JOURNAL OF BACTERIOLOGY, Aug. 2002, p. 4430–4441 Vol. 184, No. 16
0021-9193/02/$04.00+0 DOI: 10.1128/JB.184.16.4430–4441.2002

Genes of De Novo Pyrimidine Biosynthesis from the Hyperthermoacidophilic Crenarchaeote *Sulfolobus acidocaldarius*: Novel Organization in a Bipolar Operon

Thia-Lin Thia-Toong,[1] Martine Roovers,[1] Virginie Durbecq,[2] Daniel Gigot,[2] Nicolas Glansdorff,[1,2] and Daniel Charlier[1]*

Erfelijkheidsleer en Microbiologie, Vrije Universiteit Brussel,[1] and Laboratoire de Microbiologie, Université Libre de Bruxelles and Institut de Recherches Microbiologiques J.-M. Wiame,[2] B-1070 Brussels, Belgium

Received 8 March 2002/Accepted 31 May 2002

Sequencing a 8,519-bp segment of the *Sulfolobus acidocaldarius* genome revealed the existence of a tightly packed bipolar pyrimidine gene cluster encoding the enzymes of de novo UMP synthesis. The G+C content of 35.3% is comparable to that of the entire genome, but intergenic regions exhibit a considerably lower percentage of strong base pairs. Coding regions harbor the classical excess of purines on the coding strand, whereas intergenic regions do not show this bias. Reverse transcription-PCR and primer extension experiments demonstrated the existence of two polycistronic messengers, *pyrEF-orf8* and *pyrBI-orf1-pyrCD-orf2-orf3-orf4*, initiated from a pair of divergent and partially overlapping promoters. The gene order and the grouping in two wings of a bipolar operon constitute a novel organization of *pyr* genes that also occurs in the recently determined genome sequences of *Sulfolobus solfataricus* P2 and *Sulfolobus tokodaii* strain 7; the configuration appears therefore characteristic of *Sulfolobus*. The quasi-leaderless *pyrE* and *pyrB* genes do not bear a Shine-Dalgarno sequence, whereas the initiation codon of promoter-distal genes is preceded at an appropriate distance by a sequence complementary to the 3′ end of 16S rRNA. The polycistronic nature of the *pyr* messengers and the existence of numerous overlaps between contiguous open reading frames suggests the existence of translational coupling. *pyrB* transcription was shown to be approximately twofold repressed in the presence of uracil. The mechanism underlying this modulation is as yet unknown, but it appears to be of a type different from the various attenuation-like mechanisms that regulate *pyrB* transcription in bacteria. In contrast, the *pyrE-pyrB* promoter/control region harbors direct repeats and imperfect palindromes reminiscent of target sites for the binding of a hypothetical regulatory protein(s).

De novo synthesis of UMP is universally performed in six steps catalyzed by conserved enzymes in all three domains of life (Fig. 1). In lower eukaryotes, such as the yeast *Saccharomyces cerevisiae*, the carbamoylphosphate synthetase (CPSase) (EC 6.3.5.5) and aspartate carbamoyltransferase (ATCase) (EC 2.1.3.2) activities are fused in a single polypeptide chain, also containing an inactive version of dihydroorotase (DHOase) (EC 3.5.2.3). In animals the pyrimidine biosynthetic activities are highly organized; thus, the trifunctional CAD protein shows CPSase, ATCase, and DHOase activities assembled in a single polypeptide chain, and similarly, orotate phosphoribosyltransferase (OPRTase) (EC 2.4.2.10) and orotidine-5′-monophosphate decarboxylase (OMPdecase) (EC 4.1.1.23) are fused in a bifunctional protein. In bacteria and archaea the different reactions are performed by monofunctional enzymes, but some of them do interact, at least transiently. Thus, in *Pseudomonas putida* and *Pseudomonas aeruginosa*, *pyrB* overlaps a *pyrC′* gene encoding a nonfunctional DHOase, which is required for the assembly of the functional dodecameric ATCase (49), and feedback inhibition and thermostability of ATCase from the extreme thermophilic bacterium *Thermus ZO5* are conferred by coexpression of *pyrB* and the adjacent *bbc* (for "between *b* and *c*") and *pyrC* genes, respectively (58, 59). In hyperthermophilic bacteria and archaea, CPSase or carbamate kinase-like CPSase appears to interact physically with ATCase and its paralogue, ornithine carbamoyltransferase, from the arginine biosynthetic pathway, thereby allowing substrate channeling and protection of the thermolabile and potentially toxic carbamoylphosphate from the hot aqueous environment (35, 57, 43, 44, 40). Indeed, since the pathway includes several thermolabile and energy-rich substrates and precursors, some of which decompose into toxic degradation products, the very possibility of pyrimidine biosynthesis is not obvious in hot environments. Yet all investigated hyperthermophilic bacteria and archaea use the same, classical route; evidently, this situation necessitates adapted strategies.

The *pyr* genes are organized in a vast number of different ways; most microorganisms exhibit an intermediate scattering, with monocistronic and polycistronic messengers. In *Esche-*

* Corresponding author. Mailing address: Erfelijkheidsleer en Microbiologie, Vrije Universiteit Brussel, 1-av. E. Gryson, B-1070 Brussels, Belgium. Phone: 32 2 526 72 79. Fax: 32 2 526 72 73. E-mail: dcharlie@vub.ac.be.

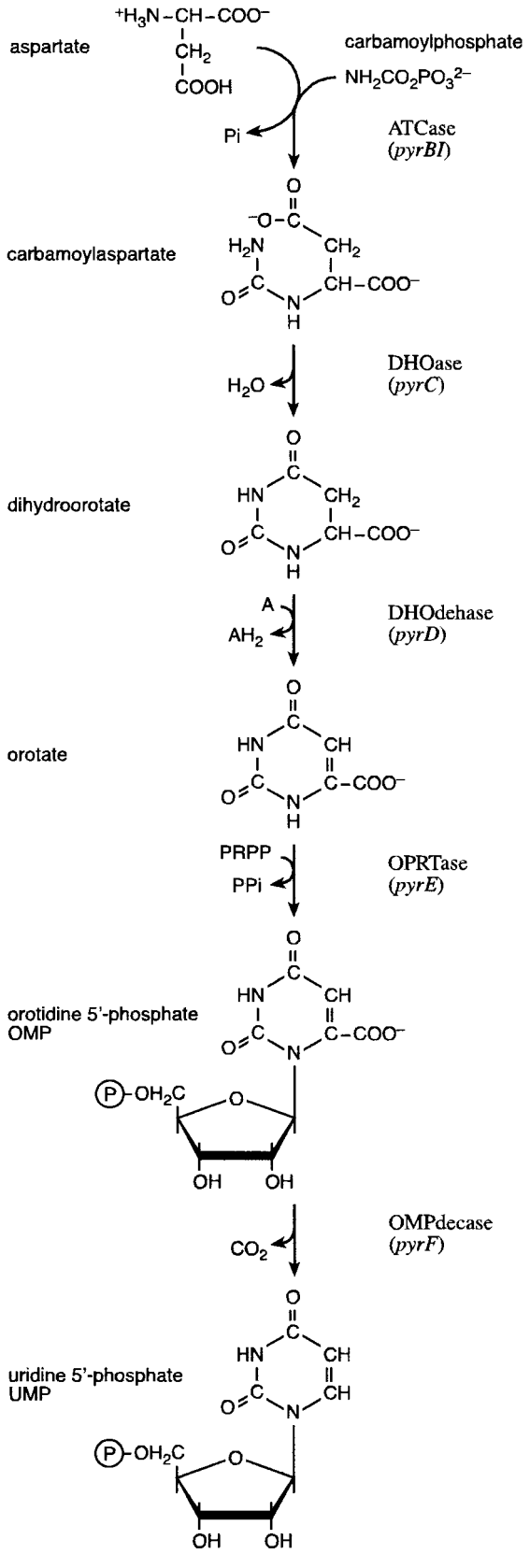

FIG. 1. De novo synthesis of UMP. Enzymes and their corresponding gene symbols are indicated.

