JOURNAL OF BACTERIOLOGY, Mar. 1995, p. 1614–1619 Vol. 177, No. 6
0021-9193/95/$04.00+0


# A Gene Encoding a Putative Membrane Protein Homologous to the Major Facilitator Superfamily of Transporters Maps Upstream of the β-Glycosidase Gene in the Archaeon *Sulfolobus solfataricus*

ANTONELLA PRISCO, MARCO MORACCI, MOSÉ ROSSI, AND MARIA CIARAMELLA*

*Institute of Protein Biochemistry and Enzymology, Consiglio Nazionale delle Ricerche, 80125 Naples, Italy*

Received 26 September 1994/Accepted 29 December 1994

**We have identified a gene encoding a putative membrane protein homologous to the major facilitator superfamily, mapping upstream of the *lacS* gene in *Sulfolobus solfataricus*. Permeases from this family mediate secondary transport and are widely distributed among eubacteria and eukaryotes; the finding of an archaeal member suggests that this mechanism of transport evolved before the divergence of the three living domains. We also report a transcriptional mapping of the gene cluster.**

Despite the interest recently devoted to *Archaea* molecular evolution and genetics, the regulation of gene expression in these organisms has not yet been extensively investigated. It is now clear that the archaeal basal transcription machinery is strikingly similar to the eukaryotic one (5, 10, 13, 19, 24, 29); however, in some cases the organization of analogous genes is colinear in *Archaea* and *Eubacteria*, with functionally related genes cotranscribed or physically linked (for a review, see reference 4).

We have previously reported studies on the *lacS* gene from the extreme thermoacidophilic archaeon *Sulfolobus solfataricus*, coding for a thermostable β-glycosidase (S-βgly) (6, 17, 18, 20; for a recent review, see reference 16). *S. solfataricus* is able to grow on β-glycosides, such as lactose and cellobiose (9), that are potential substrates for S-βgly. *S. solfataricus* mutants containing insertion sequences in the *lacS* gene have been isolated recently (27), but the function and regulation of S-βgly in vivo have not been investigated.

In bacteria, classical operons for carbohydrate utilization include genes for regulatory and transport proteins. The latter are membrane proteins that mediate substrate uptake by one of the three major transport mechanisms: ATP-dependent transport (1), secondary transport (15), and transport by the phosphoenolpyruvate-dependent phosphotransferase system or PTS (21). The *Escherichia coli* homolog of *lacS*, called *bglB* (11), maps in the cryptic operon *bgl*, immediately downstream of *bglF*, coding for a permease of the PTS (28).

The structure and regulation of genes involved in carbohydrate uptake in *Archaea* have not been studied so far. Here we report the identification, immediately upstream of *lacS*, of a gene coding for a putative membrane protein homologous to the major facilitator superfamily (MFS).

*S. solfataricus* MT4 was grown at 85°C in rich medium (23); *E. coli* HB101 (25) was the host for plasmids. DNA manipulations were performed by standard techniques (25). Sequencing was done by the dideoxy-chain termination method (26). For RNA extraction, aliquots (500 ml) of *S. solfataricus* cultures were centrifuged, resuspended in 7 ml of extraction buffer (20 mM sodium acetate [pH 5.3], 1% sodium dodecyl sulfate, 0.1% diethylpyrocarbonate), extracted at 65°C with the same volume of a mixture of phenol-chloroform-isoamyl alcohol (24:24:1) equilibrated in 20 mM sodium acetate (pH 5.3), and ethanol precipitated. Contaminating DNA was eliminated by digestion with DNase I (Worthington). Electrophoresis, Northern (RNA) blotting, hybridization, labeling, and RNase protection were done as described previously (25). For primer extensions, end-labeled primers (0.5 to 1 pmol) (about $10^5$ cpm) were ethanol precipitated together with 5 μg of RNA, resuspended in 3 μl of reverse transcriptase buffer (50 mM Tris-HCl [pH 8.3], 6 mM $MgCl_2$, 10 mM dithiothreitol, 100 mM NaCl), denatured for 5 min at 65°C, frozen on dry ice, and annealed for 45 min at 45°C; deoxynucleoside triphosphates (final concentration of 0.4 mM each), 10 U of RNAsin (Promega) and 8 U of avian leukosis virus reverse transcriptase

