Cloning and sequencing of ISC1041 from the archaeon *Sulfolobus solfataricus* MT-4, a new member of the IS30 family of insertion elements

Sergio Ammendola, Laura Politi, Roberto Scandurra*

Dipartimento di Scienze Biochimiche, Università di Roma 'La Sapienza', P.le Aldo Moro 5, 00185 Rome, Italy

Received 19 March 1998; revised version received 21 April 1998

Abstract A genomic fragment containing the insertion sequence ISC1041 has been cloned by PCR from the archaeon *Sulfolobus solfataricus* MT-4, an extremophilic microorganism which grows at 87°C. The 1038 bp ISC1041 element contains an imperfect 18 nt repeat and a long open reading frame which encodes a polypeptide of 311 amino acid residues. The translated amino acid sequence shows a significant similarity to IS30-like transposases. Structural analysis indicates that ISC1041 is a novel member of the IS30 family and displays the DDE motif not previously seen in Archaea. This motif is believed to be involved in the integration mechanism of many mobile elements. As this motif is present in several integrases and transposases which, despite the lack of overall protein homologies, share topological homologies to the DDE motif, a common ancestor has been proposed. The finding of an IS30-like transposase in the archaeal kingdom may have relevance for horizontal gene transfer.

Key words: Archaeon; IS30 family; Transposase; DDE motif

1. Introduction

Bacterial insertion sequences (IS) are genetic units carrying all the information required for transposition function [1], i.e. to translocate segments of DNA to a variety of sites in the chromosome of a host organism. IS play an important role in producing spontaneous mutations through precise DNA excision or by the generation of small, directly repeated duplication of the target DNA at the point of insertion. In addition IS mediate a variety of DNA rearrangements through their ability to generate a new joint between homologous and non-homologous DNA segments [2].

The transposition process is used during a number of biological events: for example the acquisition of bacterial genes for antibiotic resistance, the replication of certain bacteriophages, the integration of retroviruses and the intracellular movements of retroviral-like events [2–5].

Transposition events have similar breakage and joining reactions. Moreover, even though they lack extensive sequence homology, a structural similarity has been found in nucleic acid processing enzymes from very different biological sources [5,6].

The most characteristic of these conserved segments is a triad of invariant carboxylate residues, two aspartate and one glutamate known as the DDE motif, found not only in a large group of transposases (including Mu protein), retroviral integrases and other proteins such as RuvC resolvase and RNase H, but also in transposases of bacterial IS elements [6,7]. Inside IS many families of elements emerge with conserved transposase domains, common structural features and similar functional properties [8].

IS30, which is an insertion element resident in *Escherichia coli* K12 and C [9], represents the prototype of the IS30 family and shares the DDE motif [7]; members of this family have not previously found in Archaea. In the present paper we report the sequence of the insertion sequence ISC1041, found in *Sulfolobus solfataricus* MT4, a novel insertion element belonging to the IS30 family encoding a transposase sharing the DDE motif.

The presence of a transposition protein containing a DDE motif in the archaeal kingdom is further evidence towards a common ancestor and the ability of these mobile elements to pass across natural barriers [7,10,11].

2. Materials and methods

2.1. Chemicals and reagents

All restriction enzymes were purchased from Pharmacia or Gibco-BRL. PCR was performed using the Taq Perkin-Elmer kit. Oligonucleotides used as PCR primers and primers for sequencing reactions were synthesized by Pharmacia.

2.2. Strains and plasmids

Sulfolobus strains were purchased from DSM and were grown following protocols supplied by DSM. *S. solfataricus* MT4, *Pyrococcus furiosus* and *Bacillus acidocaldarius* were the kind gift of Dr. M. De Rosa (University of Naples, Italy). *Thermococcus* strain TY was the kind gift of Dr. F. Canganella (University of Viterbo, Italy). *Bacteroides fragilis* DNA and plasmid pVHO1 were supplied by Dr. A. Pantosti (Istituto Superiore di Sanità, Rome). pUC19 vector DNA was used for cloning and sequencing experiments.

2.3. PCR

To clone the structural gene encoding glutamate dehydrogenase (GDH) from *S. solfataricus* MT4 we designed the following degenerate nucleotides on the basis of *S. solfataricus* GDH amino acid sequence [12]: SSMET, CATGGAATTCATGGAGGA(G/A)GT(G/A)(T/C)T(A/G)TC, and SSGDH, CGACTCTAGATTA(T/C)A(A/G)TAT(T/C)CCTC(T/G)TGCTTTCAT.

These primers incorporated *Eco*RI and *Xba*I site sequences to facilitate subcloning of the PCR product in pUC19. The annealing step in the PCR reaction was performed at 42°C, since higher temperatures were not effective. The *Eco*RI and *Xba*I digested product was electroeluted and ligated into pUC19 vector (*Eco*RI/*Xba*I) and transformed into *E. coli* DH5α; transformants were analyzed by restriction digestion. Two different clones containing similarly sized DNA were ob-

*Corresponding author. Fax: (39) (6) 4440062.
E-mail: scandurra@axcasp.caspur.it

Abbreviations: ISC, insertion sequence of Crenoarcheota; PPE, proximal promoter element; DPE, distal promoter element

Dedicated to Prof. A. Ballio on the occasion of his 75th birthday.

The nucleotide sequence data reported in this paper will appear in the GenBank Nucleotide Sequence Databases under accession number U85710.

