Leu Ile Asp Glu Leu Asn
     50                  55                  60

GTG AAA TCA AAA ATA CAG ATC GTT AAG GGT GAT ACA AGA TAT GAG TTA       298
Val Lys Ser Lys Ile Gln Ile Val Lys Gly Asp Thr Arg Tyr Glu Leu
 65                  70                  75                  80

AGA GTA TCC AGT AAG AAA CTA TAC CAT TAT TTC GCT AAC ATG CTA GAG       346
Arg Val Ser Ser Lys Lys Leu Tyr His Tyr Phe Ala Asn Met Leu Glu
                 85                  90                  95
```

```
AGG ATA AGG TTA TTC AAT ATG CGT GAG CAA ATA GCG TTC ATA AAG GGG      394
Arg Ile Arg Leu Phe Asn Met Arg Glu Gln Ile Ala Phe Ile Lys Gly
            100                 105                 110

CTA TAT GTA GCT GAA GGA GAT AAA ACC CTC AAG AGA CTA AGA ATT TGG      442
Leu Tyr Val Ala Glu Gly Asp Lys Thr Leu Lys Arg Leu Arg Ile Trp
        115                 120                 125

AAT AAG AAT AAA GCA TTA CTA GAA ATA GTA TCG CGA TGG TTA AAT AAC      490
Asn Lys Asn Lys Ala Leu Leu Glu Ile Val Ser Arg Trp Leu Asn Asn
    130                 135                 140

CTG GGT GTA AGG AAT ACT ATT CAC TTG GAT GAT CAT AGG CAC GGT GTA      538
Leu Gly Val Arg Asn Thr Ile His Leu Asp Asp His Arg His Gly Val
145                 150                 155                 160

TAT GTA TTA AAT ATT TCA CTC AGA GAT AGA ATT AAG TTT GTT CAC ACA      586
Tyr Val Leu Asn Ile Ser Leu Arg Asp Arg Ile Lys Phe Val His Thr
                165                 170                 175

ATT CTC TCT TCA CAC CTT AAC CCC CTG CCC CCC GAG                      622
Ile Leu Ser Ser His Leu Asn Pro Leu Pro Pro Glu
            180                 185
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 188 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met His Asn Asn Glu Asn Val Ser Gly Ile Ser Ala Tyr Leu Leu Gly
  1               5                  10                  15

Leu Ile Ile Gly Asp Gly Gly Leu Tyr Lys Leu Lys Tyr Lys Gly Asn
            20                  25                  30

Arg Ser Glu Tyr Arg Val Val Ile Thr Gln Lys Ser Glu Asn Leu Ile
        35                  40                  45

Lys Gln His Ile Ala Pro Leu Met Gln Phe Leu Ile Asp Glu Leu Asn
    50                  55                  60

Val Lys Ser Lys Ile Gln Ile Val Lys Gly Asp Thr Arg Tyr Glu Leu
65                  70                  75                  80

Arg Val Ser Ser Lys Lys Leu Tyr His Tyr Phe Ala Asn Met Leu Glu
                85                  90                  95

Arg Ile Arg Leu Phe Asn Met Arg Glu Gln Ile Ala Phe Ile Lys Gly
            100                 105                 110
```

```
Leu Tyr Val Ala Glu Gly Asp Lys Thr Leu Lys Arg Leu Arg Ile Trp
        115                 120                 125

Asn Lys Asn Lys Ala Leu Leu Glu Ile Val Ser Arg Trp Leu Asn Asn
    130                 135                 140

Leu Gly Val Arg Asn Thr Ile His Leu Asp Asp His Arg His Gly Val
145                 150                 155                 160

Tyr Val Leu Asn Ile Ser Leu Arg Asp Arg Ile Lys Phe Val His Thr
                165                 170                 175

Ile Leu Ser Ser His Leu Asn Pro Leu Pro Pro Glu
            180                 185
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 194 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Desulfurococcus mobilis

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION: 1..194

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
Met His Asn Asn Glu Asn Val Ser Gly Ile Ser Ala Tyr Leu Leu Gly
1               5                   10                  15

Leu Ile Ile Gly Asp Gly Gly Leu Tyr Lys Leu Lys Tyr Lys Gly Asn
            20                  25                  30

Arg Ser Glu Tyr Arg Val Val Ile Thr Gln Lys Ser Glu Asn Leu Ile
        35                  40                  45

Lys Gln His Ile Ala Pro Leu Met Gln Phe Leu Ile Asp Glu Leu Asn
    50                  55                  60

Val Lys Ser Lys Ile Gln Ile Val Lys Gly Asp Thr Arg Tyr Glu Leu
65                  70                  75                  80

Arg Val Ser Ser Lys Lys Leu Tyr Tyr Tyr Phe Ala Asn Met Leu Glu
                85                  90                  95
```

Arg Ile Arg Leu Phe Asn Met Arg Glu Gln Ile Ala Phe Ile Lys Gly
100 105 110

Leu Tyr Val Ala Glu Gly Asp Lys Thr Leu Lys Arg Leu Arg Ile Trp
115 120 125

Asn Lys Asn Lys Ala Leu Leu Glu Ile Val Ser Arg Trp Leu Asn Asn
130 135 140

Leu Gly Val Arg Asn Thr Ile His Leu Asp Asp His Arg His Gly Val
145 150 155 160

Tyr Val Leu Asn Ile Ser Leu Arg Asp Arg Ile Lys Phe Val His Thr
165 170 175

Ile Leu Ser Ser His Leu Asn Pro Leu Pro Pro Glu Arg Ala Gly Gly
180 185 190

Tyr Thr

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 607 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Pyrobaculum organotrophum

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 1..607
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /function= "Codes for endonuclease I-Por I"
/evidence= EXPERIMENTAL
/standard_name= "Por 23S rRNA intron 1"
/label= Por_intron_1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 50..574
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /codon_start= 50
/function= "genome cleavage"
/product= "DNA endonuclease"
/evidence= EXPERIMENTAL

/standard_name= "Endonuclease from 23S rRNA intron 1 of Pb. organotrophum"
/label= I-Por_I

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
AGGGCGGTAA GGGAGCCAAG CTTAAAAATA CGGGGGTGGC AATTGTATA GTG GAT          55
                                                        Val Asp
                                                          1

ATA TTC CAG TAT GTC AAA GCT CTC AGC GCT TTT GAT GGC TAT GTT GAT        103
Ile Phe Gln Tyr Val Lys Ala Leu Ser Ala Phe Asp Gly Tyr Val Asp
          5                  10                  15

TAC AGA GTC GGC AAA AAC TAC GAA ATC GTA ATC GCC GAT ATG TCT AAG        151
Tyr Arg Val Gly Lys Asn Tyr Glu Ile Val Ile Ala Asp Met Ser Lys
     20                  25                  30

GAA TTC CTC GAG GAA ATA TGC AAC GAA CTA AGA AAA TAC GGC GTG AGT        199
Glu Phe Leu Glu Glu Ile Cys Asn Glu Leu Arg Lys Tyr Gly Val Ser
 35                  40                  45                  50

TGC GGA GTC TAC GCT TCG CGA AGA GAC AGA GCC TTT AGG CTT AGA ATA        247
Cys Gly Val Tyr Ala Ser Arg Arg Asp Arg Ala Phe Arg Leu Arg Ile
                 55                  60                  65

TAC GGC AAG GAG TCT GTC GAT AGA ATC TTG AAC TCC AGC CTT GCC CCC        295
Tyr Gly Lys Glu Ser Val Asp Arg Ile Leu Asn Ser Ser Leu Ala Pro
             70                  75                  80

GAG GTA CTA TTA GCG GCT GCA ATA GAT GCC GAG GGG AAT GTG AAG AAA        343
Glu Val Leu Leu Ala Ala Ala Ile Asp Ala Glu Gly Asn Val Lys Lys
         85                  90                  95

TAC AGA AAT CAG CCT TTT AGA ACC AGA ATT GTC GTA AAG AGT GAC ATG        391
Tyr Arg Asn Gln Pro Phe Arg Thr Arg Ile Val Val Lys Ser Asp Met
    100                 105                 110

GCT AGA CGA ATT GAA GAT GCG CTC ACT GCA CTA TCG ATA AGA TAC GTT        439
Ala Arg Arg Ile Glu Asp Ala Leu Thr Ala Leu Ser Ile Arg Tyr Val
115                 120                 125                 130

AAG ATC ACT AGG AAA GGA GGT AGA TAC ACT GAA ATC GTC GTA TCT GGT        487
Lys Ile Thr Arg Lys Gly Gly Arg Tyr Thr Glu Ile Val Val Ser Gly
                135                 140                 145

AAG GAA GAG AAC CAA AAG CTG TAT CGG GTA GTT AAA ATA CGC CAC CCC        535
Lys Glu Glu Asn Gln Lys Leu Tyr Arg Val Val Lys Ile Arg His Pro
            150                 155                 160

CAA AAA TTA CAG GTG GTG TGC CAT TAT CTT GAC CTC TAGGCCGAAC             581
Gln Lys Leu Gln Val Val Cys His Tyr Leu Asp Leu
        165                 170                 175

AAGTTGGCTC CCTTCCGCCC CGGGAG                                           607
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
Val Asp Ile Phe Gln Tyr Val Lys Ala Leu Ser Ala Phe Asp Gly Tyr
 1               5                  10                  15

Val Asp Tyr Arg Val Gly Lys Asn Tyr Glu Ile Val Ile Ala Asp Met
            20                  25                  30

Ser Lys Glu Phe Leu Glu Glu Ile Cys Asn Glu Leu Arg Lys Tyr Gly
        35                  40                  45

Val Ser Cys Gly Val Tyr Ala Ser Arg Arg Asp Arg Ala Phe Arg Leu
    50                  55                  60

Arg Ile Tyr Gly Lys Glu Ser Val Asp Arg Ile Leu Asn Ser Ser Leu
65                  70                  75                  80

Ala Pro Glu Val Leu Leu Ala Ala Ala Ile Asp Ala Glu Gly Asn Val
                85                  90                  95

Lys Lys Tyr Arg Asn Gln Pro Phe Arg Thr Arg Ile Val Val Lys Ser
            100                 105                 110

Asp Met Ala Arg Arg Ile Glu Asp Ala Leu Thr Ala Leu Ser Ile Arg
        115                 120                 125

Tyr Val Lys Ile Thr Arg Lys Gly Gly Arg Tyr Thr Glu Ile Val Val
    130                 135                 140

Ser Gly Lys Glu Glu Asn Gln Lys Leu Tyr Arg Val Val Lys Ile Arg
145                 150                 155                 160

His Pro Gln Lys Leu Gln Val Val Cys His Tyr Leu Asp Leu
                165                 170
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Pyrobaculum organotrophum

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 1..598
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /function= "Codes for endonuclease I-Por II"
/evidence= EXPERIMENTAL
/standard_name= "Por 23S rRNA intron 2"
/label= Por_intron_2

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 46..552
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /codon_start= 46
/function= "genome cleaving"
/product= "DNA endonuclease"
/evidence= EXPERIMENTAL
/standard_name= "Endonuclease from 23S rRNA intron 2 of Pb. organotrophum"
/label= I-Por_I

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
AGGCGGGTGC AGGAGGGGGT TTTTTATACC CGTTTATTTT TCTCC ATG GGC TGT          54
                                                  Met Gly Cys
                                                    1

GAT TTG GAG TAT CTA GCT GGC GTG GTT AAG GGT GAT GGG ACG CTG TAC       102
Asp Leu Glu Tyr Leu Ala Gly Val Val Lys Gly Asp Gly Thr Leu Tyr
  5                  10                  15

CAC AAC AAG AAG GCT AGG GAG TAT GTG GTT GAG ATC TAT GAT AGG GAT       150
His Asn Lys Lys Ala Arg Glu Tyr Val Val Glu Ile Tyr Asp Arg Asp
 20                  25                  30                  35

GTG GAG TAC GTG GCG ATT CTG GTT GAT ATG TTG AAG AGT TGC GGT TTA       198
Val Glu Tyr Val Ala Ile Leu Val Asp Met Leu Lys Ser Cys Gly Leu
                 40                  45                  50

AAT CCC CAT GTG AGG TCG TAT GGG AAC TAT TAC AGG GTT AGG GTG AAC       246
Asn Pro His Val Arg Ser Tyr Gly Asn Tyr Tyr Arg Val Arg Val Asn
            55                  60                  65

AGT AGG GAG TTT TAC GAA TCT GTC AGA GGC GCC ATT GAG CGG CTT CTA       294
Ser Arg Glu Phe Tyr Glu Ser Val Arg Gly Ala Ile Glu Arg Leu Leu
        70                  75                  80
```

```
GTA TCT CCG ACG GTG CCG TTT GTG CGT GGG TTG TTT GAT AGC GAC GGC        342
Val Ser Pro Thr Val Pro Phe Val Arg Gly Leu Phe Asp Ser Asp Gly
     85                  90                  95

ACT CTC TAC TTT GAC AGG AGG AAG AGG CGT CTC TAT CCC GTG GTG GAG        390
Thr Leu Tyr Phe Asp Arg Arg Lys Arg Arg Leu Tyr Pro Val Val Glu
100                 105                 110                 115

TTG GGG AAT TCG GAT TGG AGA GTT GTA AAC GCC GCT GCT GTG GTG CTT        438
Leu Gly Asn Ser Asp Trp Arg Val Val Asn Ala Ala Ala Val Val Leu
                120                 125                 130

TCG TCG TTT GGC GTT AAG TTT AGC GTT AAG AGC TAT GGG GGG AGG TTC        486
Ser Ser Phe Gly Val Lys Phe Ser Val Lys Ser Tyr Gly Gly Arg Phe
            135                 140                 145

TTC AAG TTG GTG GTT AGG GGG ACG CCG TGT TGT TTG CCA GAG TTG TCA        534
Phe Lys Leu Val Val Arg Gly Thr Pro Cys Cys Leu Pro Glu Leu Ser
        150                 155                 160

AGC CTC TCC ACC CGG TGAAGTTCTC CCCTCTTTCT TCTCTAGCCC CCTCCAGCCC        589
Ser Leu Ser Thr Arg
    165

GCCCCCCAA                                                              598
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 168 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
Met Gly Cys Asp Leu Glu Tyr Leu Ala Gly Val Val Lys Gly Asp Gly
  1               5                  10                  15

Thr Leu Tyr His Asn Lys Lys Ala Arg Glu Tyr Val Val Glu Ile Tyr
            20                  25                  30

Asp Arg Asp Val Glu Tyr Val Ala Ile Leu Val Asp Met Leu Lys Ser
        35                  40                  45

Cys Gly Leu Asn Pro His Val Arg Ser Tyr Gly Asn Tyr Tyr Arg Val
    50                  55                  60

Arg Val Asn Ser Arg Glu Phe Tyr Glu Ser Val Arg Gly Ala Ile Glu
65                  70                  75                  80

Arg Leu Leu Val Ser Pro Thr Val Pro Phe Val Arg Gly Leu Phe Asp
                85                  90                  95
```