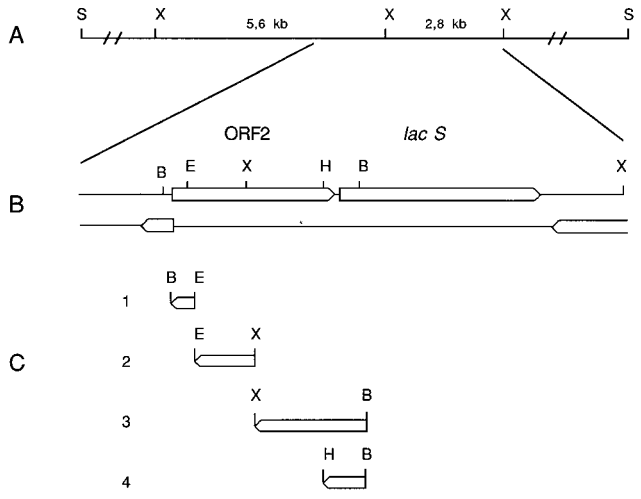


FIG. 1. (A) The 12-kb insert of the λC1 clone, showing the 2.8-kb *Xba*I fragment containing *lacS* (6) and the adjacent 5.6-kb *Xba*I fragment that was subcloned in this study. (B) Positions and relative orientations of the ORFs in the sequenced region. (C) Probes used for transcription analysis (see the legends to Fig. 4 to 6). Abbreviations for selected restriction enzyme sites: E, *Eco*RI; H, *Hin*dIII; S, *Sal*I; B, *Bgl*II; X, *Xba*I.

* Corresponding author. Mailing address: Institute of Protein Biochemistry and Enzymology, Consiglio Nazionale delle Ricerche, Via Marconi, 10, 80125 Naples, Italy. Phone: 0039/81/7257246. Fax: 0039/81/7257240 or 2396525. Electronic mail address: ROSSI1@MVX36.CSATA.IT.