PII S0014-5793(98)00530-4

tained (pUC19–27 and pUC19–22). These clones were sequenced using the Sequenase 2.0 kit (Amersham) with α-^{35}S-dATP (600 Ci/mmol) purchased from NEN-Dupont. The DNA sequence revealed that one clone corresponded to the expected gene of GDH (pUC19–27). However, on comparison of the largest open reading frame found in the other clone, pUC19–22, with the data base, significant similarity was obtained with a transposase belonging to the IS30 family.

2.4. Analysis of the genomic distribution of ISC1041

Total DNA was extracted from disrupted cells of microorganisms and digested with *Eco*RI. Digested genome DNA was subjected to electrophoresis in 0.7% agarose gel and transferred to nylon membrane (Amersham Hybond N$^+$) according to the protocol described by Sambrook et al. [13]. An *Eco*RI-*Xba*I fragment excised from pUC19.22 was randomly labeled with α-^{32}P-dATP (3000 Ci/mmol), purchased from NEN-Dupont. Low stringency hybridization of the membrane was carried out at 5×SSC, 42°C. More stringent conditions were obtained by washing the filter for 2 h in 0.2×SSC, 0.1% SDS at 70°C.

2.5. Multiple alignments

Amino acid sequences were compared with the PROSITE data base [14], SSEARCH [15] was used to search a non-redundant sequence data base obtained combining SwissProt, PIR and the coding regions from GenBank. Multiple alignment was obtained firstly using Maxhom [16] and subsequently refined using an hidden Markov model [17]. The final multiple alignment was used as an input for the PHD secondary structure prediction program [18]. The alignment of DDE motifs of RSV, HIV-1 and MuA integrase with IS30 and ISC1041 reported in Fig. 5 was performed manually but according to Kulkosky et al. [19] and Rezsöhazy et al. [8]. The alignment of inverted repeats, right and left of IS30, ISC1041 and IS4351 reported in Fig. 6, was also performed manually according to Dalrymple et al. [9].

The CodonPreference program [20] was used to detect open reading frames using the high A+T content characteristic of the third position of codons of *Sulfolobus* genes.

2.6. Phylogenetic analyses

Unrooted phylogenetic trees were constructed by distance matrix and maximum parsimony from selected positions using the programs CONSENSE, NEIGHBOR, PROTDIST, PROTPARS and SEQBOOT implemented in the Phylogeny Inference Package (PHYLIP) version 3.5c [21]. Bootstrap confidence levels attached to internal tree branches are based on 100 pseudoreplicates of the original alignment.

```
1          10         20         30         40         50         60
TCTAGATTTGGCTGCCCCGTACGGATAATTGGGTTTTGAGTAAAGTCCCATGAGGTTTCG
TATGAATCCCAATCATTCTTAGAAATATAAATTAACTTTTTTTGACCCACTTGATTAGTT
GTATCTGGGCAGAGTAAAACTTGAAGTGCGACATAACCACCTAATTAATTTAAAGGGTT
TATGGAGTATATAAAAGTTTTCATACCATCATCTTAACTTTGAAGATCGATGCTCATTA
                  A                                   B

240 ATG CTT GAG TCA AGA AAA GAA GGC TTT TCA GCC AGA AAA TTT GCT GAA CTC ATT AAA AGA
     M  L  E  S  R  K  E  G  F  S  A  R  K  F  A  E  L  I  K  R
300 CAT CCT AGT ACC GTA ATC TAT CGT GAG CTT AAA AGA AAT AGC ATC AAT GAC GTT TAT CTT
     H  P  S  T  V  I  Y  R  E  L  K  R  N  S  I  N  D  V  Y  L
360 GCT CGA TAT GCT TCT GAT AAC ACC TTC GCT AGA CTA GAC TGT GGC ACA GAA AAC TCA AAA
     A  R  Y  A  S  D  N  T  F  A  R  L  D  C  G  T  E  N  S  K
420 TGG ATT CAA TCC TCT GGA AAT TTA TTG TTG TTA AAG GCG ATC CCG TGG TTT ATG GGC CTC
     S  I  Q  S  S  G  N  L  L  L  L  K  A  I  P  W  F  M  G  L
480 TCA GGA AAT ACG GAA ACT TTA TTA ACA TGC AGA GTA CAT GTG ACC AAA GAA TGG GAG CCG
     S  G  N  T  E  T  L  L  T  C  R  V  H  V  T  K  E  W  E  P
540 TTG AAA AAA GAC TTA TTA TCC TGT CTG CGT CAT GTA AAA AAG GCG GTA AAA CGT AAC TGG
     L  K  K  D  L  L  S  C  L  R  H  V  K  K  A  V  K  R  N  W
600 AAG AAC CTA AAA AAG ATT CTA TAT TAC AGG ATT AAT AAC TAT TTC AGA TGC GAT GGA CCC
     K  N  L  K  K  I  L  Y  Y  R  I  N  N  Y  F  R  C  D  G  P
660 AGC CGA AGT TCA AGA AAG AAA AAT ACC GGA CAT TGG GAA CGT GAT TTA ATT AAA TGG AAA
     S  R  S  S  R  K  K  N  T  G  H  W  E  R  D  L  I  K  W  K
720 GAC AAT AAA AGT TCG ATA CGA ACA TTT ATT GAC CTT AAG ATA CAC GGC TCT GTA TCT TCG
     D  N  K  S  S  I  R  T  F  I  D  L  K  I  H  G  S  V  S  S
780 GAA CAT TAC CTG ATG GCA AAA GGC AGA ATA CGT GGA AGG GCT TTA ACT GAA GCT CTG AAA
     E  H  Y  L  M  A  K  G  R  I  R  G  R  A  L  T  E  A  L  K
840 TAT TTA CCT GCA GAA CGT GCT AAA ACG TTG ACC TAT GAC CGT GGA CGC GAG ATG GCA GAA
     Y  L  P  A  E  R  A  K  T  L  T  Y  D  R  G  R  E  M  A  E
900 CAT AAA ATA CTT GAA GAA GAT TTA GGC ATA GAT GTA TAT TTC TGT GAC CCA CAT TCA CCC
     H  K  I  L  E  E  D  L  G  I  D  V  Y  F  C  D  P  H  S  P
960 TGG CAA AAA GGC ACA TGC GAA AAT ATG AAT GGT TTA ATT AGG CAA TAT TTA CCT AAA GGG
     W  Q  K  G  T  C  E  N  M  N  G  L  I  R  Q  Y  L  P  K  G
1020 ATT GAT TTA AAT CAG GCA GAT CAG CAT TAT TTA AAT CAA GTT GCC ATG TCA CTG AAT ACT
     I  D  L  N  Q  A  D  Q  H  Y  L  N  Q  V  A  M  S  L  N  T
1080 CGT CCT AGA AAA GCG TTA GTT GGC TTC ACC ATT GAG AAA TTT GCT CAG CTT GTT GTT ATC
     R  P  R  K  A  L  V  G  F  T  I  E  K  F  A  Q  L  V  V  I
1140 ATA AGA CTT TTC AAA CTG TCG CAC CTC ATG TTT TGA ATTC
     I  R  L  F  K  L  S  H  L  M  F  -
```