richia coli and *Salmonella enterica* serovar Typhimurium, the *pyr* genes are all scattered throughout the chromosome; only *carA* and *carB*, encoding the glutamine amidotransferase and catalytic subunits of CPSase, and *pyrB* and *pyrI*, encoding the catalytic and regulatory subunits of ATCase, are grouped into two small operons (for a review, see reference 42). The other extreme occurs in *Bacillus subtilis*, where 10 cistrons, comprising the pyrimidine biosynthetic genes, the gene for the regulator (*pyrR*), and the gene for a uracil permease (*pyrP*), are grouped in a compact, unipolar operon (45). A similar situation is found with the extreme thermophilic gram-positive bacterium *Bacillus caldolyticus* (17, 18). In the completely sequenced genomes of several euryarchaeota, including methanogens, halophiles, and hyperthermophilic *Thermococcales*, the pyrimidine genes are highly dispersed, with only *pyrB* and *pyrI* showing more systematic clustering. An intermediate situation prevails with *Aeropyrum pernix* (three small clusters, *pyrBI*, *pyrDC*, and *pyrFE*) and *Pyrobaculum aerophilum* (*pyrFB* and *pyrDEI* clusters), the only crenarchaeotes not belonging to the *Sulfolobales* genus for which the entire genome sequences have been established (31, 15).

In *Bacteria*, pyrimidine-specific control of gene expression is exerted by a multitude of distinct mechanisms. Remarkably, even within a single organism like *E. coli*, the scattered genes and operons are noncoordinately regulated by multiple mechanisms, including UTP-sensitive transcription attenuation (*pyrBI* and *pyrE*), UTP-sensitive reiterative transcription (*pyrBI* and *carAB*), nucleotide-pool-dependent translational control (*pyrC* and *pyrD*), mRNA stability (*pyrF*) or still, as demonstrated for the *carAB* operon, complex transcription initiation control imposed by the concerted action of several multifunctional proteins, some of which combine catalytic and regulatory activities (5, 6, 20, 26, 33, 42). The pyrimidine operon of *B. subtilis* is submitted to transcriptional attenuation exerted by PyrR, an RNA binding protein that also has residual uracil phosphoribosyltransferase activity (54, 56). UTP-dependent binding of PyrR to three untranslated regions of the polycistronic mRNA disrupts the formation of the antiterminator stem-loop structures (3). An analogous mechanism appears to be operative in the gram-positive bacteria *B. caldolyticus* (17, 18), *Lactobacillus plantarum* (12), *Lactococcus lactis* (38, 39), and *Enterococcus faecalis* (36). An alternative PyrR-dependent attenuation mechanism was proposed in the extreme thermophilic gram-negative bacterium *Thermus* strain Z05 (58). In this organism, transcriptional attenuation would result from the coupling of transcription and translation of a short open reading frame (ORF), whose ribosome binding site is occluded by pyrimidine-dependent binding of PyrR to the cognate mRNA. PyrR binding would prevent translation of the leader peptide, thus promoting the formation of the terminator structure that leads to reduced expression of the downstream genes. Data on modulation of pyrimidine biosynthetic enzyme specific activities in archaea are extremely scarce and to the best of our knowledge have been reported only for *Sulfolobus acidocaldarius* (23; also this work); no information is yet available on the mechanism(s) that imposes pyrimidine-specific control of gene transcription in archaea.

In this study we present the isolation, organization, and characterization of the pyrimidine gene cluster of *S. acidocaldarius*, an aerobic thermoacidophilic sulfur-oxidizing crenar-

TABLE 1. Strains and plasmids used in this work

Strain or plasmid	Genotype	Source or reference
S. acidocaldarius DSM639	Wild-type	Deutsche Sammlung von Microorganismen
E. coli		
XL1-Blue	F′::Tn*10 proA*$^+$*B*$^+$ *lacI*^{q} Δ(*lacZ*)*M15*/*recA1 endA1 gyrA96* (Nal^r) *thi hsdR17*(r$^-$ m$^+$) *supE44 relA1 lac*	Stratagene
JM101	F′ *traD36 lacI*^{q} Δ(*lacZ*) *M15 proA*$^+$*B*$^+$/*supE thi* Δ(*lac-proAB*) *strA slcB15 endA hspR4*	41
HMS174 (DE3)/pLysS	F$^-$ *recA1 hsdR* (r$^-$ m$^+$) Rif^r (DE3)/pLysS (Cm^r)	Novagen
Plasmids		
pUC18	Ap^r, *lacZ* cloning vector	60
pGEM-7 Zf(+)	Ap^r, cloning vector	Promega
pKK223-3	Ap^r, IPTG-inducible expression vector	Pharmacia
pET3a	Ap^r, IPTG-inducible and T7 RNA-polymerase dependent expression vector	Novagen
pSPYR3	9.6-bp *Pst*I genomic fragment of *S. acidocaldarius* in pKK223-3	11
pSPYRE	1,493-bp *Sac*I genomic fragment of *S. acidocaldarius* cloned as *Bam*HI PCR product in pGEM7 Zf(+)	This study
pSPYRF	1,316-bp *Eco*RI genomic fragment of *S. acidocaldarius* cloned as *Bam*HI PCR product in pGEM7 Zf(+)	This study
pET3PYRE	*pyrE* coding region of *S. acidocaldarius* in pET3a	This study

chaeote that grows optimally around 80°C and pH 3.0 (4). In several aspects, *S. acidocaldarius* is the best-studied member of the *Sulfolobales*; its genome sequence has not yet been established, whereas those of the relatives *Sulfolobus solfataricus* P2 and *Sulfolobus tokodaii* strain 7 have recently been determined (52, 32). We demonstrate that the pyrimidine gene cluster of *S. acidocaldarius* constitutes a bipolar operon transcribed from two major promoters, an unprecedented organization for pyrimidine genes that turns out to be characteristic of *Sulfolobus*. We determined transcription initiation sites and polycistronic messenger compositions and demonstrate pyrimidine-specific modulation of transcription initiation. We performed amino acid sequence comparisons and analyzed in silico the nucleotide sequence in terms of transcription and translation initiation and stop signals.

MATERIALS AND METHODS

Strains, media, and growth conditions. Genotypes and descriptions of strains and plasmids are given in Table 1. *S. acidocaldarius* (type strain DSM 639) was grown aerobically at 75°C on a rotary shaker platform in either complex medium or minimal medium as described previously (11, 13). Growth conditions for *E. coli* were described previously (19). Ampicillin was used at a concentration of 25 μg/ml, kanamycin at 30 μg/ml, tetracycline at 15 μg/ml, and uracil at 50 μg/ml.