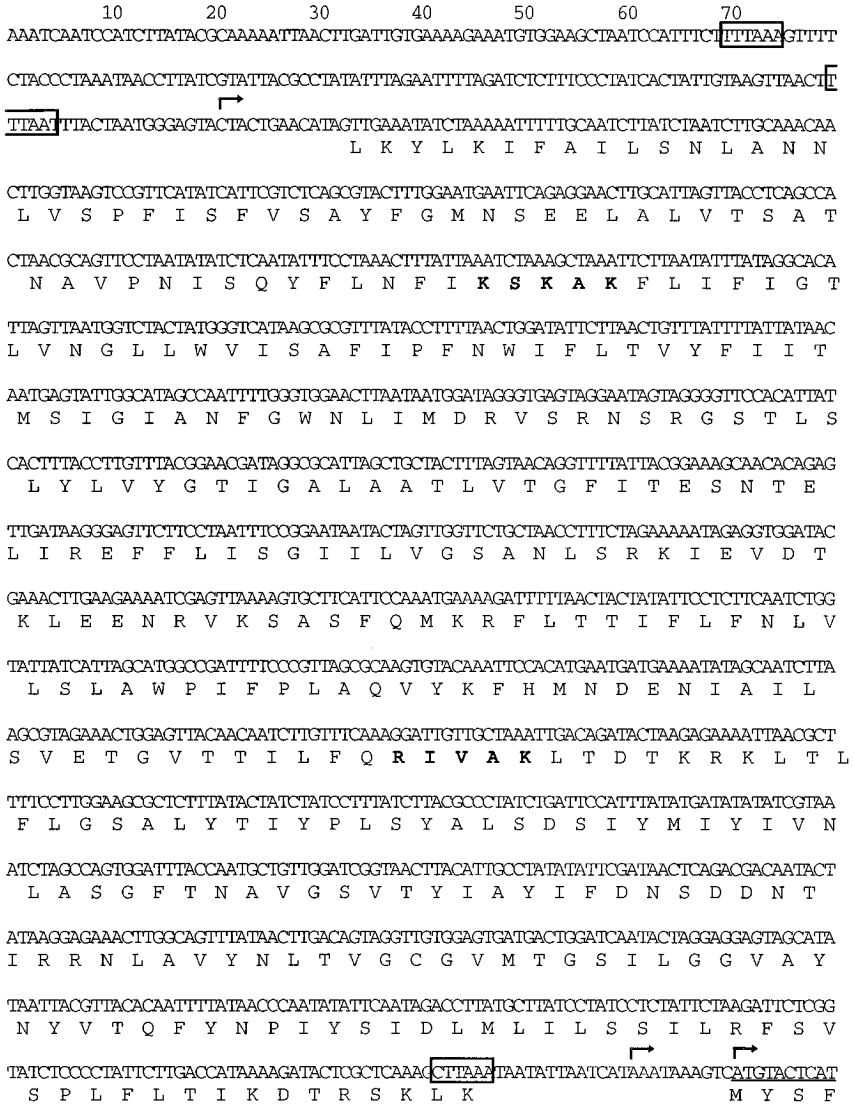


FIG. 2. Nucleotide and deduced amino acid sequences of ORF2 and surrounding regions (GenBank accession number L36201). Consensus sequences for the archaeal promoter are boxed. The lysine-rich sequence motifs conserved in MFS proteins are indicated in bold type. The 5′ end of *lacS* is underlined. The arrows indicate the 5′ ends of *lacS* and ORF2 transcripts, as mapped by RNase protection and primer extension experiments.

were added. The reaction mixture was incubated at 48°C for 30 min; samples were denatured and loaded on sequencing gels. The sequence of pORF was compared with those in the Swiss Prot, PIR, and GenEMBL libraries by using the programs FASTA, Tfasta, and BLAST of the University of Wisconsin Genetics Computer Group. Sequence alignments were obtained with the program PILEUP (8).

**Cloning and sequence analysis of ORF2.** In a previous report (6), we found, upstream of the *lacS* gene, the 3′ end of an open reading frame (ORF), contained in a 2.8-kb *Xba*I *S. solfataricus* genomic fragment partially sequenced. We first completed the sequencing of the 411 bp from the region previously sequenced up to the 5′ *Xba*I site; then, by using a probe containing the 5′ end of the 2.8-kb *Xba*I fragment, we subcloned the adjacent 5.6-kb *Xba*I genomic fragment from the λC1 clone (Fig. 1A); within this, we sequenced a region of about 1.2 kb. We found a complete ORF (ORF2), and, further upstream, an ORF of 125 codons on the opposite strand (Fig. 1B). ORF2 potentially codes for a 378-amino-acid protein (pORF) starting from a TTG codon (Fig. 2). The predicted amino acid composition of pORF is 45% hydrophobic, and its hydrophobicity pattern is typical of those of integral membrane proteins. A computer search for sequence homology showed a significant sequence similarity between pORF and members of the major facilitator superfamily or MFS, including over 50 symporters, antiporters, and uniporters found so far in eubacteria and eukaryotes (15). Comparison of the sequence of pORF to that of the *E. coli* arabinose/$H^+$ symporter, AraE, reveals 24% identity and 51% similarity; this degree of sequence homology is comparable to that found among evolutionarily distant members of the MFS (3, 14). More impressive than the sequence homology is the structural similarity between these two proteins, as shown by their hydropathy plots (Fig. 3A and B); both proteins contain two blocks of six putative transmembrane domains, separated by a hydrophilic loop that is larger in the *E. coli* protein. pORF shows a similar