Fig. 1. Complete nucleotide sequence of the 1179 bp insertion element ISC1041 from *S. solfataricus* MT4 and the translated amino acid sequence of the transposase (from position 240 to 1173). Left inverted repeat (from position 136 to 153) and right inverted repeat (from position 1156 to 1173) are reported in bold. The putative promoter elements A (distal or DPE) and B (proximal or PPE) are underlined. Sequence shown is 5′→3′.

3. Results

3.1. DNA analysis of insertion sequence

The plasmid pUC19–22 contained a DNA fragment of 1179 bp including an insertion element of 1038 bp (Fig. 1). This insertion sequence begins at position 136 with a 18 nt imperfect inverted repeat (IR-L) and stops at 1173 bp after a 18 nt imperfect inverted repeat (IR-R). IR-R differs from IR-L in two mismatches found in positions 1164 and 1168. Before nucleotide 136 we found the sequence GAGT: the same se-

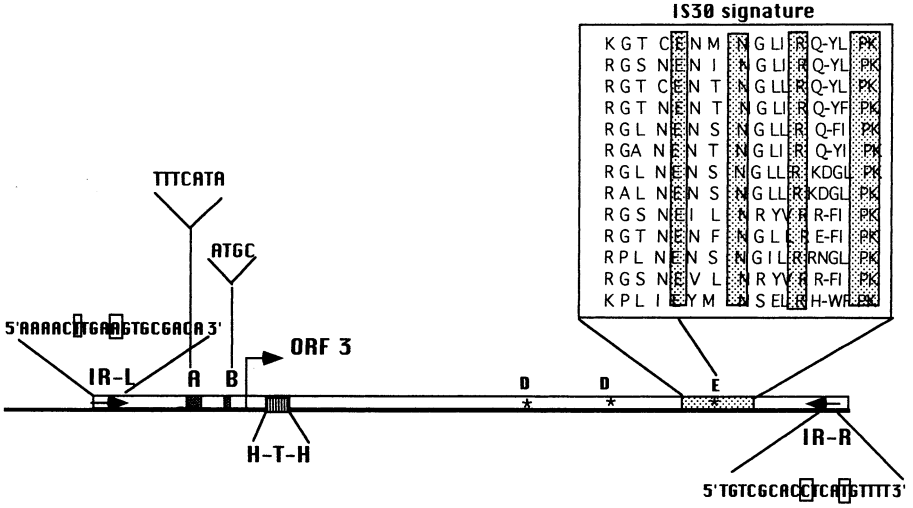

Fig. 2. Schematic representation of the cloned DNA fragment from *S. solfataricus* carrying the insertion element ISC1041. The insertion element is delimited by two arrows representing the left and right inverted repeats (IR-L and IR-R) whose nucleotide sequences are reported, boxing the mismatched nucleotides. The nucleotide sequences of the putative promoter elements are shown as A and B. The ORF3 encoding the transposase of ISC1041 includes: (i) a putative DNA binding helix-turn-helix motif (H-T-H); (ii) the amino acid residues of the DDE motif, highlighted by stars; (iii) the amino acid sequence corresponding to the IS30 signature presented in single letter amino acid code; the sequences reported refer to the multiple alignment and in the same order as in Fig. 4. The invariant residues are shaded.

quence is found at the end of the ISC1041 element (1173) with a single mismatch, GAAT instead of GAGT. The mismatch A→G is probably due to the primer SSMET used for PCR, the sequence GAGT could be the duplication target site. The start codon ATG of the gene encoding the transposase (ORF3) was found at position 240 and the putative archaeal promoter is characterized by box B (proximal promoter element, PPE) from positions 230 to 233, and box A (distal promoter element, DPE) from positions 199 to 205 [22]. The stop codon TGA was found at position 1173. On the strand complementary to the ORF3 sequence no significant open reading frames were found.