DNA preparations and manipulations, sequencing strategy, and amino acid sequence comparisons. Plasmid DNA extraction was based on the alkaline sodium dodecyl sulfate lysis method and performed with the commercial Qiaprep Spin Miniprep kit (Qiagen). Oligonucleotides were purchased from Gibco BRL (Table 2). Nuclease digestion, ligation, dephosphorylation, and phosphorylation of DNA fragments and oligonucleotides were performed with commercial enzymes and buffers according to the manufacturer's instructions (Roche). Competent cells were prepared by the $CaCl_2$ method (10). DNA sequencing was performed by the enzymatic dideoxy chain-terminating method with double-stranded plasmid DNA as the template (47). The nucleotide sequence of the 6,845-bp *Pst*I fragment of plasmid pSPYR3 was determined for both strands by subcloning of the *Pst*I-*Xba*I borders (540 and 1,000 bp) and internal *Xba*I subfragments (290, 400, 2,200, 2,400 bp) and the generation of nested series of deletions by partial *Exo*III nuclease digestion. The correct order of the *Xba*I subfragments in the *Pst*I clone was determined by sequencing the boundaries using as a template pSPYR3 plasmid DNA and as primers oligonucleotides designed on the basis of the established sequences of the subclones. Amino acid sequences of *S. acidocaldarius* enzymes were used as queries to retrieve similar sequences from databases using the BLASTP program. Multiple and pairwise alignments of amino acid sequences were generated using the Clustal W program (55).

Reverse transcriptase primer extension. Total RNA was prepared with the Life Technologies procedure using the Trizol reagent (8) from 200-ml cultures of *S. acidocaldarius* cells grown in complex medium, minimal medium, and minimal medium supplemented with uracil (50 μg/ml) and arrested in the exponential phase. Total RNA (100 μg) was mixed with about 40,000 cpm of 5′-^{32}P-end-labeled oligonucleotide primer, and after overnight hybridization at 42°C this was elongated with 10 U of avian myeloblastosis virus reverse transcriptase (Roche) at 40°C for 1 h as described previously (7). Chain-terminating DNA sequencing reactions of the noncoding strand obtained with pSPYR3 plasmid DNA as the template and the same 5′-end-labeled nucleotides used as the primer were applied as reference ladders.

Reverse transcription-PCR (RT-PCR). Total RNA was extracted with the RNeasy Midi kit (Qiagen) according to the manufacturer's instructions from *S. acidocaldarius* cells grown in minimal medium and harvested in the exponential phase. To eliminate all possible traces of contaminant DNA, a supplementary DNase I treatment was performed by incubating 1 U of enzyme (RNase free; Roche) per μg of RNA at 37°C for 30 min. DNase I was inactivated by incubation at 75°C for 5 min. cDNA synthesis was performed in a total volume of 20 μl, with 1.0 μg of RNA, 30 pmol of oligonucleotide, 1.0 mM concentration (each) of the four deoxynucleoside triphosphates, 10 mM dithiothreitol, and 50 U of Expand Reverse Transcriptase (Roche) in the commercial buffer and in the presence of 20 U RNase inhibitor (Roche). RNA and primer were heated at 65°C for 10 min and then mixed with the other components and incubated at 42°C for 90 min. The reaction was stopped by cooling on ice. Controls were performed in the absence of Expand reverse transcriptase. Three-microliter cDNA aliquots were used as a template in the PCR amplification step with different combinations of oligonucleotide pairs (0.6 μM each) in a total volume of 50 μl, with a 0.2 mM concentration (each) of the four deoxynucleoside triphosphates and 1.5 U of *Pfu* DNA polymerase (Promega) and in the commercial buffer. Initial denaturation was for 5 min at 94°C, followed by 30 cycles of synthesis comprising 1 min of denaturation at 94°C, 30 s of annealing at 50°C, and elongation at 72°C for 2 min per kilobase. The amplification was ended with a 10-min elongation at 72°C. Seventeen-microliter aliquots of the various PCRs were analyzed by agarose gel electrophoresis.

Enzyme assays. ATCase activity was measured using the colorimetric method of Prescott and Jones on dialyzed cell extracts as described previously (11). For OPRTase assays, bacterial pellets from 100-ml cultures were resuspended in 3 ml of 100 mM Tris-HCl, pH 8.8, and sonicated for 6 min. Cell debris was removed

TABLE 2. Oligonucleotide used in this work

Oligonucleotide	Sequence	Use
E3Sac	5′-CGGGATCCCTGCTTCGTATACTAACAAGG-3′	Inverse PCR, *pyrE*
E5Bam	5′-CGGGATCCGCTATGGATATTGTCTTACCAC-3′	Inverse PCR, *pyrE*
PYRF1	5′-CGGGATCCCATTCTCAGCGACGAGATTTGAC-3′	Inverse PCR, *pyrF*
PYRF2	5′-CGGGATCCGCTCATATCTGATTATATGGTG-3′	Inverse PCR, *pyrF*
PYRENde	5′-GGAATTCCATATGGATTTCGTGAAAGCTCTACTTG-3′	Amplification coding region, *pyrE*
PYREBam	5′-CGGGATCCCTAGCTTTTTCCAATATTTTTCACC-3′	Amplification coding region, *pyrE*
PYRB	5′-CTGTTAGGGCAAATATGTCCTC-3′	Primer extention
PYRC	5′-GAGATCAACTGATGCAGGCAG-3′	Primer extention
PYRD	5′-GGTTCCAATGGTGAATAAGTTAGG-3′	Primer extention
PYRE	5′-GGCTAACTTTACCTGATGTTAAC-3′	Primer extention
ORF1	5′-GCCCAAGAGCTTTTTCCCCT-3′	Primer extention
ORF4	5′-CACAGAAAAAGGATCTATAATTGC-3′	Primer extention
1a	5′-GTGGACAAGACTACTAAGGC-3′	RT-PCR
1b	5′-CCTAAGACATTAAGTACTGCG-3′	RT-PCR
2a	5′-GAGATGTGAGTATTGTGAGAC-3′	RT-PCR
2b	5′-CGTAATCTGCAACTGTCAGG-3′	RT-PCR
3a	5′-GTTATAGCCTGTGTTGAGGGAC-3′	RT-PCR
3b	5′-GAGATCAACTGATGCAGGCAG-3′	RT-PCR
4a	5′-GATAGTTGAAGGTAAATTGGC-3′	RT-PCR
4b	5′-GGTTCCAATGGTGAATAAGTTAAC-3′	RT-PCR
5a	5′-CCAAGACCTCCTTCTAGCAATG-3′	RT-PCR
5b	5′-GTCAGTATATAAGGTAAGCG-3′	RT-PCR
10b	5′-CACAGAAAAAGGATCTATAATTGC-3′	RT-PCR

by centrifugation. The assay mixture contained 300 μmol of Tris-HCl (pH 8.8), 0.25 μmol of sodium orotate, 25 μmol of $MgCl_2$, and 0.4 μmol of 5′-phospho-α-D-ribosyl-pyrophosphate (PRPP) and bacterial extract in a final volume of 3 ml. The reaction was monitored by measuring the decrease in absorbance measured at 295 nm. Protein was measured by the method of Lowry et al. (37).

Nucleotide sequence accession number. The 8,519-bp-long sequence has been deposited in the EMBL data bank under accession number AJ459777.