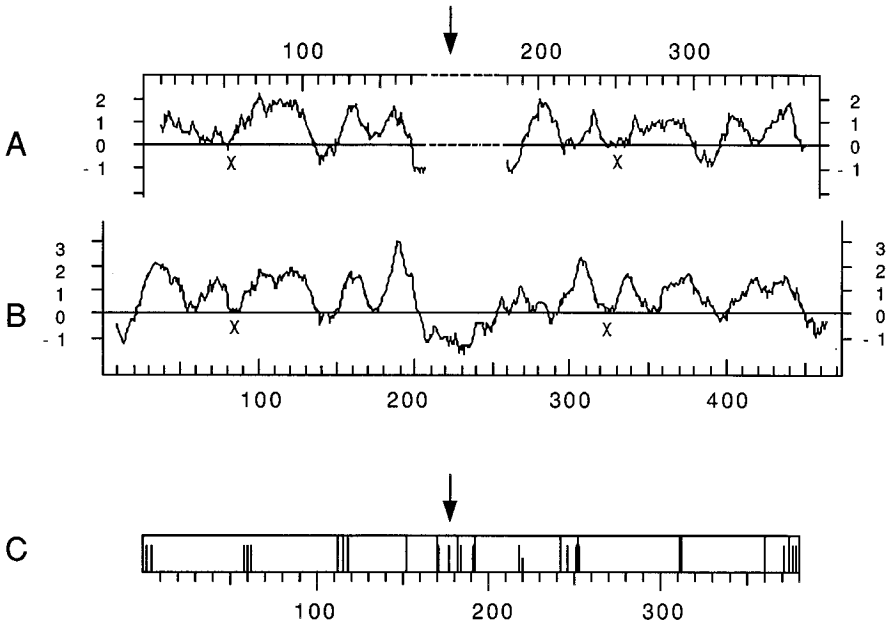


FIG. 3. Hydropathic profiles of pORF (A) and AraE (B); hydropathy values for a window of 19 amino acid residues were averaged and plotted by the position of the middle residue (12). X indicates the position of the MFS motifs; the arrow indicates the axis of symmetry of pORF. pORF can be considered a duplication of two portions, from positions 1 to 180 and 181 to 378, which show about 20% identity and 47% similarity with each other; the hydrophobicity patterns and the position of the MFS motifs overlap. (C). In pORF, as in bacterial integral membrane proteins, positively charged amino acids (indicated by vertical lines) are clustered in hydrophilic loops.

degree of sequence and structural similarity with other MFS proteins. The signature of the MFS family is a sequence motif found twice in almost all the members, whose consensus is (**K**/**R**)$X_{2-3}$(**K**/**R**) (15); in pORF, the motifs **K**SKA**K** and **RI**-VA**K** were found in corresponding positions (Fig. 2 and 3).

The organization of the ORF2-*lacS* region resembles the structure of bacterial operons; however, this region seems not to be the archaeal counterpart of the *E. coli bgl* operon. First, no potential regulatory genes were found in the region sequenced, suggesting that either the cluster is not regulated, or regulatory gene(s) are not strictly linked. Second, the β-glucoside permease BglF, whose gene maps in a position corresponding to ORF2 (with respect to the structural genes *bglB* and *lacS*) is a phoshotransferase of PTS, rather then a symporter, and also has regulatory functions (2). We noticed that S-βgly, unlike BglB, is not able to hydrolyze phosphorylated sugars (18); therefore, its substrate(s) are not expected to be phosphorylated during transport. Transport by PTS has been demonstrated so far only in *Eubacteria* and could be a mechanism peculiar to this living domain; in contrast, permeases of the MFS are widely distributed among eubacteria and eukaryotes. After the finding of an archaeal member of the family, it is tempting to speculate that secondary transport is a more ancient mechanism and evolved before the separation of the three domains. MFS proteins show a 12-transmembrane segment structure with internal symmetry; it has been suggested that MFS genes have evolved by duplication of a six-helix ancestor gene (15). The same internal symmetry can be recognized in pORF (Fig. 3), suggesting that the duplication of the ancestor gene occurred before the divergence of the three domains.

**Transcription analysis.** By inspecting the sequences, potential archaeal terminator sequences (22) were found upstream of ORF2 and downstream of *lacS*, but not between the two genes, suggesting that they might be cotranscribed. To test this possibility, RNA prepared from *S. solfataricus* cultures at different growth stages were analyzed by Northern blotting with antisense RNA probes specific for *lacS* and ORF2 (Fig. 4A and B, respectively). Transcripts of about 1.1 and 1.5 kb were detected, whose sizes are in agreement with the lengths of the corresponding genes; no abundant larger transcripts were evident. Both transcripts show a peak of expression at the exponential phase of growth, but their relative abundance is very