Ser Asp Gly Thr Leu Tyr Phe Asp Arg Arg Lys Arg Arg Leu Tyr Pro
100 105 110

Val Val Glu Leu Gly Asn Ser Asp Trp Arg Val Val Asn Ala Ala Ala
115 120 125

Val Val Leu Ser Ser Phe Gly Val Lys Phe Ser Val Lys Ser Tyr Gly
130 135 140

Gly Arg Phe Phe Lys Leu Val Val Arg Gly Thr Pro Cys Cys Leu Pro
145 150 155 160

Glu Leu Ser Ser Leu Ser Thr Arg
165

P A T E N T C L A I M S

1. A thermostable sequence-specific DNA endonuclease, c h a r a c t e r i z e d in that it has the amino acid sequence encoded by an archaeal type intron of a stable RNA (ribosomal RNA or transfer RNA) or protein gene or is an enzymatically active variant thereof in which one or more amino acid residues have been deleted, inserted or substituted by other amino acids.

2. An endonuclease according to claim 1, c h a r a c t e r i z e d in that it has the amino acid sequence encoded by an archeal type intron of a rRNA gene or is an enzymatically active variant thereof in which one or more amino acid residues have been deleted, inserted or substituted by other amino acids.

3. An endonuclease according to claim 1 or 2 which has the amino acid sequence encoded by the 622 bp intron in the 23S rRNA gene of Desulfurococcus mobilis with the nucleotide sequence stated in SEQ ID No. 1.

4. An endonuclease according to claim 3 which is produced by expression of RNA transcribed from DNA having the nucleotide sequence stated in SEQ ID No. 1, said endonuclease having the amino acid sequence stated in SEQ ID No. 2.

5. An endonuclease according to claim 3 which is produced by expression of cyclic RNA transcribed from DNA having the nucleotide sequence stated in SEQ ID No. 1, the coding sequence running from nucleotide 59 past nucleotide 622 to end at nucleotide 21 (stop codon), said endonuclease having the amino acid sequence stated in SEQ ID No. 3.

6. An endonuclease according to claim 1 or 2 which has the amino acid sequence encoded by the 607 bp intron 1 in the 23S rRNA gene of Pyrobaculum organotrophum with the nucleo-

tide sequence stated in SEQ ID No. 4, said amino acid sequence being as stated in SEQ ID No. 5.

7. An endonuclease according to claim 1 or 2 which has the amino acid sequence encoded by the 598 bp intron 2 in the 23S rRNA gene of Pyrobaculum organotrophum with the nucleotide sequence stated in SEQ ID No. 6, said amino acid sequence being as stated in SEQ ID No. 7.

8. The use of a DNA endonuclease according to any one of the preceding claims as an endonuclease tool for gene analysis and manipulation.