RESULTS

Cloning and nucleotide sequence determination of the pyrimidine gene cluster of *S. acidocaldarius*. We previously reported the isolation of the *S. acidocaldarius pyrB* and *pyrI* genes on a 6.9-kb genomic *Pst*I fragment (Fig. 2) by complementation of an *E. coli* mutant deficient for ATCase (11). A Southern blot analysis performed under stringent conditions indicated that this fragment indeed originates from *S. acidocaldarius* (data not shown). The complete 6,845-nucleotide (nt)-long sequence was established, and subsequently extended with another 1,674 bp (Fig. 2, left-hand border *Eco*RI-*Pst*I segment) by sequencing two clones obtained by inverse PCR with *Sac*I- and *Eco*RI-digested genomic DNA, respectively. Sequence analysis of the 8,519-bp-long contig and comparisons with databases revealed the presence of 13 complete ORFs and two truncated ones, some of which are homologous to pyrimidine genes of other archaea, bacteria, and eukaryotes (see below). Besides the previously identified contiguous *pyrBI* genes (11), the fragment also contains the pyrimidine genes *pyrC*, *pyrD*,

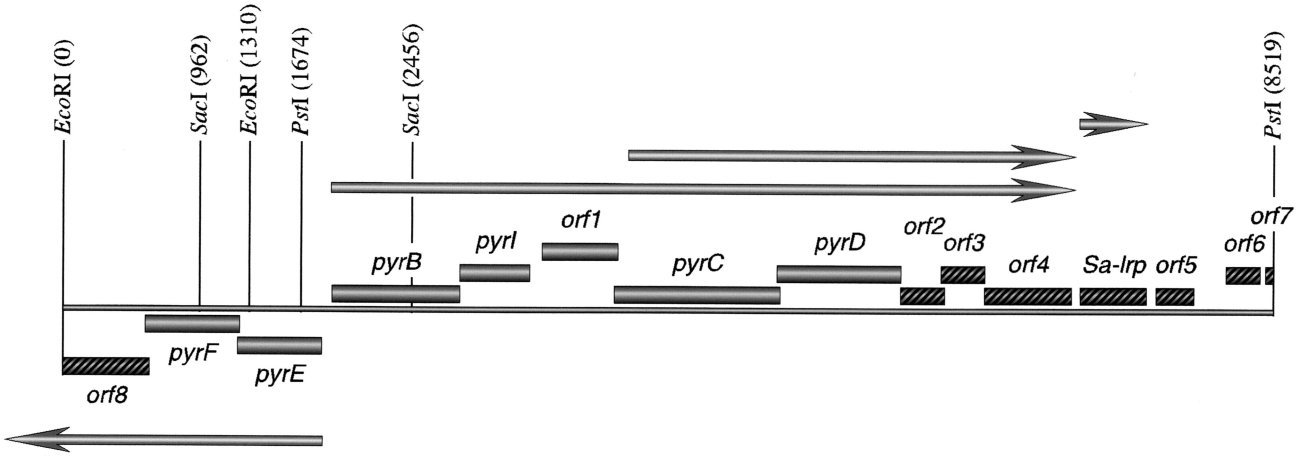

FIG. 2. Schematic drawing of the pyrimidine gene cluster of *S. acidocaldarius* and surrounding genes. Grey boxes represent pyrimidine biosynthetic genes. Cross-hatched boxes indicate potential ORFs with unknown function (*orf2* to *orf8*) and the transcriptional regulator Sa-Lrp. Arrows represent mRNA molecules and indicate the start, approximate stop, and orientation of transcription. Positions of the recognition sites for the restriction enzymes *Eco*RI, *Sac*I, and *Pst*I are indicated (in nucleotides).

FIG. 3. Nucleotide sequences of the border regions of adjacent genes in the divergent pyrimidine gene cluster. Potential start codons are indicated in bold capitals and are overlined; stop codons are indicated in bold capitals and underlined. Potential SD sequences are indicated in bold capitals.

pyrE, and *pyrF* encoding DHOase, dihydroorotate dehydrogenase (DHOdehase), OPRTase, and OMPdecase, respectively (Fig. 2). *pyrI* and *pyrC* are separated by the 177-amino-acid (aa)-long *orf1*, involved in the electron transport associated with the formation of orotate (see below). *pyrD* is followed by six relatively short ORFs (see below), one of which corresponds to *Sa-lrp*, encoding an Lrp-like archaeal transcription regulator (7, 13). Thus, the entire 8,519-bp-long *Eco*RI-*Pst*I segment of the *S. acidocaldarius* genome contains the information for the *pyrE*, *pyrF,* and truncated *orf8* gene products on one strand, and for the *pyrB*, *pyrI*, *orf1*, *pyrC*, *pyrD*, *orf2*, *orf3*, *orf4*, *Sa-lrp*, *orf5*, *orf6*, and truncated *orf7* gene products on the complementary strand (Fig. 2). The G+C content of the cluster is 35.3%, comparable to that of the entire genome (36 to 38%) (61, 24).

This divergent *pyr* gene cluster encodes the functional enzymes responsible for de novo UTP synthesis, as demonstrated by sequence analysis of mutant alleles and enzyme assays. The *pyrB* allele from mutant DG64, affected in ATCase activity (23), proved to bear a substitution of phenylalanine (TTT) for the strictly conserved proline (CCT, position 130), right in a highly conserved stretch (HPTQ) that is part of the active site. Similarly, we have shown that the *pyrE* alleles of mutants DG96 and MR450, affected in the OPRTase activity (23), bear the same single A-T base pair deletion in an A_7 stretch that appears to be a hot spot for *pyrE* mutations (22). The mutation results in a frame shift and the introduction of a premature stop codon (TGA) at position 191. Enzyme assays demonstrated that the *S. acidocaldarius pyrE* gene encodes a functional and thermophilic OPRTase. Indeed, cell extracts of *E. coli* strain HMS174(DE3)pLysS transformed with plasmid pET3PYRE, induced with isopropyl-β-D-thiogalactopyranoside (IPTG), and heated for 5 min at 55, 60, 70, or 80°C exhibited a threefold-higher OPRTase activity when assayed at 66°C than at 37°C, whereas no activity was measurable in extracts of noninduced controls (not shown). Moreover, the cloned *S. acidocaldarius pyrE* gene can complement a *pyrE* mutant of *Pyrococcus abyssi* for growth at 90°C (S. Lucas, L. Toffin, Y. Zivanovic, D. Charlier, H. Moussard, P. Forterre, D. Prieur, and G. Erauso, submitted for publication).

Amino acid sequence comparisons. The sequence analysis, enzyme purification, and characterization, as well as the phylogeny, of *S. acidocaldarius* ATCase have been published previously (11, 34). Briefly, *S. acidocaldarius* ATCase is allosterically regulated, belongs to the class B enzymes composed of catalytic (*pyrB*) and regulatory (*pyrI*) chains, and is clearly homologous to other ATCases of archaeal, bacterial, or eukaryotic origin. Now that new sequences are available, the best fits are with *S. tokodaii* (70 and 66% amino acid sequence identity for the catalytic and regulatory chains, respectively) and *S. solfataricus* (69 and 59%) followed by *P. aerophilum* (53 and 51%), *A. pernix* (50 and 35%), and *Pyrococcus horikoshii* (50 and 47%).