A schematic representation of ISC1041 is presented in Fig. 2.

3.2. Analysis of the genomic distribution of ISC1041

In order to ascertain the presence of similar insertion elements within the genome of several thermophilic and mesophilic microorganisms, we used the Southern blotting technique. The *Eco*RI-*Xba*I pUC19.22 insert was used to probe genomic DNA from *Bacillus acidocaldarius, Pyrococcus furiosus, Thermococcus* strain TY, plasmid pVH01 carrying transposon Tn*4351* from *B. fragilis* and the DNA from several strains of *Sulfolobus*. Under high stringent conditions no hybridization was obtained with the DNA of these microorganisms or with the following *Sulfolobus* strains: *S. solfataricus* (DSM 1616), *S. acidocaldarius* (DSM639), *S. shibatae* (DSM5389).

In low stringency conditions some positive signals were observed, as reported in Fig. 3A.

In higher stringent conditions (5×SSC, 65°C) (Fig. 3B), hybridization of total genomic DNA from *S. solfataricus* MT4 with total PCR product revealed, after a 20 day exposure, two significant bands with apparent weights of approximately 3 kb and 2 kb (lane 1). Under identical experimental conditions the same filter hybridized with the pUC19.27 insert, revealing, after overnight exposure, one band of about 3 kb (lane 2), whereas hybridization with the pUC19.22 insert revealed one band of about 2 kb (lane 3).

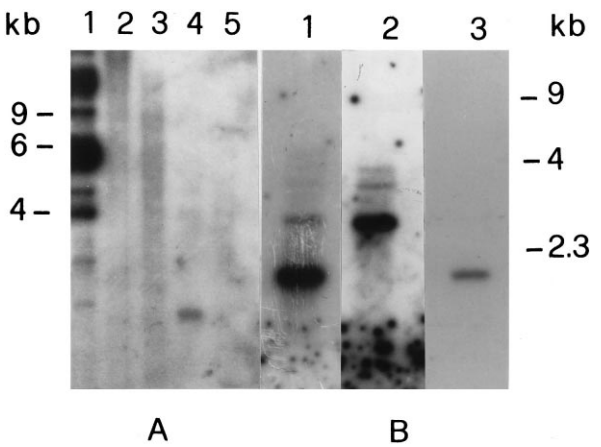

Fig. 3. Southern blots. A: The following genome DNAs were hybridized at 5×SSC, 42°C, using as a probe the pUC19–22 insert. Lane 1: pVH01, uncut; lane 2: *Bacillus acidocaldarius*; lane 3: *Pyrococcus furiosus*; lane 4: *Sulfolobus solfataricus* MT4; lane 5: *Thermococcus* strain TY. B: Hybridization at 5×SSC, 65°C of total genomic DNA from *S. solfataricus* MT4 with total PCR product (lane 1); with the pUC19–27 insert in the same filter and in the same conditions (lane 2) and with the pUC19–22 insert in the same filter and in the same conditions (lane 3).

3.3. Transposase structure analysis

The translated sequence of ORF3 encodes a polypeptide of 36 200 Da. The deduced amino acid sequence was compared with the PROSITE pattern data base which revealed the presence of the signature of a transposase belonging to the IS30 family, i.e. R-G-X(2)-E-X(2)-N-G-[LIVM](2)-R-[QE]-[LIVMFY](2)-P-K, with a single mismatch consisting of an Arg to Lys substitution (as reported in Fig. 4, highlighted by stars). Most of the proteins identified by this method are 310–380 amino acids long and are conserved in their C-terminal region. Fig. 4 shows a multiple alignment of the protein which highlights the presence of the helix-turn-helix motif suggested by Stalder [23], with a score of 0.82 in position 12–31 of the IS30 sequence. Our alignment, performed using the PHD prediction program and with a standard deviation of 1.8, confirms the presence of a helix-turn-helix motif in all the IS30-like transposases but shifted between positions 30 and 50.

Our results suggest that the ORF3 polypeptide is a novel transposase of the IS30 family. The level of sequence identity with the transposase of *E. coli* IS30 (about 35%) indicates that these proteins share a similar fold, which we suggest is mostly helical (see Fig. 4) [24].

Comparison of the sequence of the transposase encoded in ISC1041 with the sequences of the transposases of the IS30 family reported in Fig. 4 shows the presence of the helix-turn-helix motif (between positions 30 an 50) and the signature of the IS30 family (between positions 319 and 336, highlighted by stars). Furthermore, the adopted alignment based on the IS30 of *E. coli* allows comparison of the ISC1041 sequence with some integrases or transposases of other families such as HIV-1, RSV and MuA (Fig. 5).

This alignment was done according to Kulkosky et al. [19] and Rezsöhazy et al. [8] who found highly conserved amino acids among prokaryotic and eukaryotic mobile elements. These conserved amino acids belong to a domain where the DDE motif is found at positions 155, 213 and 247 of ISC1041 (Fig. 5). The DDE motif is defined as an amino acid sequence in which D is separated from D by about 58 amino acids and D from E by about 35 amino acids, and is found in many recombinases such as avian sarcoma virus integrase, HIV integrase, Mu transposase, ribonuclease H, RuvC resolvase and the Holliday junction resolving enzyme. Despite the lack of overall protein similarities, all these recombinases share the topological homologies of the DDE motif [5,6,25,26].