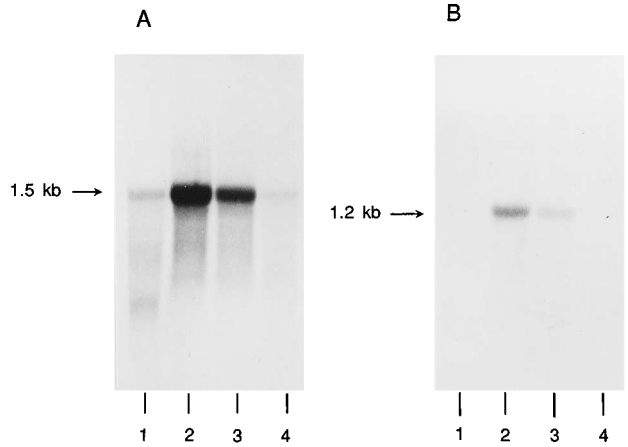


FIG. 4. Northern blots showing *lacS* (A) and ORF2 (B) transcripts. *S. solfataricus* RNAs were extracted at different growth stages: lanes 1, optical density at 600 nm ($OD_{600}$) of 0.3; lanes 2, $OD_{600}$ of 0.4; lanes 3, $OD_{600}$ of 0.6; lanes 4, $OD_{600}$ of 0.8. The same amount of RNA (2 μg) was loaded on each lane; each filter was hybrydized with riboprobes labeled at the same specific activity and exposed for the same time span. Longer exposures of filter B revealed that ORF2 transcript is also present in lane 1. Riboprobes were prepared from the 185-bp *Hin*dIII-*Bgl*II fragment (probe 4 of Fig. 1C) for *lacS* and the 420-bp *Eco*RI-*Xba*I fragment (probe 2 of Fig. 1C) for ORF2.

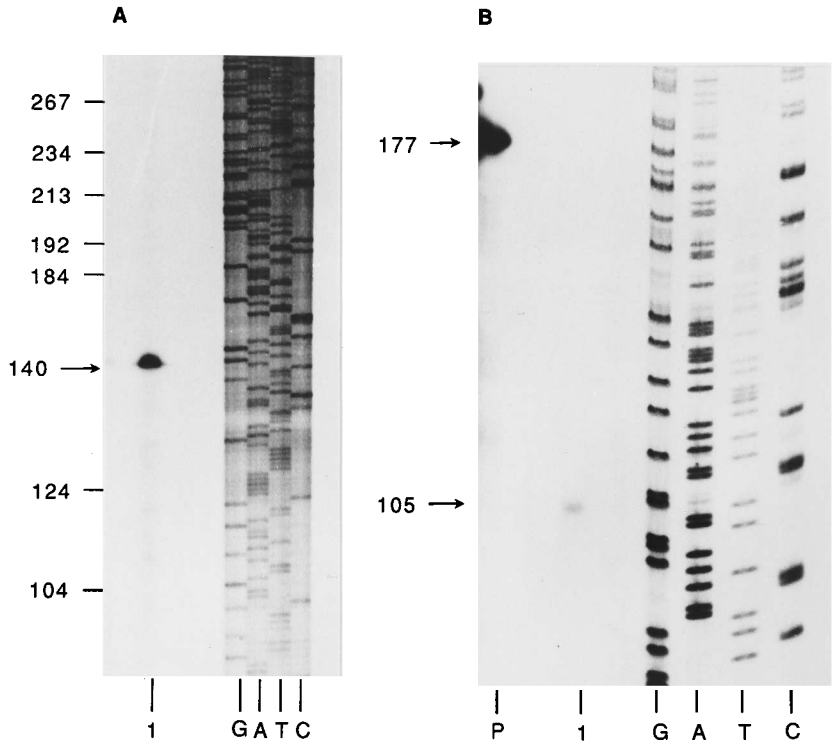


FIG. 5. 5′ mapping of ORF2 RNA. (A) Primer extension. Five micrograms of total RNA (template) and 0.5 pmol of the 17-mer annealing to the sequence from +125 to +141 with respect to the start codon (lane 1) were used. (B) RNase protection. Five micrograms of total RNA was hybridized to $10^4$ cpm of the $^{32}$P-labeled riboprobe prepared from the 160-bp *Bgl*II-*Eco*RI fragment (probe 1 of Fig. 1C), annealing to the sequence from −60 to +100 with respect to the start codon (lane 1). RNA was extracted from cultures grown at an optical density at 600 nm of 0.4. The arrows show the lengths of bands corresponding to the 5′ end. The numbers indicate molecular weight markers (in nucleotides). P, undigested probe; G, A, T, C, sequence ladder obtained with the same primer used for primer extension.