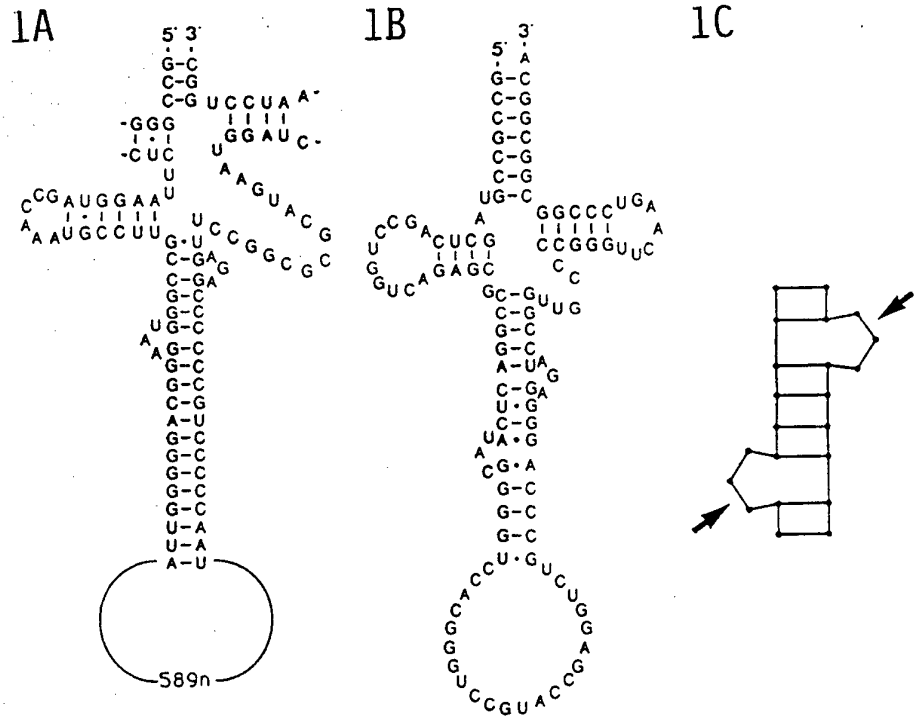


FIG. 1

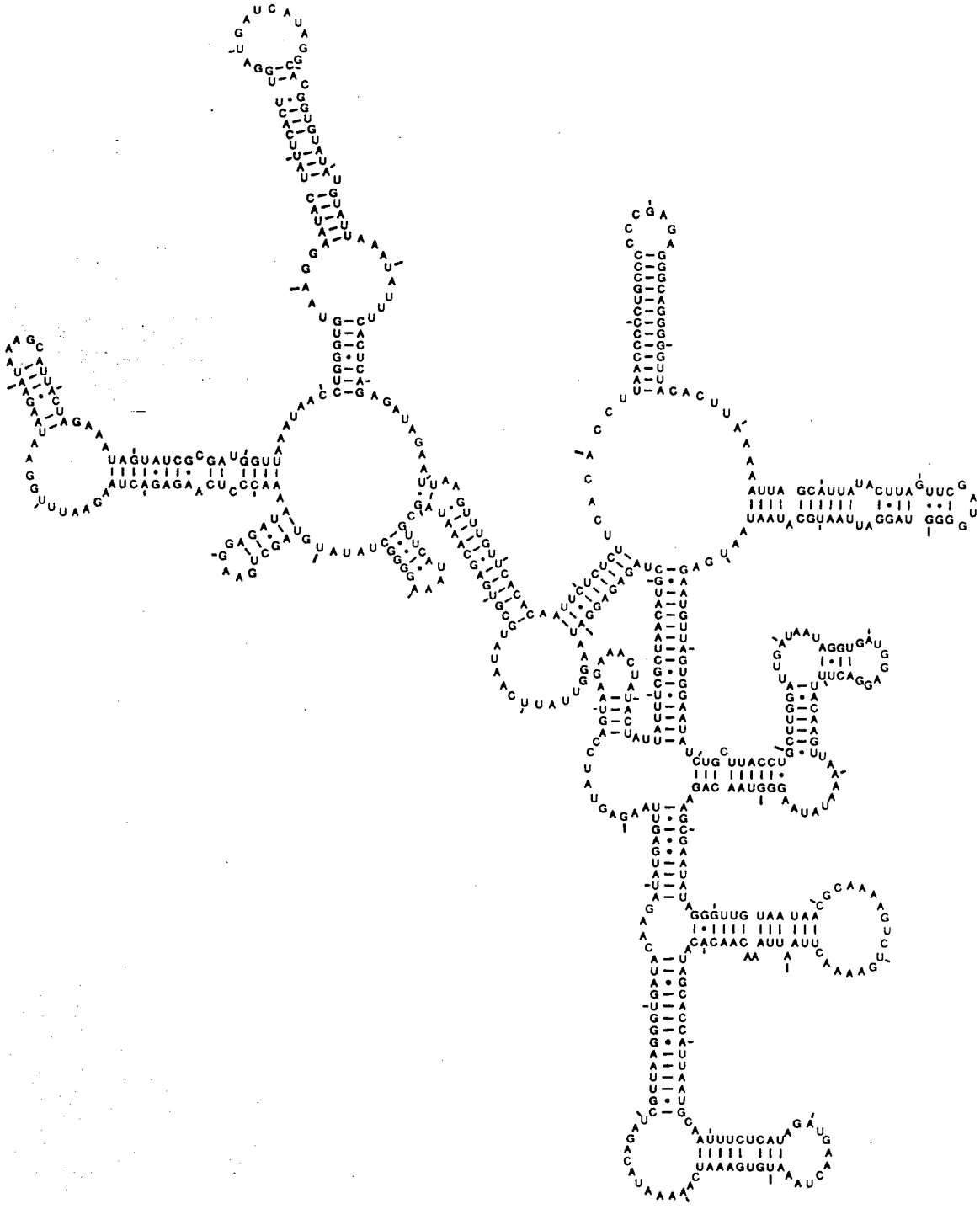


FIG. 2

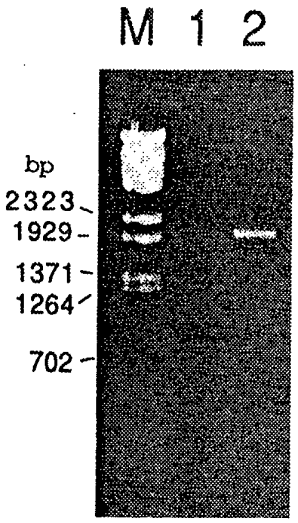


FIG. 3

```
EXON 1 - TAGGTCCCCG CGAGCCCGTA AGGGCGGTAA GGGAGCCAAG CTTAAAAATA  30

CGGGGGTGGC AATTGTATAG TGGATATATT CCAGTATGTC AAAGCTCTCA GCGCTTTTGA  90
                          V  D  I   F  Q  Y  V   K  A  L   S  A  F
TGGCTATGTT GATTACAGAG TCGGCAAAAA CTACGAAATC GTAATCGCCG ATATGTCTAA 150
D  G  Y  V   D  Y  R   V  G  K   N  Y  E  I   V  I  A   D  M  S
GGAATTCCTC GAGGAAATAT GCAACGAACT AAGAAAATAC GGCGTGAGTT GCGGAGTCTA 210
K  E  F  L   E  E  I   C  N  E   L  R  K  Y   G  V  S   C  G  V
CGCTTGCGCA AGAGACAGAG CCTTTAGGCT TAGAATATAC GGCAAGGAGT CTGTCGATAG 270
Y  A  C  A   R  D  R   A  F  R   L  R  I  Y   G  K  E   S  V  D
AATCTTGAAC TCCAGCCTTG CCCCCGAGGT ACTATTAGCG GCTGCAATAG ATGCCGAGGG 330
R  I  L  N   S  S  L   A  P  E   V  L  L  A   A  A  I   D  A  E
GAATGTGAAG AAATACAGAA ATCAGCCTTT TAGAACCAGA ATTGTCGTAA AGAGTGACAT 390
G  N  V  K   K  Y  R   N  Q  P   F  R  T  R   I  V  V   K  S  D
GGCTAGACGA ATTGAAGATG CGCTCACTGC ACTATCGATA AGATACGTTA AGATCACTAG 450
M  A  R  R   I  E  D   A  L  T   A  L  S  I   R  Y  V   K  I  T
GAAAGGAGGT AGATACACTG AAATCGTCGT ATCTGGTAAG GAAGAGAACC AAAAGCTGTA 510
R  K  G  G   R  Y  T   E  I  V   V  S  G  K   E  E  N   Q  K  L
TCGGGTAGTT AAAATACGCC ACCCCCAAAA ATTACAGGTG GTGTGCCATT ATCTTGACCT 570
Y  R  V  V   K  I  R   H  P  Q   K  L  Q  V   V  C  H   Y  L  D
CTAGGCCGAA CAAGTTGGCT CCCTTCCGCC CCGGGAGAGG GTGTGTACGG GGGCTGA -  607
L  *
                    EXON 2 (722 nt)

      - CAGCCGCCAA GGGTGGGGCT AGGCGGGTGC AGGAGGGGGT TTTTTATACC  30

CGTTTATTTT TCTCCATGGG CTGTGATTTG GAGTATCTAG CTGGCGTGGT TAAGGGTGAT  90
                   M   G  C  D  L   E  Y  L   A  G  V   V  K  G  D
GGGACGCTGT ACCACAACAA GAAGGCTAGG GAGTATGTGG TTGAGATCTA TGATAGGGAT 150
  G  T  L   Y  H  N   K  K  A  R   E  Y  V   V  E  I   Y  D  R  D
GTGGAGTACG TGGCGATTCT GGTTGATATG TTGAAGAGTT GCGGTTTAAA TCCCCATGTG 210
  V  E  Y   V  A  I   L  V  D  M   L  K  S   C  G  L   N  P  H  V
AGGTCGTATG GGAACTATTA CAGGGTTAGG GTGAACAGTA GGGAGTTTTA CGAATCTGTC 270
  R  S  Y   G  N  Y   Y  R  V  R   V  N  S   R  E  F   Y  E  S  V
AGAGGCGCCA TTGAGCGGCT TCTAGTATCT CCGACGGTGC CGTTTGTGCG TGGGTTGTTT 330
  R  G  A   I  E  R   L  L  V  S   P  T  V   P  F  V   R  G  L  F
GATAGCGACG GCACTCTCTA CTTTGACAGG AGGAAGAGGC GTCTCTATCC CGTGGTGGAG 390
  D  S  D   G  T  L   Y  F  D  R   R  K  R   R  L  Y   P  V  V  E
TTGGGGAATT CGGATTGGAG AGTTGTAAAC GCCGCTGCTG TGGTGCTTTC GTCGTTTGGC 450
  L  G  N   S  D  W   R  V  V  N   A  A  A   V  V  L   S  S  F  G
GTTAAGTTTA GCGTTAAGAG CTATGGGGGG AGGTTCTTCA AGTTGGTGGT TAGGGGGACG 510
  V  K  F   S  V  K   S  Y  G  G   R  F  F   K  L  V   V  R  G  T
CCGTGTTGTT TGCCAGAGTT GTCAAGCCTC TCCACCCGGT GAAGTTCTCC CCTCTTTCTT 570
  P  C  C   L  P  E   L  S  S  L   S  T  R   *
CTCTAGCCCC CTCCAGCCCG CCCCCCAAGC CCGCCCATTA AAGGGGAA - EXON 3     598
```

Intron 1 (Domain IV)

Intron 2 (Domain V)