In addition a large conserved sequence (called the 'integrase domain') spanning over 90 amino acids was found inside ISC1041 from Lys-147 to Lys-260; this domain is conserved in transposases of several IS30-related elements, in transposases of the IS3, IS15 and IS4 families and in the integrases of many retroelements [8]. This integrase domain together with the DDE signature may suggest a common step in the transposition mechanism of otherwise unrelated mobile elements.

If the inverted repeats IR-L and IR-R are aligned with IR of related IS30-like transposases (IS30, IS4351), a number of similarities can be identified (Fig. 6), in particular dinucleotides AA, GC and AC, which have been demonstrated to be involved in binding of transposases to DNA [23], and this illustrates a potentially similar catalytic mechanism of these enzymes.

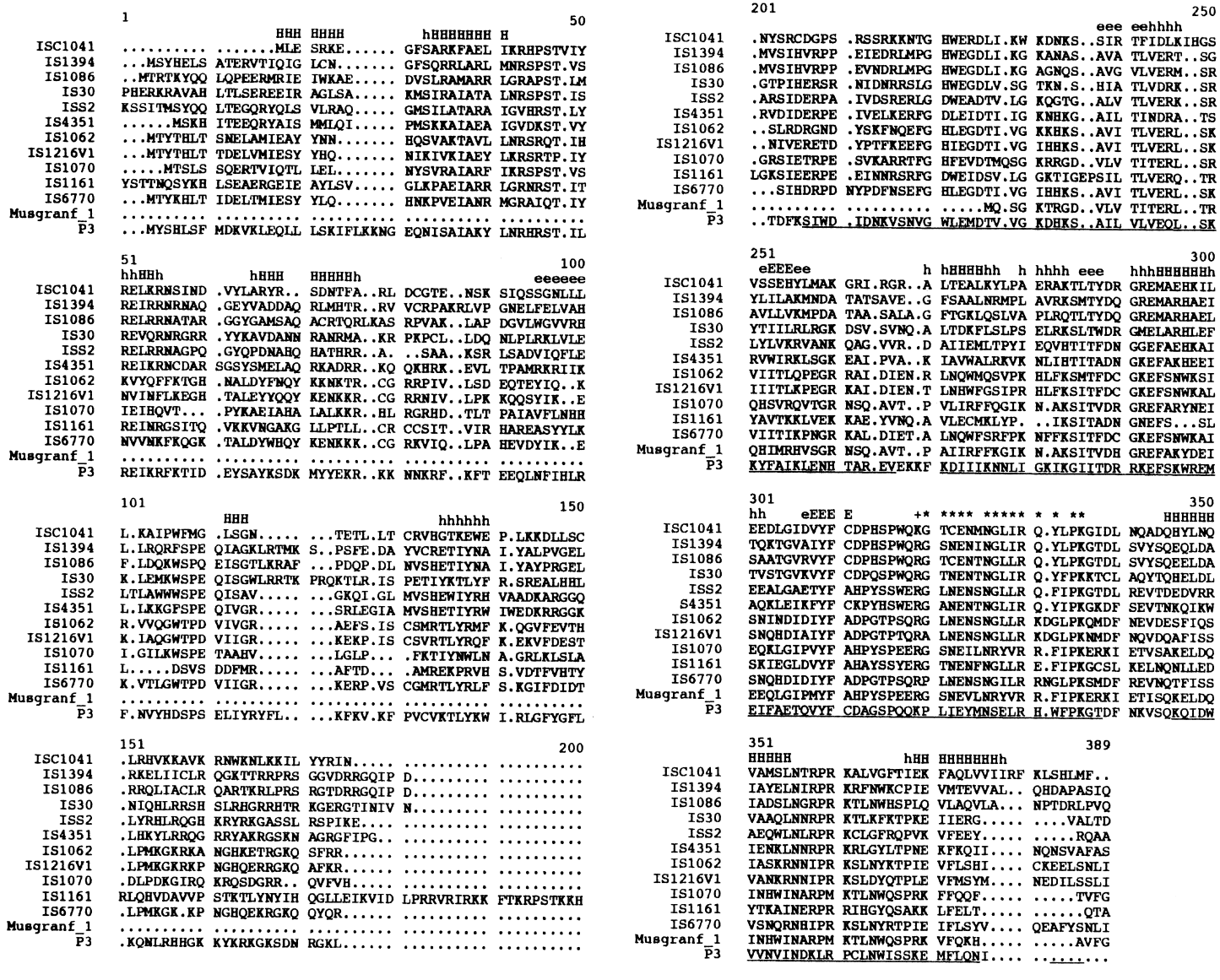

Fig. 4. Multiple alignment of the primary structures of the transposases. ISC1041 from *S. solfataricus* MT4; IS1394 from *Pseudomonas alcaligenes*; IS1086 from *Alcaligenes eutrophus*; IS30 from *E. coli*; ISS2 from *Aeromonas salmonicida*; IS4351 from *Bacteroides fragilis*; IS1062 from *Enterococcus faecalis*; IS1216V1 from *Enterococcus faecium*; IS1070 from *Leuconostoc lactis*; IS1161 from *Streptococcus salivarius*;IS6770 from *Enterococcus faecalis*; Musgranf_1 = Mouse granuloma associated factor; P3 protein encoded by gene 3 from *Spiroplasma citri* phage SPV1-R8A2B. IS1086, IS30, IS4351 and P3 sequences are from SwissProt, while others are from GenBank. The symbols 'h' and 'e' represent predicted helices and strands, respectively. Upper case symbols denote a prediction accuracy above 5 as defined by PHD [16]. Residues matching the PROSITE pattern for transposase sequence are denoted with *; + denotes the single mismatch (K instead of R). The large conserved sequence spanning over 90 amino acids between positions 147 and 260 is the 'integrase domain' [8]. The first 48 amino acids of IS30 are not shown. The continuous line at the bottom of the lanes indicates the sites of the first 11 IS sequences used for the construction of phylogenetic trees.