Sequence comparisons of active DHOases and silent DHOase domains of multifunctional enzymes indicate that DHOases fall into two main phylogenetically distinct groups (14). Type I includes the silent DHOase-like sequences of several fluorescent *Pseudomonas* species, the monofunctional active enzyme of mesophilic and thermophilic gram-positive bacteria, the active monofunctional *Ura2* gene product of *S. cerevisiae*, active multifunctional hamster CAD, and all the archaeal DHOases (monofunctional) identified till now. Type II enzymes are predominantly found in bacteria, gram-negatives and cyanobacteria (active enzymes), but also the inactive DHOase-like sequence embedded in the multifunctional *Ura4* gene product from *S. cerevisiae* belongs to the Type II. The best scores for *S. acidocaldarius* DHOase are with the crenarchaeotes: *S. tokodaii* (69% identity), *S. solfataricus* (57%), *A. pernix* (34%), and *P. aerophilum* (31%). These enzymes are clearly Type I enzymes (14), and *S. acidocaldarius* DHOase also shows the conserved stretch (consensus: PGLV) of 4 aa (PASV, positions 45 to 48), typical for Type I enzymes, that immediately precedes the aspartate residue of the first DHOase signature, DLHVHVRGA (positions 49 to 57 in *S. acidocaldarius* DHOase). Sequence analyses strongly suggest that *S. acidocaldarius pyrC* translation initiates at a TTG codon and overlaps

TABLE 3. Function and characteristics of ORFs in the 8,519-bp genomic segment of *S. acidocaldarius*

Gene	Function	Length of ORF (in aa)	Potential initiation codon	Stop codon	Potential SD sequence	Pu:Py ratio	G+C (mol %)	5′-overlap (in nt)	3′-overlap (in nt)
orf8	Hypothetical protein	202 (truncated)	ATG	?[a]	GAaGGTaA	1.164	36.2	26	?
pyrF	OMPdecase	219	TTG	TGA TAA	GGTGA	1.266	32.6	14	26
pyrE	OPRTase	197	ATG	TAG	None	1.309	36.2		14
pyrB	ATCase catalytic	299	TTG	TGA	None	1.342	35.0		
pyrI	ATCase regulatory	164	ATG	TGA	GGTGA	1.800	34.8		
orfI	DHOdehase electron transfer	177	TTG	TGA	TGATCC	1.414	36.4		29
pyrC	DHOase	389	TTG	TAA	AGGTGA	1.432	37.0	29	23
pyrD	DHOdehase catalytic	291	TTG	TGA	GAGGTG	1.473	37.6	23	2
orf2	Hypothetical protein	103	ATG	TAA	GAG-TGA	1.836	38.2	2	29
orf3	Hypothetical protein	102	ATG	TGA	GAGG	1.597	34.0	29	2
orf4	Conserved hypothetical	203	ATG	TGA	None	1.061	37.6	2	
Sa-lrp	Transcriptional regulator	145	ATG	TAA	None	1.450	31.8		
orf5	Hypothetical Zn finger	99	ATG	TAA	GGTGAT	2.000	40.0		
orf6	Hypothetical protein	80	ATG	TAA	None	2.076	28.8		
orf7	Hypothetical protein	18 (truncated)	ATG	?	GGTG	1.040	50.9		?

[a] ?, unknown.

the C terminus of *orf1* by 10 aa (Fig. 3), rather than at the in-phase ATG codon 99 (nt) downstream. Several arguments support this proposal: the TTG codon is preceded by a good Shine-Dalgarno (SD) sequence (Table 3), whereas the ATG codon is not; downstream initiation would result in a truncated DHOase missing 34 aa residues that show good sequence conservation with other DHOase sequences; moreover, this 102-nt-long stretch separating the TTG and ATG codons shows all the characteristics of a coding region (Pu/Py ratio, 2.90; G+C content, 34.3%) rather than of an intergenic part (see also below, gene organization in the divergent gene cluster) and finally, a similar overlap of PyrC with the C terminus of Orf1 appears to exist in *S. solfataricus*, though not in *S. tokodaii*.

The formation of orotate constitutes the sole redox reaction in de novo UMP biosynthesis. On the basis of phylogenetic studies and enzyme characterizations, DHOdehase sequences have been subdivided into two major groups (29, 53). Class I enzymes (predominantly present in gram-positive bacteria and archaea) are cytosolic homodimers (*pyrD* encoded) that use fumarate as the electron acceptor (type 1A) or heterotetramers (*pyrD-pyrK*, *pyrD-pyrDII*) (1, 30) that use NAD^+ as the electron acceptor (type 1B), whereas class II enzymes (gram-negative bacteria, higher eukaryotes) are monomeric enzymes attached to the membrane that use different kinds of quinones. Class I and class II enzymes use a different amino acid as the catalytic base, cysteine and serine, respectively. Sequence alignments and experimental data indicate that *S. solfataricus* (53) and also *S. acidocaldarius* (by sequence similarity only) DHOdehase belongs to the class I enzymes, but the *S. solfataricus* enzyme is unable to use any of the natural electron acceptors used in known DHOdehase types and uses serine as catalytic base, unique for a cytosolic DHOdehase (53). This serine residue (position 120 of *S. solfataricus* DHOdehase) is also conserved in the *S. acidocaldarius* (position 121) and *S. tokodaii* enzymes. The exact nature of the natural electron acceptor of the *Sulfolobus* enzymes is still unknown.

In gel filtration experiments, the catalytic subunit (*pyrD*) of *S. solfataricus* DHOdehase was found to comigrate with another protein that turned out to be the product of *orf1* (53). *orf1* shows only very weak amino acid sequence similarity to the *pyrK* (*pyrDII*)-encoded electron transfer subunits and is much shorter, but it bears an iron-sulfur cluster. Since *orf1* from *S. acidocaldarius* (177 aa) and that from *S. solfataricus* (208 aa) show 38.9% sequence identity (41% to *S. tokodaii 7* [197 aa]), *Sulfolobus* DHOdehases appear to be heteromeric enzymes constituted by a catalytic subunit and a new type of electron transfer subunit (53).

The best fits of *S. acidocaldarius* OPRTase are with the crenarchaeotes *S. tokodaii* (57% identity), *S. solfataricus* (55%), *A. pernix* (39%), and *P. aerophilum* (36%). The PRPP binding and active site is well conserved (VIVVDD-VATTGGS, residues 108 to 120 in the *S. acidocaldarius* enzyme).

The best fits of *S. acidocaldarius* OMPdecase are with archaeal homologues, the best being with *S. tokodaii* (60% identity), *S. solfataricus* (48%), *A. pernix* (43%), *M. jannashii* (32%), and *P. abyssi* (31%).

orf2 (encoding 103 aa), *orf3* (encoding 102 aa), and *orf6* (encoding 80 aa) encode very small, hypothetical proteins of unknown function. Since they are conserved and similarly organized in *S. solfataricus* and *S. tokodaii*, they are most likely expressed. The fact that *orf2*, *orf3*, and *orf4* are cotranscribed with the *pyr* genes (see below) and show the classical excess of purines on the coding strand and the typical G+C value of coding parts (Table 3) further supports this proposal. Sa-Lrp has been purified from the original host (13). Orf4 (204 aa) is well conserved among very divergent archaea and bacteria, but its function is unknown. Orf5 (99 aa) shows a Zn finger motif and is homologous to the hypothetical DNA-directed RNA polymerase subunit M of *S. solfataricus* and *S. tokodaii*. The 18-aa-long truncated Orf7 is homologous to the N termini of 506- and 489-aa-long conserved hypothetical proteins with unknown function from *S. solfataricus* and *S. tokodaii*, respectively. This ORF is embedded in the same genetic environment in *S. acidocaldarius* and *S. tokodaii*, but not in *S. solfataricus*. Similarly, the 203-aa-long truncated Orf8 is homologous to a hypothetical protein of unknown function from *S. solfataricus*

TABLE 4. Characteristics of intergenic regions

Location	Length (nt)	Pu/Py ratio	G+C (mol %)
pyrE-pyrB	71	1.03	26.8
pyrB-pyrI	1		
pyrI-orf1	85	1.073	34.1
orf4-Sa-lrp	60	1.14	18.3
Sa-lrp-orf5	66	0.784	24.2
orf5-orf6	133	0.847	27.8
orf6-orf7	96	1.133	31.4

(NP342138) and *S. tokodaii* (NP377445), but these are not contiguous to *pyrF*.