different (the *lacS* RNA is at least 10-fold more abundant than ORF2 RNA). Both genes are transcribed in cultures grown in complete medium as well as in minimal medium containing glucose (unpublished results). A screening on minimal medium containing different carbohydrates as the carbon source is required to establish whether transcription of the two genes is coordinated and somehow regulated.

We have mapped the 5′ ends of the *lacS* and ORF2 transcripts by RNase protection and primer extension experiments. For ORF2, both techniques identified a unique 5′ end, located 10 bp from the first TTG (Fig. 5). The 5′ ends of *lacS* RNA are heterogeneous, with two major sites identified both by primer extension (Fig. 6A) and by RNase protection (Fig. 6B), mapping at −9 and +1 with respect to the A of the start codon, respectively. In the primer extension experiment, an additional band corresponding to a 5′ end mapping at −168, was evident, and is likely to be a technical artifact. Consensus sequences for the archaeal promoter (10) can be found at a canonical distance from the 5′ end of ORF2 and from position +1 of *lacS* (Fig. 2); these 5′ ends therefore possess the prerequisites to be true transcription start sites; surprisingly, position −168 of *lacS*, which was not identified by RNase protection, is at a proper distance from a promoter sequence, whereas position −9, identified by both techniques, is not. Taken together, transcription analysis data show that, unlike bacterial operons, the two genes may be transcribed from independent promoters. By reverse PCR, we could also detect a rare transcript(s) that transverses the junction between the two genes (not shown); therefore, we cannot rule out the possibility that an additional polycistronic messenger is transcribed from a common promoter and either processively degraded from the 5′ to the 3′ end or specifically cleaved; in this latter view, the 5′ end mapped at −9 in *lacS*, which is not at a proper distance from a promoter sequence, could be a processing site. It is also possible that transcripts comprising sequences of the two genes arise by read-through of ORF2 transcripts, caused by the lack of a strong terminator sequence. Additional experiments are required to elucidate the role of each transcript.

**Heterologous expression of pORF.** In order to gain information about the function of pORF, we tried to obtain its expression in *E. coli*; however, the protein is strongly toxic, inducing block of growth and filamentation after 30 min of induction (not shown). This phenotype suggests that insertion into the membrane of pORF interferes with essential function(s). The archaeal membrane is peculiar among those of all living organisms in its lipidic components (7); it is not unlikely that a protein evolved to functionally interact with such a particular environment is dangerous in a different context.

Additional experiments are required to elucidate the role of pORF. The study of the function of protein in vivo requires molecular genetic tools that are still not available for thermophilic *Archaea*; on the other hand, the high toxicity of the protein rules out both the complementation of *E. coli* transport mutants and easy protein production. Possible approaches are in vivo transport studies by using different labeled substrates and eventually toxic analogs and in vitro studies on reconstituted membranes using the protein purified from *S. solfataricus*.

**Nucleotide sequence accession number.** The nucleotide sequence of the ORF2 region has been assigned GenBank accession number L36201.