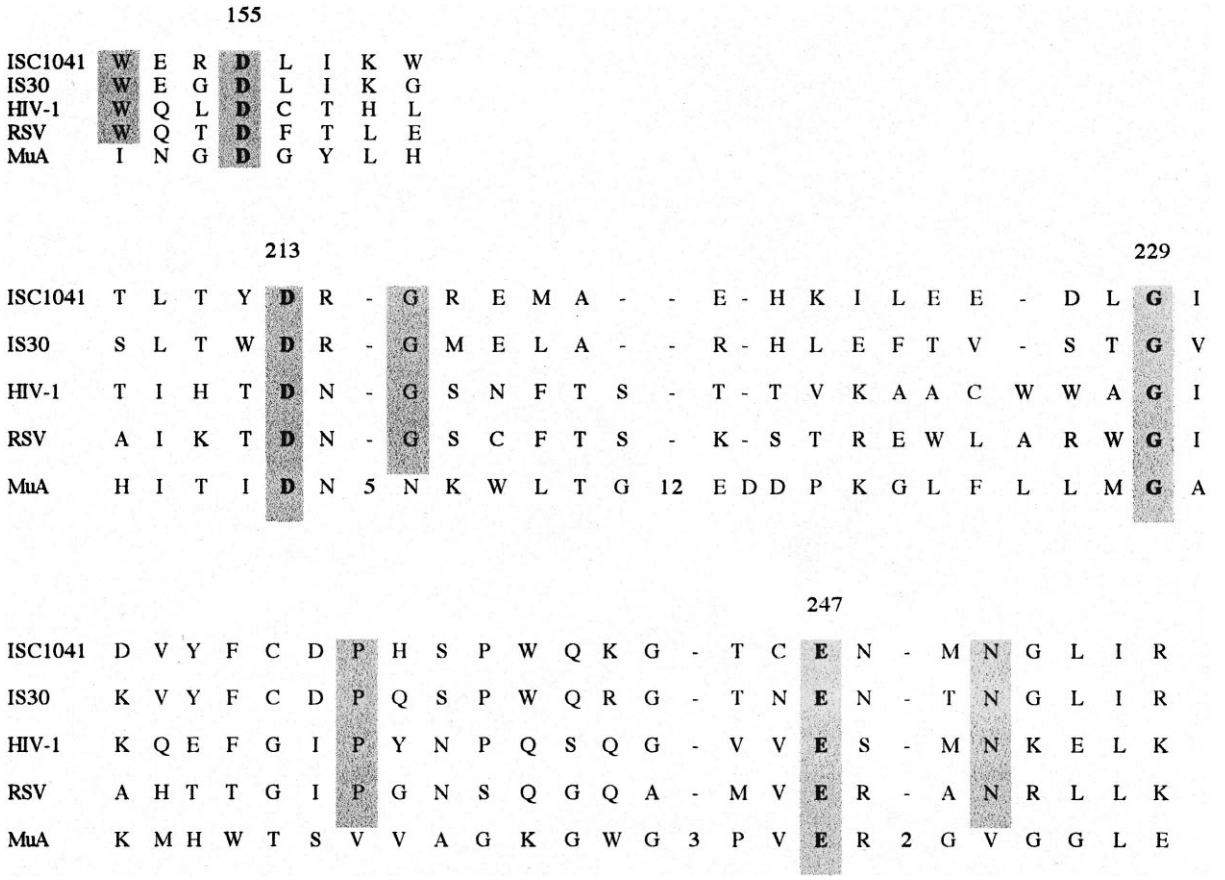

Fig. 5. Manual multiple alignment of the DDE motif of the crystallized HIV-1 integrase, RSV integrase and MuA transposase with ISC1041 and IS30. Manual multialignment was performed following Kulkosky et al. [19]. Invariant residues are shaded. Positions are ISC1041 numbering.

Finally we want to point out that the IS30-like elements are highly conserved among unicellular organisms. As shown in Fig. 4 some similarity has been found between IS30-like transposases and mouse granuloma factor. Interestingly in the murine plasmacytomas and precursor cells an illegitimate translocation t(12:15) of the c-*myc* gene has been found [27]. If contamination by mycoplasma DNA is excluded and if the section of mouse granuloma factor which shows similarity to IS30 transposases is contained within the catalytic domain, then mouse granuloma factor could be involved in the translocation t(12:15).

Phylogenetic trees of the transposases encoded by IS reported in Fig. 4 were calculated from 163 sites in selected positions of the multiple alignment using both distance matrix (Fig. 7A) and maximum parsimony (Fig. 7B) methods. By both criteria the sequence ISC1041 is firmly associated with the grouping comprising IS1086, IS30 and IS1394 and is separated from other IS clusters by a statistically well supported node (80–97 bootstrap confidence level, BCL). However, given the low BCL attached to the internal node joining ISC1041 and IS30, the relative branching order of these two IS sequences within the grouping cannot be established. In fact all three alternative topologies shown in Fig. 7C are equally likely.