FIG. 4

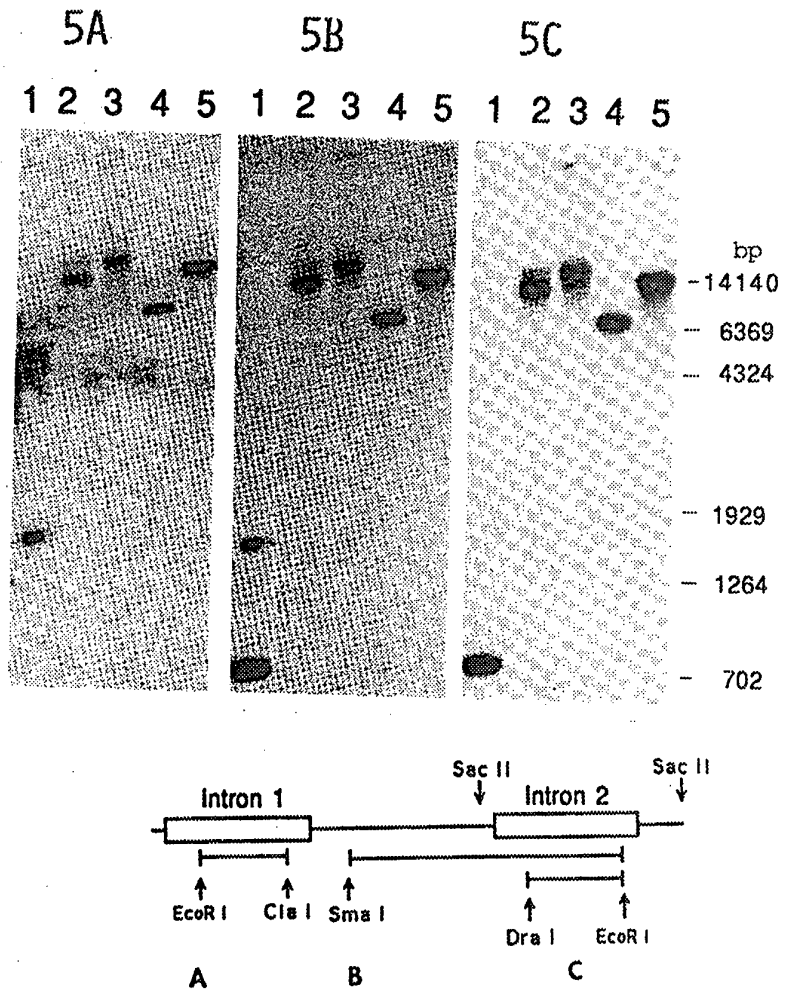
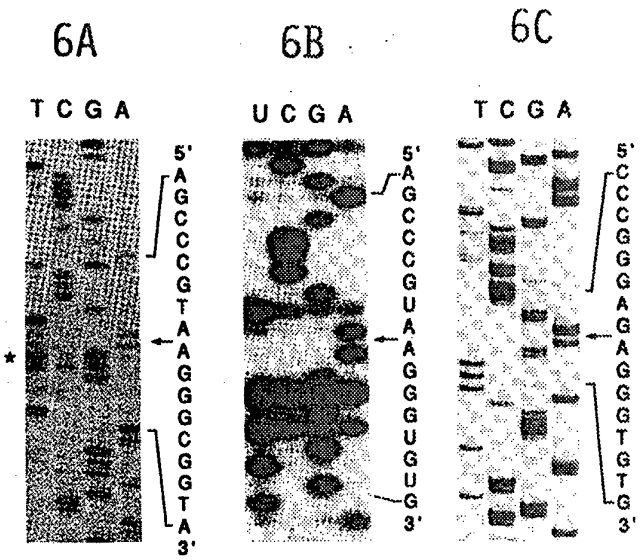

5A
5B
5C
1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
bp
14140
6369
4324
1929
1264
702
Sac II
Sac II
Intron 1
Intron 2
EcoR I
Cla I
Sma I
Dra I
EcoR I
A
B
C

FIG. 5


6A
6B
6C
T C G A
U C G A
T C G A
5' AGCCCCGTAAGGGCGGTA 3'
5' AGCCCCGUAAGGGUGUG 3'
5' CCCGGGAGAGGGTGTG 3'

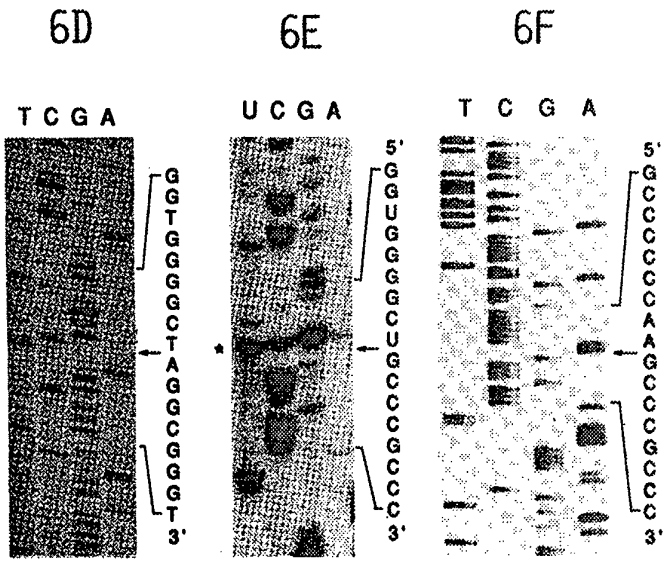

6D
6E
6F
T C G A
U C G A
T C G A
5' GGTGGGGCTAGGCGGGT 3'
5' GGUGGGGCUGCCCCGCCC 3'
5' GCCCCCCCAAGCCCCGCCC 3'

FIG. 6

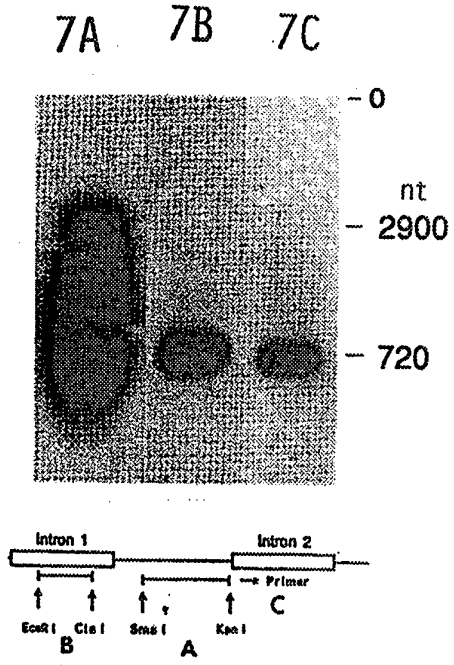

7A
7B
7C
0
nt
2900
720
Intron 1
Intron 2
Primer
C
EcoR I
Cla I
Sma I
Kpn I
B
A

FIG. 7

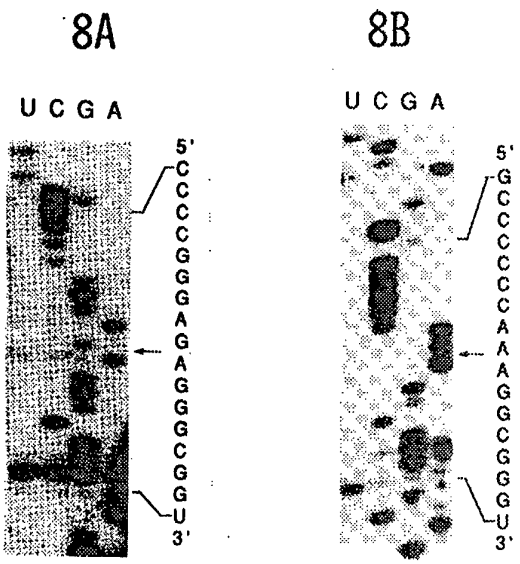
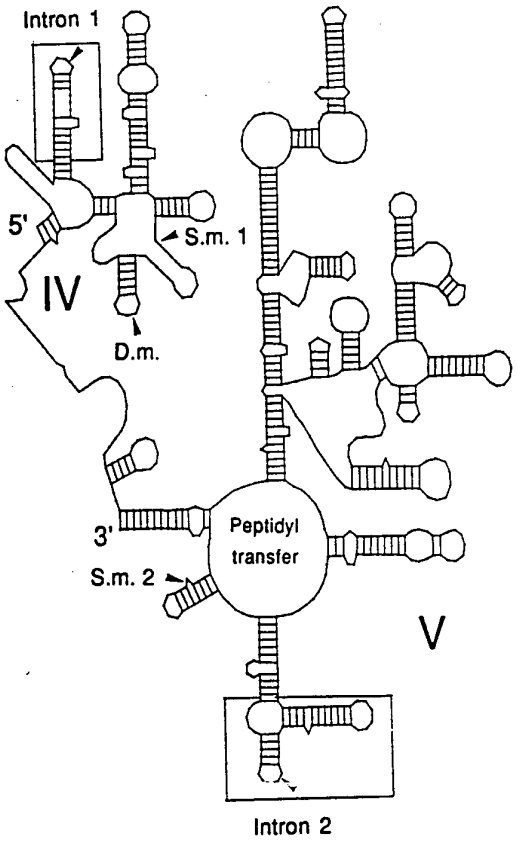
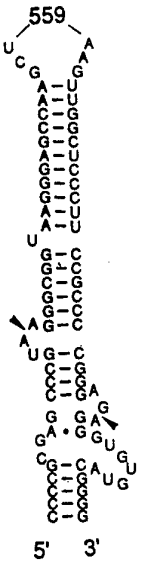

8A
8B
U C G A
U C G A
5'
C C C C G G G A G A G G G C G G U 3'
5'
G C C C C C A A G G C G G G U 3'

FIG. 8

9A


Intron 1
5'
S.m. 1
IV
D.m.
3'
S.m. 2
Peptidyl
transfer
V
Intron 2

9B


559
5'
3'