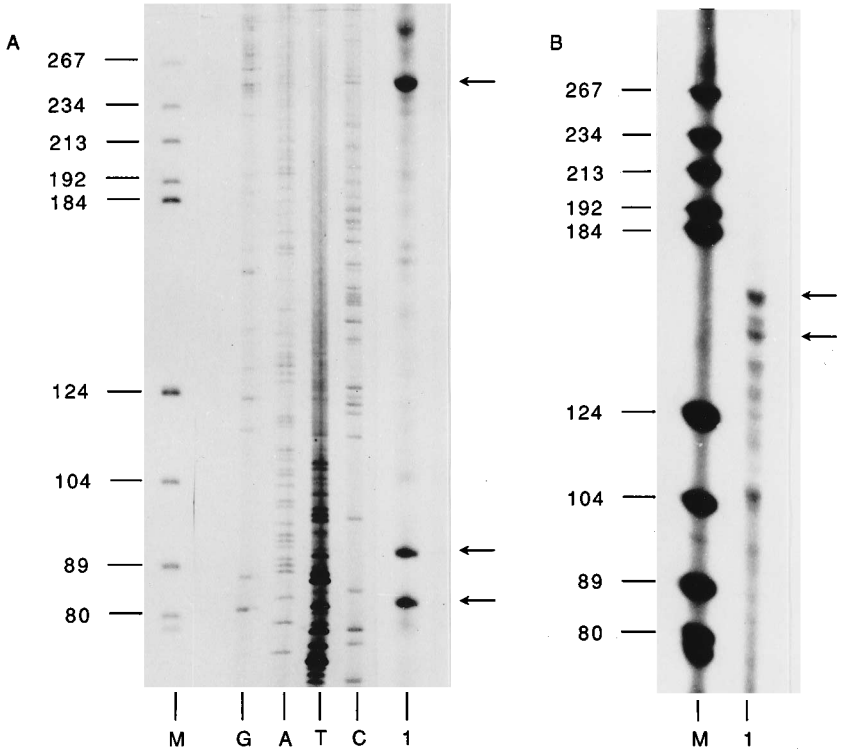


FIG. 6. 5′ mapping of *lacS* RNA. (A) Primer extension. Five micrograms of total RNA (template) and 0.5 pmol of the 20-mer annealing to the sequence from +63 to +83 with respect to the start codon (lane 1) were used. (B) RNase protection. Five micrograms of total RNA was hybridized to $10^4$ cpm of the $^{32}$P-labeled riboprobe prepared from the 806-bp *Xba*I-*Bgl*II fragment (probe 3 of Fig. 1C), annealing to the sequence from −652 to +154 with respect to the start codon (lane 1). RNA was extracted from cultures grown at an optical density at 600 nm of 0.4. The other symbols are explained in the legend to Fig. 5.

We thank Arnold Driessen for help in gene bank search, Mario De Rosa and Antonio Ottombrino for *S. solfataricus* biomass, Eduardo Patriarca for useful discussion, Gianni Imperato for technical assistance, and Ottavio Piedimonte for the art work.

This work was partially supported by CNR Target Project Biotechnology and Bioinstrumentation and by CNR P. F. Ingegneria Genetica.