4. Discussion

PCR mediated cloning of the GDH encoding gene was successful only at low stringency conditions. This result was due not only to degeneracy of primers but also to DNA structures. PCR was then unspecific (Fig. 3A) and oligonucleotides were able to prime other DNA fragments such as that containing the interesting insertion element ISC1041 belonging to the IS30 family. The primary structure of the cloned DNA showed that SSMET could prime at the 3′ end of ISC1041 whereas SSGDH primed at the 5′ end. Since in *S. solfataricus* only one transposase has been characterized [28], although the *S. solfataricus* genome is known to be rich in mobile elements [29], the availability of the primary structure of an archaeal IS30-like transposase was considered useful to

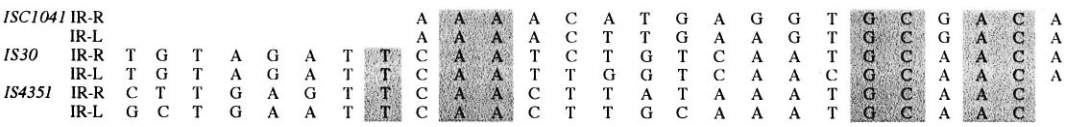

Fig. 6. Nucleotide sequence of the IR-L and IR-R of ISC1041, IS30 and IS4351. Alignment was performed manually. Invariant nucleotides are shaded.

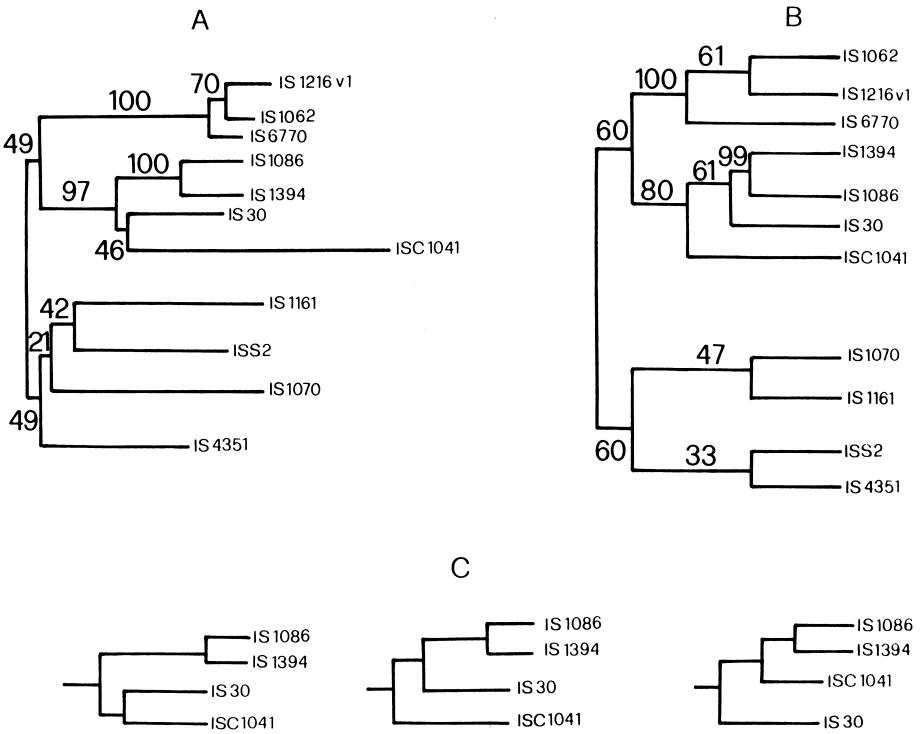

Fig. 7. Evolutionary trees of the 11 insertion sequences, reported in Fig. 4, inferred from the 163 sites underlined at the bottom of lanes in Fig. 4. The numbers shown are percentages of 100 bootstrap replicates in which the same internal branch was recovered. A: Distance matrix tree inferred by the program NEIGHBOR with evolutionary distances calculated by invoking the Dayhoff model of amino acid substitution of the PHYLIP program PROTDIST. Scale length represents 0.1 substitutions per site. B: Parsimony tree inferred from the protein data set with the program PROTPARS; the tree requires 835 substitutions (neglecting synonymous changes).

study the phylogenetic relationship of this family of mobile elements with related elements [10,11].

The transposase encoded by the ISC1041 element has a high degree of similarity with the transposase encoded by IS1086 from *Alcaligenes eutrophus* [30] and with those of some human pathogenic bacteria such as IS30 from *E. coli* [9] and IS4351 from *B. fragilis* [31].

Insertion sequences found in the genome of the *Sulfolobus* species do not show homologies, proving the surprising plasticity of this genome. Many insertion sequences such as ISC1439, ISC1228, ISC1078, ISC1225 have been mapped into the 156 kb sequenced genome of *S. solfataricus* P2 strain [29]. The IS we found in *S. solfataricus* MT4 does not show similarity with any of these, nor with ISC1217 described in *S. solfataricus* P1 [28], demonstrating that the present IS element is novel and it is the first report on an IS30-like transposase in Archaea. On the other hand dendrogram comparison shows that IS30 and ISC1041 may share a common ancestor.