9C

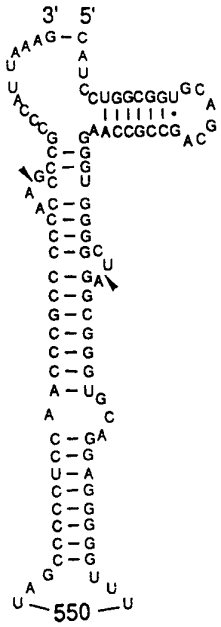

3'
5'
550

FIG. 9

Eukaryotes
Consensus sequence
Archaeal ORF's
Dc.mobilis
Pb.organo- 1
trophum 2
MHNNENVSGISA
VDIFQ
MGCDLE
YLAGFVDGDG
YLLGLIIGDG
YVKALSAFDG
YLAGVVKGDG
GLYKLKYK 78aa
YVDYRVGK 62aa
TLYHNKKA 65aa
WL·GF·DADG
FIKGLYVAEG
LLAAAIDAEG
FVRGLFDSDG
DKTLKRLRIW 80aa
NVKKYRNQPF 69aa
TLYFDRRKRR 76aa

FIG. 10

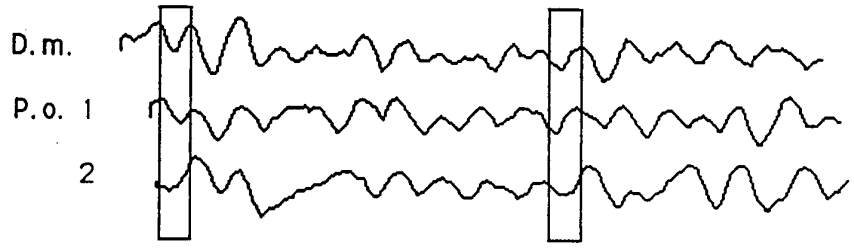
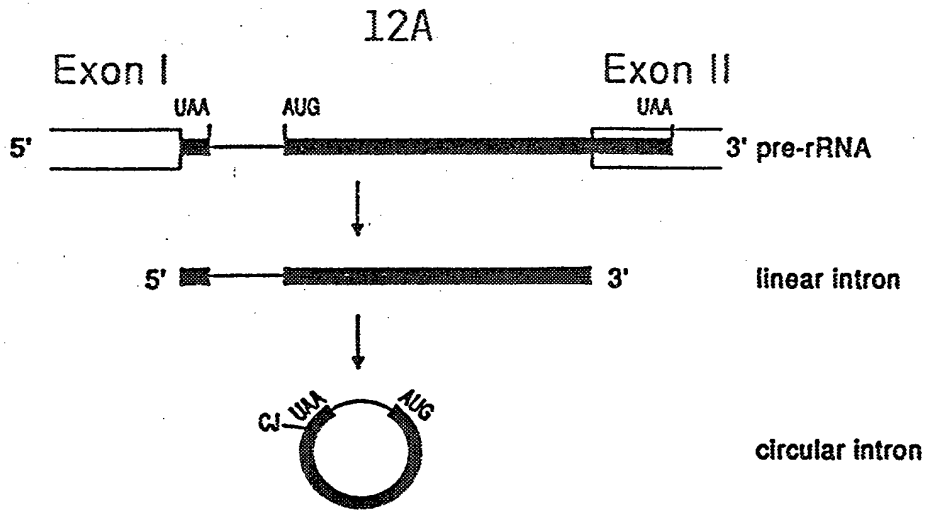
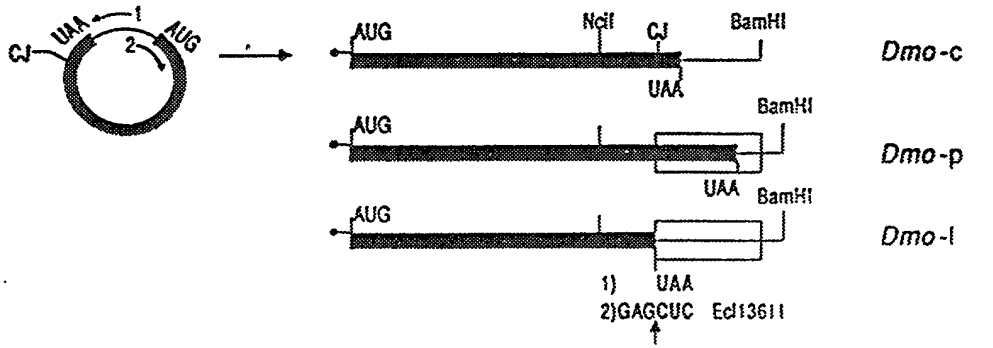


FIG. 11

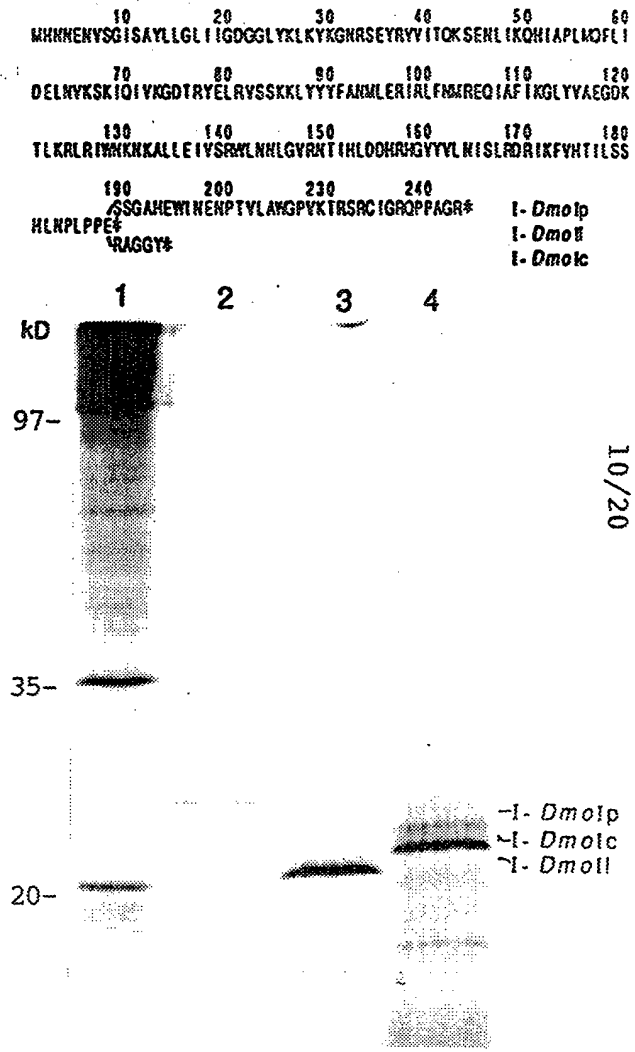
12A



12B



12C



12D

FIG. 12

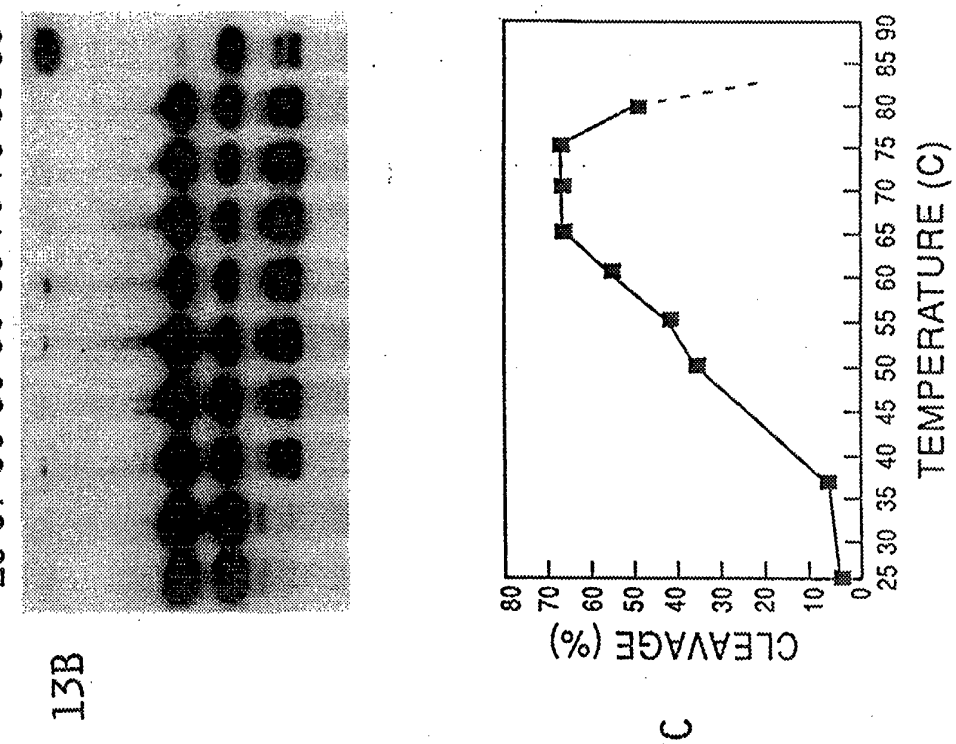
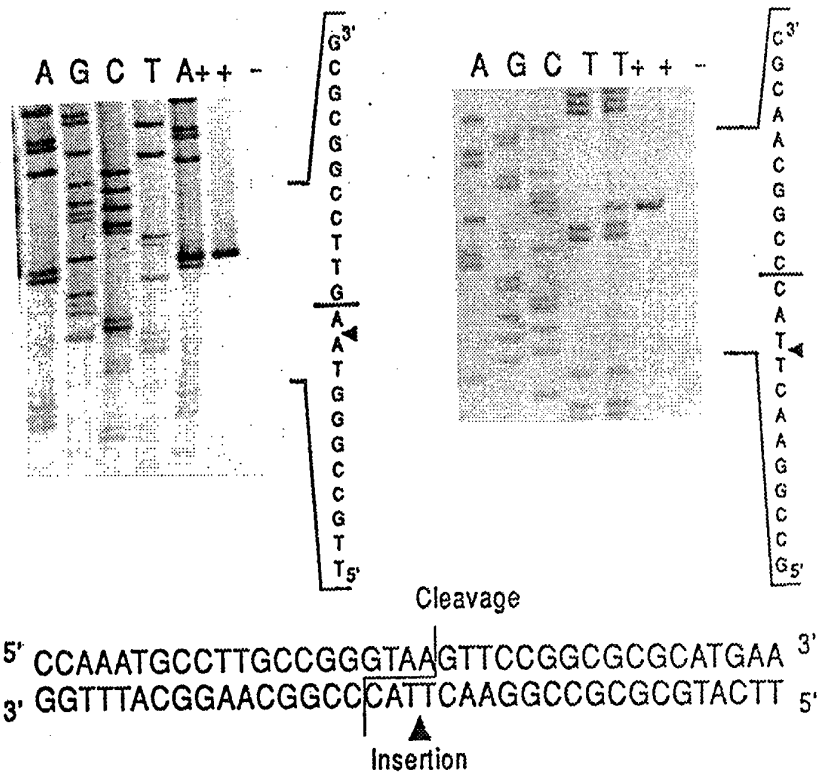