Gene organization in the divergent pyrimidine cluster. The genetic information comprised within the 8,519-bp stretch containing the divergent *pyr* cluster is extremely tightly packed. ORFs cover more than 94% of the sequence, and the gene density of 1.643 per kb largely exceeds those determined for the complete *S. solfataricus* P2 and *S. tokodaii 7* genomes (1.013 and 1.049, respectively). Considering only those genes clearly involved in pyrimidine biosynthesis, the density becomes 1.333 ORFs per kb, which is still high with respect to gene densities calculated for archaeal genomes. In seven instances, adjacent ORFs overlap over a short distance, varying between 2 and 29 nt (Table 3; Fig. 3). Of these, five show an overlap at both the N and C termini, resulting in 1.5% of sequence information being translated into two different proteins. In three other instances, the stop and start codons of contiguous ORFs are separated by a short nucleotide stretch of variable length comprised of between 1 and 133 nt (Table 4).

Of the 15 ORFs, 10 (66.7%) likely start with an ATG codon and five (33.3%) start with TTG, an observation in good agreement with the statistics on the utilization of ATG (65%), TTG (28%), and GTG (14%) as initiation codons in a 156-kb sequence of the related *S. solfataricus* (51). The most frequent stop codon utilized in the divergent gene cluster is TGA (53.8%), followed by TAA (38.5%) and TAG (7.7%) (Table 3); *pyrF* shows a pair of adjacent TGA and TAA stop codons. This occurrence of stop codons is unlike the situation reported for the 156-kb sequence of *S. solfataricus* (51) where TAA is predominant (50%). The coding parts show a purine bias (Pu/Py ratios varying between 1.164 and 2.076 for the individual genes; mean value of 1.416 for the 15 ORFs) (Table 3) close to the typical 55% A+G (ratio of 1.222) found in all sequenced genomes, including archaea, bacteria, and eukaryotes, except for *orf4* (Pu/Py ratio of 1.061). In contrast, the intergenic regions with Pu/Py ratios of between 0.78 and 1.133 do not show this bias (Table 4).

Experimental identification of the transcription starts indicate that the 5′ mRNA leader sequences of the two polycistronic *pyr* messengers (*pyrE-F-orf8* and *pyrB-I-orf1-pyrC-D-orf2-orf3-orf4*; see below) and of monocistronic *Sa-lrp* (13) are very short and do not harbor a ribosome binding site upstream of the initiation codon. A similar situation might prevail for *orf6*. Within the polycistronic mRNAs (see below), however, the situation is different and the initiation codon is nearly always preceded at an appropriate distance by a sequence that shows extensive complementarity to the 3′ end of the 16S rRNA (Table 3; Fig. 3). This is not the case for *orf4*, but as the ATG initiation codon overlaps the stop codon of *orf3*, translation might possibly proceed without the need for reassociation of ribosomes on the polycistronic messenger.

Polycistronic messenger analyses and potential transcription stop signals. The existence of polycistronic *pyr* messengers was demonstrated by RT-PCR experiments (Fig. 4). To demonstrate that two or more genes are transcribed as a single mRNA molecule, we used total RNA as template, reverse transcriptase, and pairs of oligonucleotides hybridizing to complementary strands of different genes, in PCR amplification reactions. In the first step, the oligonucleotide complementary to the mRNA of the downstream gene was used to synthesize cDNA. In the second amplification reaction, the oligonucleotide complementary to the cDNA and located in a different gene was added (for details, see Materials and Methods). In the sole presence of RNA as a template, the amplification reaction will be possible only if the mRNA molecule indeed spans the different genes. Thus, we have demonstrated that *pyrF* and *pyrE* are part of the same mRNA molecule, initiated at the *pyrE* promoter (see below for the initiation site). Indeed, using the pair of oligonucleotides 5a (hybridizing to the N-terminal part of the *pyrF* messenger, for cDNA synthesis) and 5b (hybridizing to the C terminus of *pyrE*, for further amplification), we detected the expected 226-bp fragment (Fig. 4a and b, lane 9), whereas no amplification at all was detected in the absence of reverse transcriptase in the assay (lane 10). Similarly, we have demonstrated cotranscription of *pyrI* and *pyrB* (oligonucleotides 1b and 1a; 204-bp fragment, lanes 1 and 2), of *orf1* and *pyrI* (oligonucleotides 2b and 2a; 231-bp fragment, lanes 3 and 4), of *pyrC* and *orf1* (oligonucleotides 3b and 3a; 193-bp fragment, lanes 5 and 6), and of *pyrD* and *pyrC* (oligonucleotides 4b and 4a; 217-bp fragment, lanes 7 and 8). In further combinations we demonstrated cotranscription of *pyrI*, *orf1*, and *pyrC* (oligonucleotides 3b and 2a, 816-bp fragment; Fig. 4c, lanes 3 and 4), of *pyrB*, *pyrI*, *orf1*, and *pyrC* (oligonucleotides 3b and 1a; 1,329-bp fragment, Fig. 4c, lanes 1 and 2) and of *pyrC*, *pyrD*, *orf2*, *orf3*, and *orf4* (oligonucleotides 10b and 4a; 1,600-bp fragment, Fig. 4d, lanes 1 and 2). These various combinations already strongly suggested that transcription initiated at the *pyrB* promoter (see below) might proceed till the end of *orf4*. This was further proven by an RT-PCR experiment using the pair of oligonucleotides 10b and 1a, which produced the expected 3,861-bp fragment (Fig. 4e). Previously, members of our group had shown that the *Sa-lrp* gene is transcribed as a monocistronic messenger that most likely stops at the type I transcriptional stop (5′-TTTTTATT) located 1 nt downstream of the TAA stop codon and that there is only very little readthrough from *orf4* into *Sa-lrp* (13). Thus, the bipolar pyrimidine operon from *S. acidocaldarius* is transcribed as two polycistronic messengers initiated from the divergent pair of *pyrE* and *pyrB* promoters (see below). Transcription initiated at the *pyrB* promoter does not proceed into *Sa-lrp* and most likely stops at the type I transcriptional stop signal 5′-TTTTTT, located 22 nt downstream of the TGA stop codon of *orf4* (13). Similar thymine stretches (5′-TTTTCTT and 5′-TTTTATTTT) are found 70 and 68 nt downstream of the TAA stop codons of *orf5* and *orf6*, respectively. Such a run af T residues is not present in the region 3′ of the *pyrF*, *pyrE*, *pyrB*, *pyrI*, *orf1*, *pyrC*, *pyrD*, *orf2*, and *orf3* genes and occurs in four instances only in the 7,636-nt-