The multiple alignments of IR of some IS30-like elements indicates a conserved function and a possible common evolutionary origin of these different mobile genetic elements. The well-conserved region among transposases of the IS30 family is putatively the catalytic domain.

Although the catalytic domain of IS30 transposase has not been identified, a large domain, called the 'integrase domain', is diffused and conserved among different integrases and transposases. It has been suggested that this domain plays an important role in the transposition process [8]. Moreover, among nucleic acid processing enzymes, a larger degree of conservation has been observed with the sequence defining the DDE motif. This motif has been found conserved in many integrases and recombinases sharing biochemical and structural similarities that are topologically related. A transposase superfamily that performs polynucleotidyl transferase reactions for a number of different biological roles has been proposed [5,6,25]. The DDE motif has never been found in the archaeal kingdom; therefore our identification of a new member of this superfamily may yield new information about an ancestral nucleic acid processing protein. The existence in the archaeal kingdom of a thermophilic transposase sharing the DDE signature not only strengthens the idea of an ancestral bacterial sequence containing this signature [10], but also shows the ability of the IS elements to perform intragenomic transposition and mobilize genetic material among different organisms.

Acknowledgements: The authors wish to thank Dr. Anna Tramontano, Dipartimento di 'Biocomputing', Istituto di Ricerca in Biologia Molecolare P. Angeletti, Pomezia, Italy for the alignments of DNA and protein sequences, and Dr. P. Cammarano for the construction of the phylogenetic trees. The skilful technical assistance of Paola Di Giamberardino is gratefully acknowledged. This research was funded by the EU Project Extremophiles as cell factories (contract ERB-BIO4-CT96-0488), by Italian MURST and by CNR Target Project on 'Biotechnology'.

References

[1] Campbell, A., Berg, D.E., Botstein, D., Lederberg, E.M., Novick, R.P., Starlinger, P. and Szybalski, W. (1975) Gene 5, 197–206.

[2] Galas, D.J. and Chandler, M. (1989) in: Mobile DNA, Bacterial Insertion Sequences (Berg, D.E. and Howe, M.M., Eds.), pp. 109–162. American Society for Microbiology, Washington, DC.

[3] Clewell, D.B. and Gawron-Burke, C. (1986) Annu. Rev. Microbiol. 40, 635–659.
[4] Neu, H.C. (1992) Science 257, 1064–1073.
[5] Craig, N.L. (1995) Science 270, 253–254.
[6] Grindley, N.D.F. and Leschziner, A.E. (1995) Cell 83, 1063–1066.
[7] Skalka, A.M. (1993) Gene 135, 175–182.
[8] Rezsöhazy, R., Hallet, B., Delcom, J. and Mahillon, J. (1993) Mol. Microbiol. 9, 1283–1295.
[9] Dalrymple, B., Caspers, P. and Arber, W. (1984) EMBO J. 3, 2145–2149.
[10] Capy, P., Vitalis, R., Langin, T., Higuet, D. and Bazin, C. (1996) J. Mol. Evol. 42, 359–368.
[11] Capy, P., Langin, T., Higuet, D., Maurer, P. and Bazin, C. (1997) Genetica 100, 63–72.
[12] Maras, B., Consalvi, V., Chiaraluce, R., Politi, L., De Rosa, M., Bossa, F., Scandurra, R. and Barra, D. (1992) Eur. J. Biochem. 203, 81–87.
[13] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[14] Bairoch, A. (1993) Nucleic Acids Res. 21, 3097–3103.
[15] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[16] Sander, C. and Schneider, R. (1991) Proteins 9, 56–68.
[17] Hubbard, T.J.P. and Park, J. (1995) Proteins 23, 398–402.
[18] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[19] Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P.G. and Skalka, A.M. (1992) Mol. Cell. Biol. 12, 2331–2338.
[20] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[21] Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package), Version 3.5c, University of Washington, Seattle, WA.
[22] Reiter, W.D., Hudepohl, U. and Zillig, W. (1990) Proc. Natl. Acad. Sci. USA 87, 9509–9513.
[23] Stalder, R., Caspers, P., Olasz, F. and Arber, W. (1990) J. Biol. Chem. 7, 3757–3762.
[24] Chotia, C. and Lesk, A.M. (1986) EMBO J. 5, 823–826.
[25] Polard, P. and Chandler, M. (1995) Mol. Microbiol. 15, 13–23.
[26] Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Cragie, R. and Davies, D.R. (1994) Science 266, 1981–1986.
[27] Muller, J.R., Potter, M. and Janz, S. (1994) Proc. Natl. Acad. Sci. USA 91, 12066–12070.
[28] Schleper, C., Röder, R., Singer, T. and Zillig, W. (1994) Mol. Gen. Genet. 243, 91–96.
[29] Sensen, C.W., Klenk, H.P., Singh, R.K., Allard, G., Y-Chan, C.C., Liu, Q.Y., Penny, S.L., Young, F., Schenk, M.E., Gaasterland, T., Doolittle, W.F., Ragan, M.A. and Charlebois, R.L. (1996) Mol. Microbiol. 22, 175–191.
[30] Dong, Q., Sadouk, A., Van Der Lelie, D., Taghavi, S., Ferhat, A., Nuyten, J.M., Borremans, B., Mergeay, M. and Toussaint, A. (1992) J. Bacteriol. 174, 8133–8138.
[31] Rasmussen, J.L., Odelson, D.A. and Macrina, F.L. (1987) J. Bacteriol. 169, 3573–3580.