13A
kb
7.2
3.4
1.9
1.5
1 2 3 4
Vector
Target
Products
13B
25 37 50 55 60 65 70 75 80 85°C
kb
7.2
3.4
1.9
1.5
13C
CLEAVAGE (%)
TEMPERATURE (C)

FIG. 13

14A



14B

Property	Bacteria	Archaea	Eukarya
Cellular	No	Yes	Yes
Cleavage	Distal	Proximal	Proximal
Extension type	3' or 5'	3'	3'
Extension length	2-4 nt	4 nt	4 nt
LAGLI-DADG	No	Yes	Yes

FIG. 14

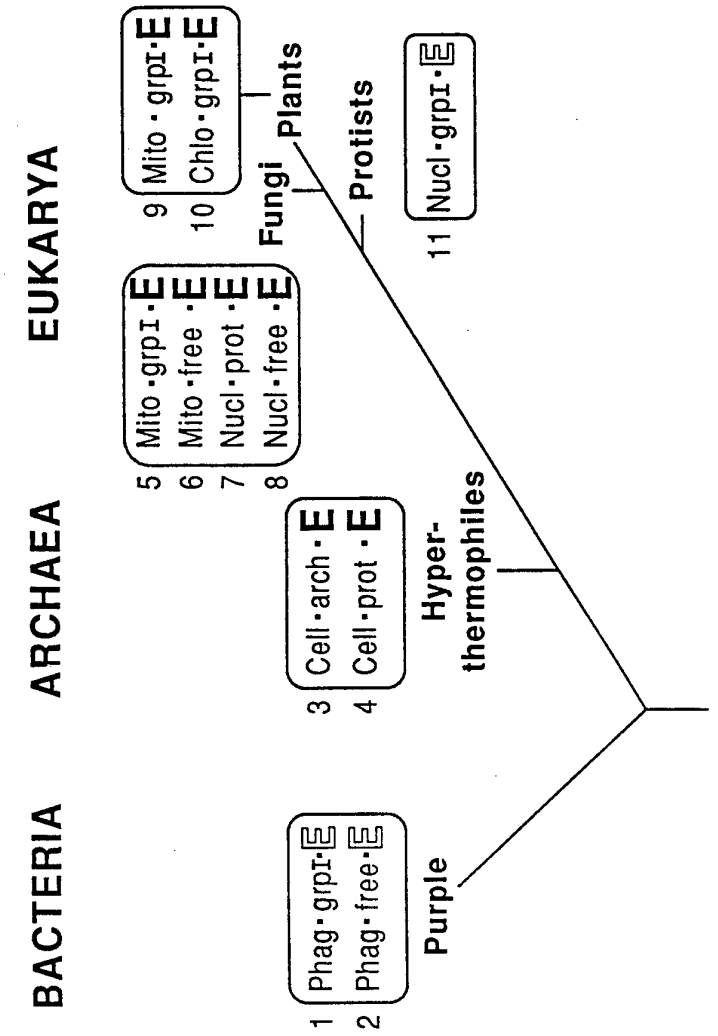

BACTERIA
ARCHAEA
EUKARYA
1 Phag·grpI·E
2 Phag·free·E
3 Cell·arch·E
4 Cell·prot·E
5 Mito·grpI·E
6 Mito·free·E
7 Nucl·prot·E
8 Nucl·free·E
9 Mito·grpI·E
10 Chlo·grpI·E
11 Nucl·grpI·E
Purple
Hyper-thermophiles
Fungi
Plants
Protists

FIG. 15

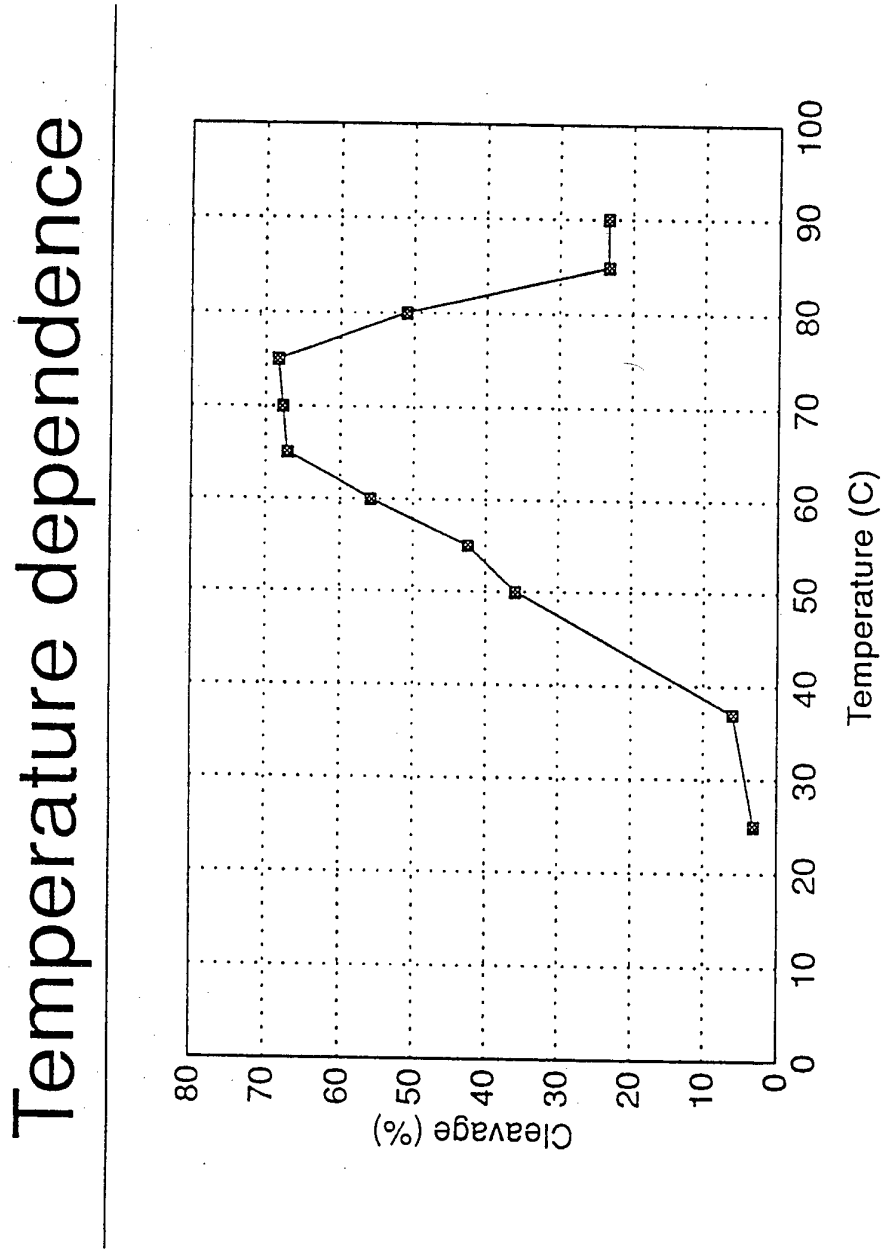

Temperature dependence
Cleavage (%)
0
10
20
30
40
50
60
70
80
0
10
20
30
40
50
60
70
80
90
100
Temperature (C)