## REFERENCES

1. **Ames, G. F., C. S. Mimura, and V. Shyamala.** 1990. Bacterial periplasmic permeases belong to a family of transport proteins operating from *E. coli* to human: traffic ATPases. FEMS Microbiol. Rev. **75:**429–446.
2. **Amster-Choder, O., F. Houman, and A. Wright.** 1989. Protein phosphorylation regulates transcription of the β-glucoside utilization operon in *E. coli*. Cell **58:**847–855.
3. **Cheng, Q., and C. A. Michels.** 1989. The maltose permease encoded by the MAL61 gene of *Saccharomyces cerevisiae* exhibits both sequence and structural homology to other sugar transporters. Genetics **123:**477–484.
4. **Ciaramella, M., R. Cannio, M. Moracci, F. M. Pisani, and M. Rossi.** Molecular biology of extremophiles. World J. Microbiol. Biotechnol., in press.
5. **Creti, R., P. Londei, and P. Cammarano.** 1993. Complete nucleotide sequence of an archaeal (*Pyrococcus woesei*) gene encoding a homolog of eukaryotic transcription factor IIB (TFIIB). Nucleic Acids Res. **21:**2942.
6. **Cubellis, M. V., C. Rozzo, P. Montecucchi, and M. Rossi.** 1990. Isolation and sequencing of a new β-galactosidase-encoding archaebacterial gene. Gene **94:**89–94.
7. **De Rosa, M., A. Trincone, B. Nicolaus, and A. Gambacorta.** 1991. Archaebacteria: lipids, membrane structures and adaptation to environmental stresses, p. 61–87. *In* G. di Prisco (ed), Life in extreme conditions. Springer, Berlin.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1988. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
9. **Grogan, D. W.** 1989. Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. J. Bacteriol. **171:**6710–6719.
10. **Hain, J., W. D. Reiter, U. Hudepohl, and W. Zillig.** 1992. Elements of an archaeal promoter defined by mutational analysis. Nucleic Acids Res. **20:** 5423–5428.
11. **Henrissat, B.** 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280:**309–316.
12. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
13. **Langer, D., and W. Zillig.** 1993. Putative tfIIs gene of *Sulfolobus acidocaldarius* encoding an archaeal transcription elongation factor is situated directly downstream of the gene for a small subunit of DNA-dependent RNA polymerase. Nucleic Acids Res. **21:**2251.
14. **Maiden, M. C. J., E. Davis, S. A. Baldwin, D. C. M. Moore, and P. J. F. Henderson.** 1987. Mammalian and bacterial sugar transport proteins are homologous. Nature (London) **325:**641–643.
15. **Marger, M. D., and M. H. Saier.** 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends Biochem. Sci. **18:**13–20.
16. **Moracci, M., M. Ciaramella, R. Nucci, L. H. Pearl, I. Sanderson, A. Trincone, and M. Rossi.** 1994. Thermostable β-glycosidase from *Sulfolobus solfataricus*. Biocatalysis **11:**89–103.
17. **Moracci, M., A. La Volpe, J. F. Pulitzer, M. Rossi, and M. Ciaramella.** 1992. Expression of the thermostable β-galactosidase gene from the archaebacterium *Sulfolobus solfataricus* in *Saccharomyces cerevisiae* and characterization of a new inducible promoter for heterologous expression. J. Bacteriol. **174:** 873–882.
18. **Nucci, R., M. Moracci, C. Vaccaro, N. Vespa, and M. Rossi.** 1993. Exoglucosidase activity and substrate specificity of the β-glycosidase isolated from the extreme thermophile *Sulfolobus solfataricus*. Biotechnol. Appl. Biochem. **17:**239–250.
19. **Ouzounis, C., and C. Sander.** 1992. TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes. Cell **71:**189–190.
20. **Pisani, F. M., R. Rella, C. Rozzo, C. A. Raia, R. Nucci, A. Gambacorta, M. De Rosa, and M. Rossi.** 1990. Thermostable β-galactosidase from the archaebacterium *Sulfolobus solfataricus*. Purification and properties. Eur. J. Biochem. **187:**321–328.
21. **Postma, P. W., J. W. Lengeler, and G. R. Jacobson.** 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase system of bacteria. Microbiol. Rev. **57:**543–594.
22. **Reiter, W. D., P. Palm, and W. Zillig.** 1988. Transcription termination in the archaebacterium *Sulfolobus*: signal structures and linkage to transcription initiation. Nucleic Acids Res. **16:**2445–2459.

23. **Rella, R., C. A. Raia, M. Pensa, F. M. Pisani, A. Gambacorta, and M. De Rosa.** 1987. A novel archaebacterial $NAD^+$-dependent alcohol dehydrogenase. Purification and properties. Eur. J. Biochem. **167:**475–479.
24. **Rowlands, T., P. Baumann, and S. P. Jackson.** 1994. The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. Science **264:**1326–1328.
25. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
27. **Schleper, C., R. Roder, T. Singer, and W. Zillig.** 1994. An insertion element of the extremely thermophilic archaeon *Sulfolobus solfataricus* transposes into the endogenous β-galactosidase gene. Mol. Gen. Genet. **243:**91–96.
28. **Schnetz, K., and B. Rak.** 1988. Regulation of the *bgl* operon of *Escherichia coli* by transcriptional antitermination. EMBO J. **7:**3271–3277.
29. **Zillig, W., P. Palm, D. Langer, H. P. Klenk, M. Lanzendorfer, U. Hudepohl, and J. Hain.** 1992. RNA polymerases and transcription in Archaebacteria, p. 79–88. *In* M. J. Danson, D. W. Hough, and G. G. Lunt (ed.), The Archaebacteria: biochemistry and biotechnology. The Biochemical Society, London.