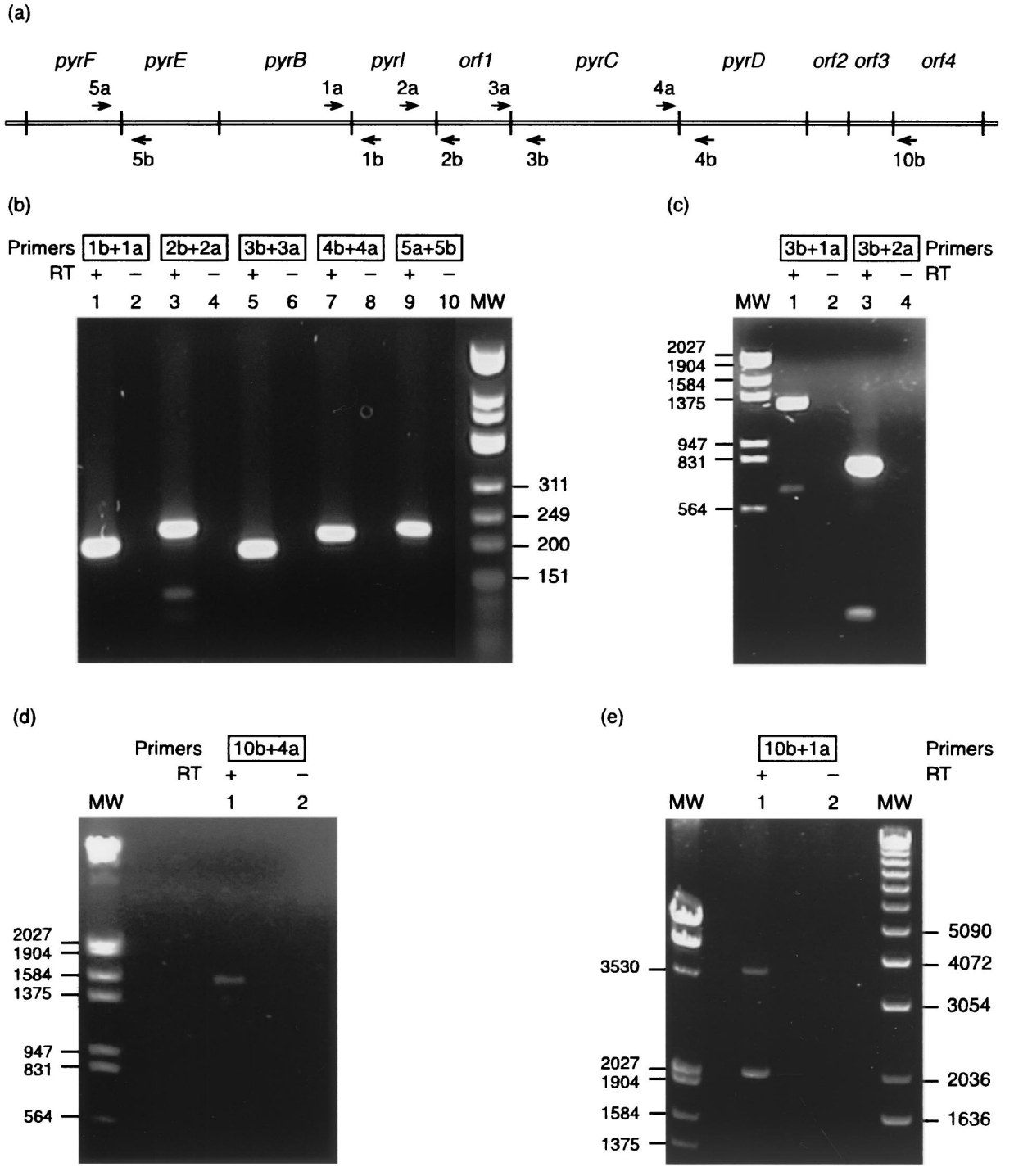

FIG. 4. (a) Schematic presentation of the bipolar pyrimidine operon. The approximate position of oligonucleotides used as primers in the RT-PCRs and their polarity are indicated by small arrows. (b, c, d, and e) Analysis by agarose gel electrophoresis of double-stranded DNA fragments generated in the RT-PCRs performed *with S. acidocaldarius* total RNA, Expand reverse transcriptase (when indicated), *Pfu* DNA polymerase, and different combinations of primers, as indicated. MW, molecular size markers.

long coding part of the cluster (one in *pyrF*, two in *pyrB*, one in *orf4*; they are far more numerous on the noncoding strand). The presence of these thymine runs at the 3′ end of mono- and polycistronic messengers appears therefore statistically significant. No potential transcriptional stop signal (type I or type II) could be identified at the 3′ end of *pyrF*. This observation, and the physical overlap of the coding parts of *pyrF* and truncated *orf8*, suggest that transcription initiated at *pyrE* proceeds into