FIG. 16

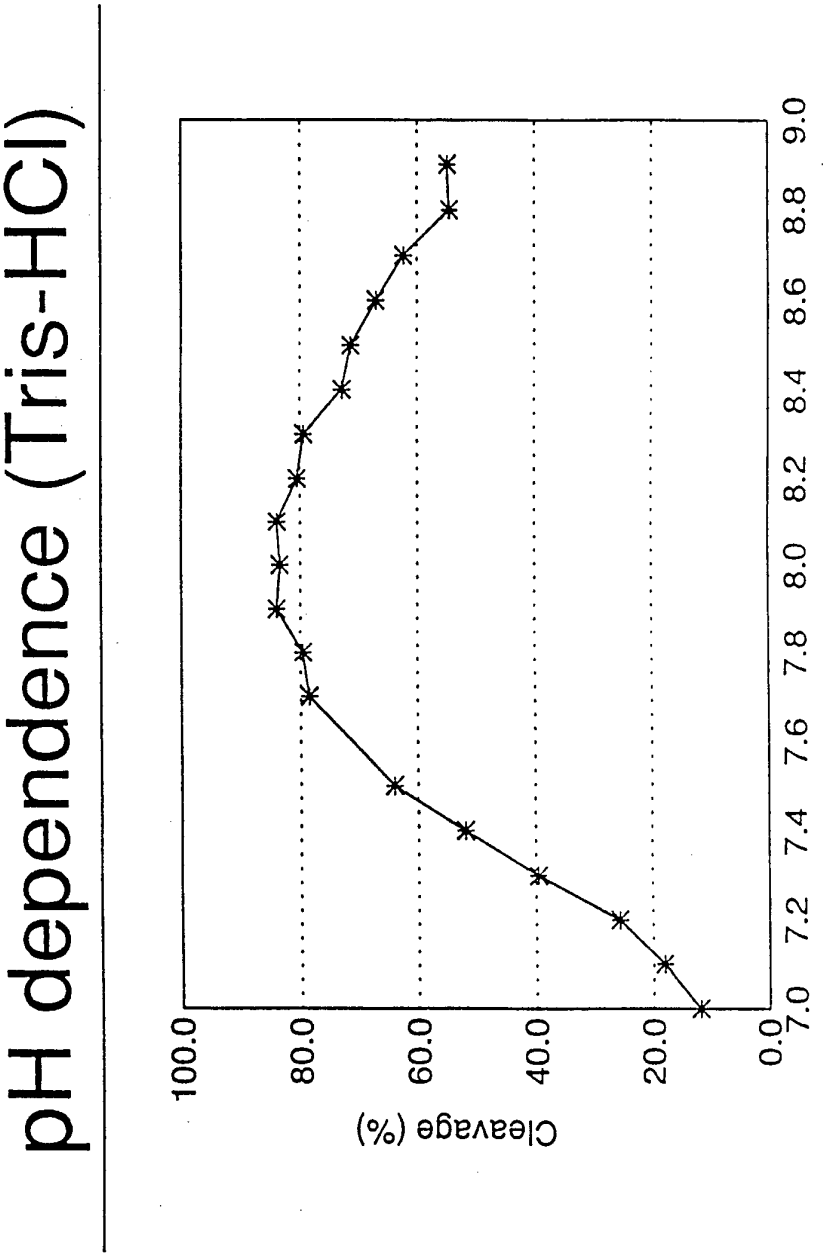

pH dependence (Tris-HCl)
Cleavage (%)
100.0
80.0
60.0
40.0
20.0
0.0
7.0
7.2
7.4
7.6
7.8
8.0
8.2
8.4
8.6
8.8
9.0
pH

FIG. 17

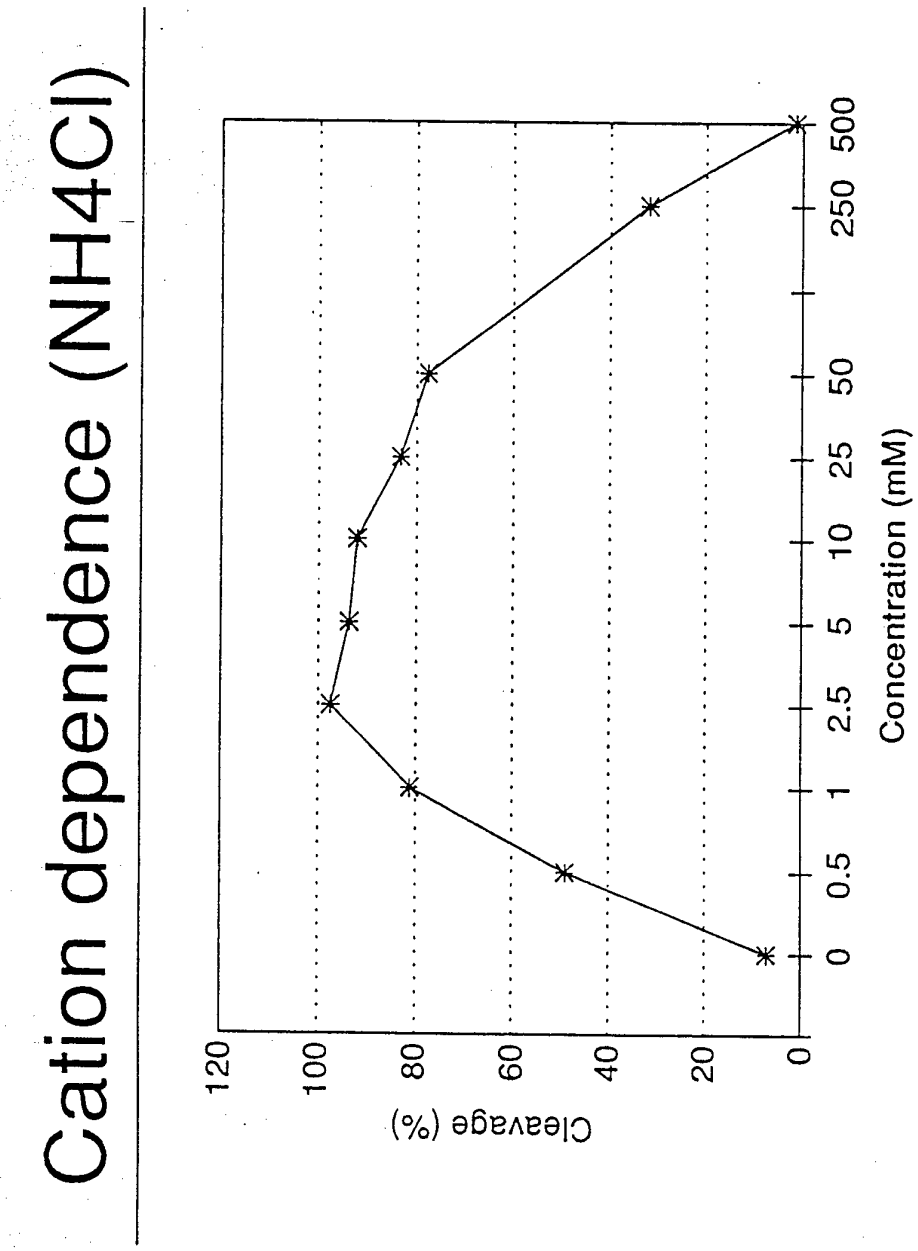

Cation dependence (NH4Cl)
Cleavage (%)
120
100
80
60
40
20
0
0
0.5
1
2.5
5
10
25
50
250
500
Concentration (mM)

FIG. 18

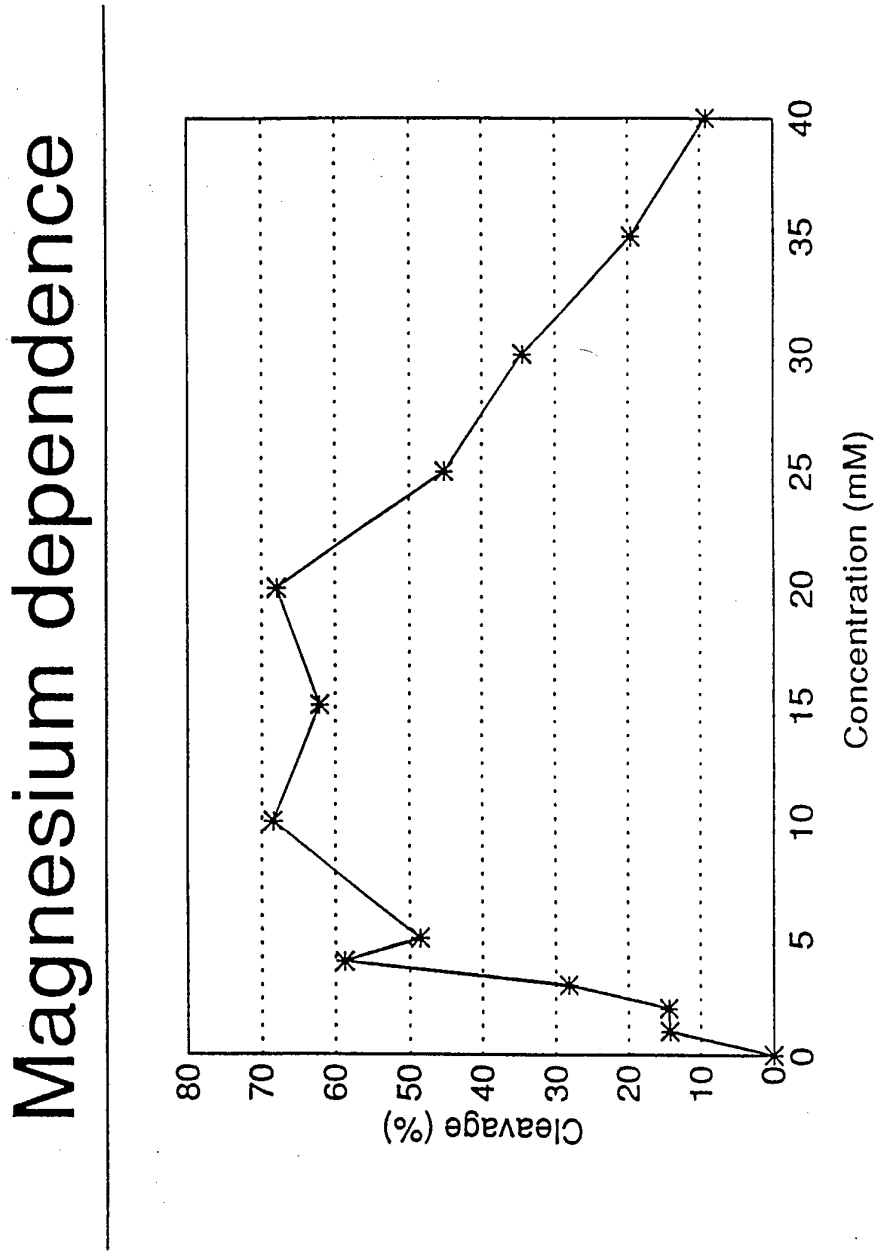
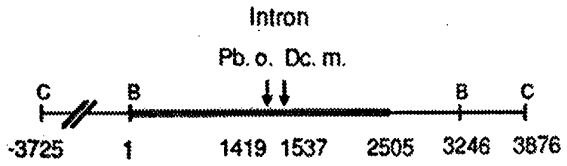

Magnesium dependence
Cleavage (%)
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25
30
35
40
Concentration (mM)

FIG. 19

20A



20B

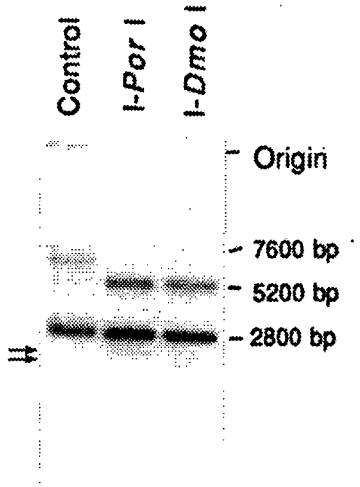
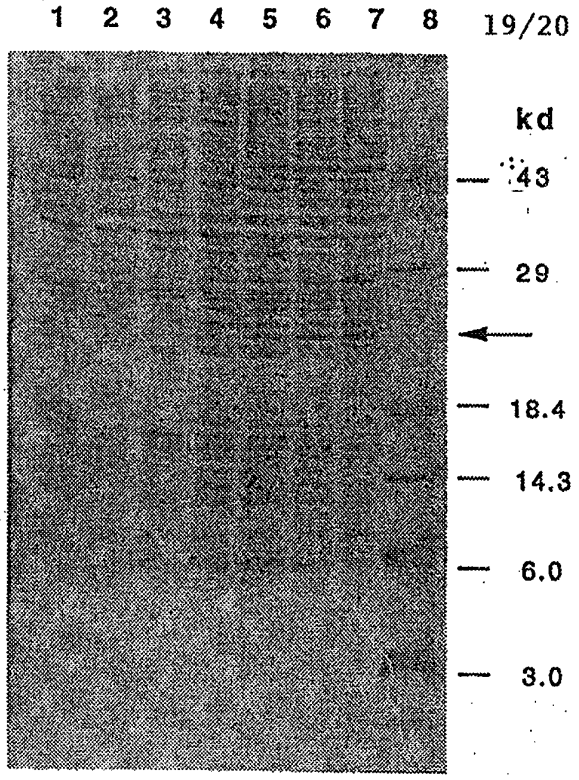
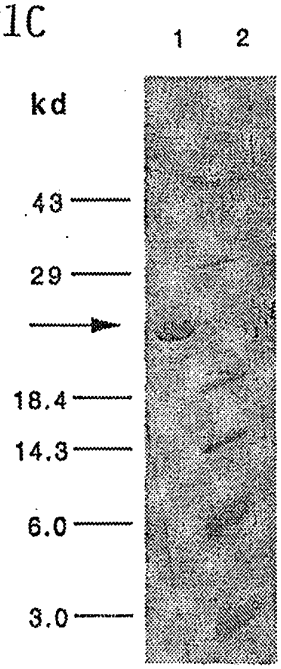


FIG. 20

21A



21C



21B

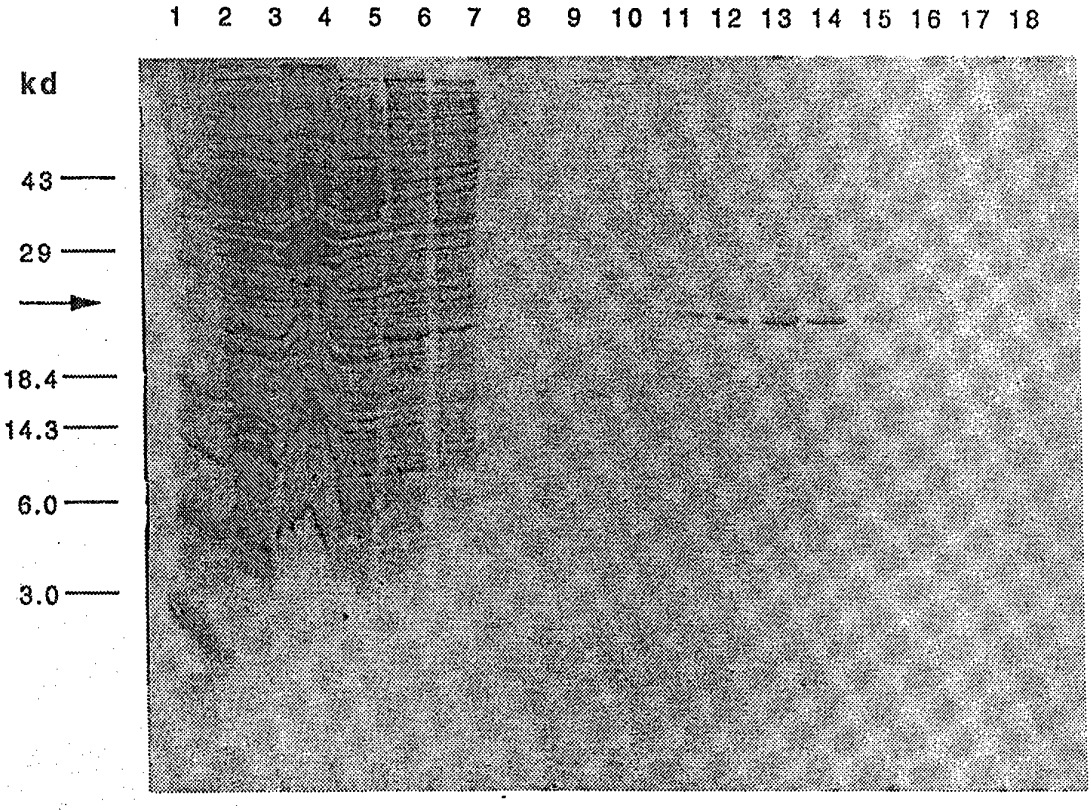


FIG. 21

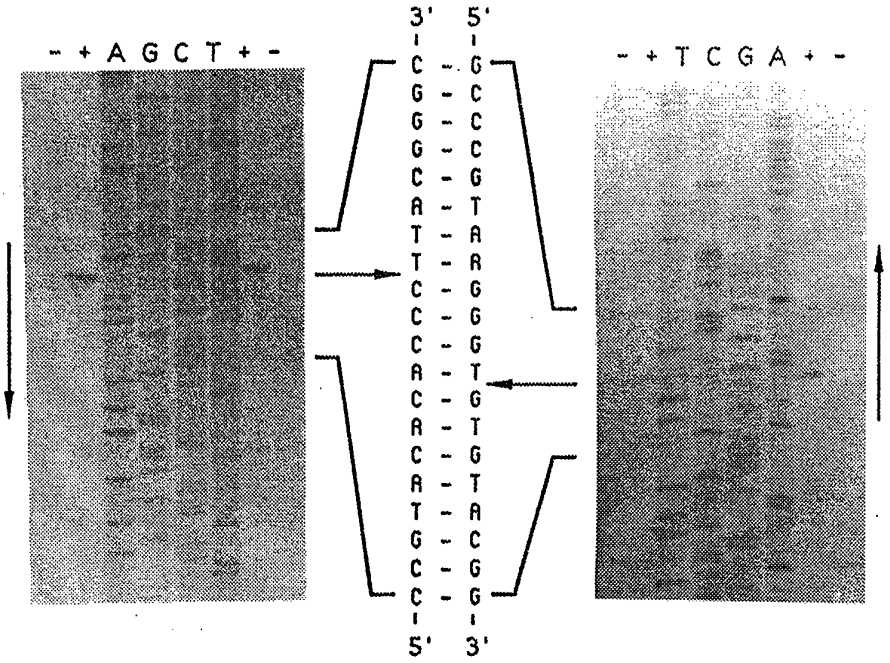

- + A G C T + -
- + T C G A + -
3' 5'
5' 3'
CCCCGCGAGCCCGTAAGGGTGTGTACGGGGGCTGAA
GGGGCGCTCGGGCATTCCCACACATGCCCCCGACTT

FIG. 22

INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK 93/00264

A. CLASSIFICATION OF SUBJECT MATTER

IPC5: C12N 9/22, C12N 15/55
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC5: C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CA, WPI

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Chemical Abstracts, Volume 114, No 19, 13 May 1991 (13.05.91), (Columbus, Ohio, USA), Thompson, L D, "Substrate recognition properties of the tRNATrp intron endonuclease from the archaebacterium Halobacterium volcanii", page 355, THE ABSTRACT No 181107f, Diss. Abstr. Int. B. 1991, 51 (7), 3259 --	1-2
X	Cell., Volume 54, August 1988, J Kjems et al., "Novel splicing Mechanism for the Ribosomal RNA Intron in the Archaebacterium Desulfurococcus mobilis", page 693 - page 703, see fig. 10 and page 11, right column --	1-8

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 November 1993	11 -11- 1993
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Yvonne Sösteen Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK 93/00264

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Nature, Volume 318, December 1985, J Kjems et al., "An intron in the 23S ribosomal RNA gene of the archaebacterium Desulfurococcus mobilis" page 675 - page 677	1-8
A	Patent Abstracts of Japan, Vol 6,No 78, C-102, abstract of JP, A, 57-12994A (MITSUBISHI KASEI KOGYO K.K.), 22 January 1982 (22.01.82)	1-8
A	Can. J. Microbiol., Volume 35, 1989, J Kjems et al., "Comparison of transfer RNA and ribosomal RNA intron splicing in the extreme thermophile and archaebacterium Desulfurococcus mobilis" page 36 - page 42	1-8
A	Chemical Abstracts, Volume 114, No 1, 7 January 1991 (07.01.91), (Columbus, Ohio, USA), Wernette, C M et al., "Purification of a site-specific endonuclease, I-Sce II, encoded by intron 4alpha of the mitochondrial coxI gene of Saccharomyces cerevisiae", page 261, THE ABSTRACT No 2532w, J. Biol. Chem. 1990, 265 (31), 18976-18982	1-8
A	Chemical Abstracts, Volume 114, No 13, 1 April 1991 (01.04.91), (Columbus, Ohio, USA), Thierry, A et al., "Cleavage of yeast and bacteriophage T7 genomes at a single site using the rare cutter endonuclease I-Sce I", page 188, THE ABSTRACT No 116135, Nucleic Acids Res. 1991, 19 (1), 189-190	1-8