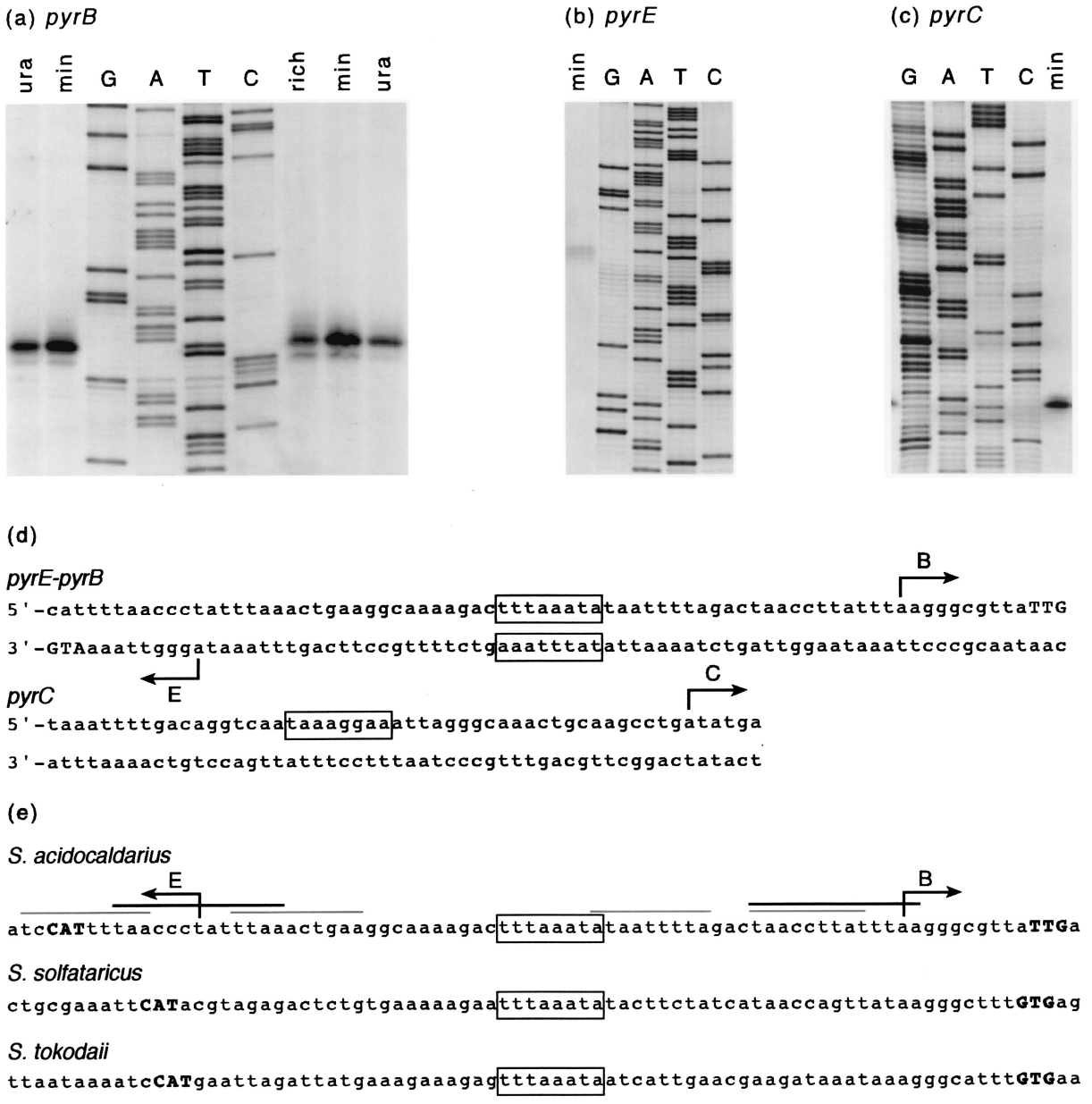

FIG. 5. (a) Quantitative determination of *pyrB* transcripts and mapping of the potential transcription initiation site by primer extension. Lanes 1 and 2, *pyrB* primer extension reactions with 100 μg of total RNA extracted from *S. acidocaldarius* cells grown in minimal medium supplemented with uracil (50 μg/ml) and minimal medium, respectively, and arrested in the exponential phase. Lanes 7 to 9, primer extension reactions with 100 μg of total RNA from independent cultures grown on complex medium, minimal medium, and minimal medium supplemented with uracil, as indicated. G, A, T, and C, chain terminating DNA sequencing reactions of the noncoding strand obtained with the same ^{32}P-labeled oligonucleotide used to perform the extension reactions. (b) Mapping of the potential transcription initiation site of *pyrE* by primer extension. The extension reaction was performed with 100 μg of total RNA extracted from cells grown in minimal medium. (c) Mapping of the potential transcription initiation site located within the *pyrC* gene. The extension reaction was performed with 100 μg of total RNA extracted from cells grown in minimal medium. (d) Nucleotide sequences of the *pyrE-pyrB* intergenic region and of the potential internal promoter located in the *pyrC* coding region. Arrows indicate the start site and orientation of transcription. Potential TATA promoter elements are boxed. (d) Comparison of the *pyrE-pyrB* intergenic regions of *S. acidocaldarius*, *S. solfataricus*, and *S. tokodaii*. Sequences have been aligned with respect to the strictly conserved TATA box elements. The most important direct repeats (black bars) and imperfect palindromic sequences (grey bars) present in the *S. acidocaldarius* sequence are indicated. Translation initiation codons are in bold capitals (complementary strand for *pyrE*), and arrows indicate the start sites and orientations of transcription.

orf8. In contrast, *orf5* and *orf6* might produce monocistronic messengers.

Transcription initiation and promoter sequences in the bipolar pyrimidine operon. Start points of transcription were determined by primer extension using total RNA extracted from *S. acidocaldarius* cells grown in minimal medium and 5′-^{32}P-end-labeled oligonucleotides complementary to the *pyrE*, *pyrB*, *pyrC*, *pyrD*, *orf1*, and *orf4* messengers (Table 2). The autoradiograph (Fig. 5a) indicates that *pyrB* transcription is initiated with an A residue, 10 nt upstream of the TTG initi-

ation codon (Fig. 5d). Similarly, *pyrE* transcription, proceeding in the opposite direction, was found to initiate with an A residue 9 nt upstream of the ATG initiation codon (Fig. 5b). Consequently, the start sites for divergent transcription are separated by a 52-nt-long stretch only (Fig. 5d). Its G+C content of 21.2% (26.8% for the extended 71-nt-long noncoding region) is considerably lower than that of the complete cluster (35.3%) but comparable to the mean value of 25% for intergenic regions in *S. solfataricus* (50). No transcription start could be identified in front of *orf1*, *pyrD*, and *orf4* (not shown), but a cDNA product corresponding to a potential initiation at an A residue within the *pyrC* coding part was detected (Fig. 5c and d). The start site of the monocistronic *Sa-lrp* gene was determined previously (7, 13); transcription initiation of *orf5*, *orf6*, and of the truncated *orf7* and *orf8* has not been investigated. Three of the identified transcripts (*pyrE*, *pyrB*, and *Sa-lrp*) initiate with a purine residue that is preceded by a thymine, thus creating the pyrimidine-purine dinucleotide required for correct start site selection of archaeal promoters (25). A good match to the consensus TATA box of archaeal promoters is the 5′-TTTAAATA sequence, ideally centered around −26.5 nt upstream of the *pyrB* start site (Fig. 5d). This sequence is preceded (positions −36 to −31) by a sequence stretch, 5′-AAAGAC, that shows good sequence conservation with the transcription factor *B* responsive element sequence of archaeal promoters, a purine-rich stretch with a consensus RNWAAW (R = purine; W = A or T) (2). The potential TATA box element of the *pyrE* promoter (5′-TATTTAAA), comprising between −23 and −30 upstream of the *pyrE* transcription start, largely overlaps the TATA box element of the *pyrB* promoter (Fig. 5d). Therefore, this region is apparently shared by the divergent pair of *pyrE* and *pyrB* promoters. The hypothetical TATA box element (5′-TAAAGGAA) comprising between −23 and −30 nt upstream of the potential *pyrC* transcription start showed only weak sequence identity with the consensus (Fig. 5d). Therefore, the detected cDNA product might correspond to a discrete degradation product of the longer mRNA molecule initiated at *pyrB*, or it could result from a premature arrest of the reverse transcriptase. The lack of a stable hairpin structure in the region makes the latter unlikely, however.

Regulation of *pyrB* transcription initiation. To determine the effect of nutrient composition on the abundance of *pyrB* transcription, we performed primer extension experiments with total RNA extracted from *S. acidocaldarius* cells grown on complex medium or on minimal medium either devoid of or supplemented with uracil. The densitometric analysis of the autoradiograph (Fig. 5a) indicated that the *pyrB* transcript is about 2.0-fold more abundant in the absence than in the presence of uracil and about 2.5-fold more abundant on minimal medium than on complex medium (mean for two assays performed with independent RNA preparations). Similarly, an approximately twofold difference was also observed in the ATCase specific activities measured in cell extracts of *S. acidocaldarius* cells grown in the presence or absence of uracil. Therefore, expression of the *S. acidocaldarius* pyrimidine biosynthetic genes is regulated, at least in part, at the transcriptional level.

DISCUSSION

In this study we present the sequence determination and analysis of an 8,519-bp segment of the *S. acidocaldarius* genome carrying an extremely tightly packed divergent pyrimidine gene cluster that exhibits a novel organization. Primer extension and RT-PCR experiments indicate that the pyrimidine genes are transcribed as two polycistronic messengers, *pyrBI-orf1-pyrCD-orf2-orf3-orf4* and *pyrEF*. The latter most likely also includes *orf8*, as suggested by the physical overlap of the *pyrF* and hypothetical *orf8* cistrons and the lack of an identifiable transcription stop signal at the 3′ end of *pyrF*. A similar clustering of *pyr* genes occurs in the entirely sequenced genomes of *S. solfataricus* P2 (52) and *S. tokodaii* strain 7 (32) and appears therefore characteristic of *Sulfolobus*. Beyond the *orf5* and *pyrF* genes, however, the similarity in genome organization vanishes. Also novel is the gene order *orf1-pyrC-pyrD*, in which *pyrC* is intercalated between the electron transfer and catalytic subunit of DHOdehase; in bacteria, *pyrK* (*pyrDII*) and *pyrD* are generally contiguous.

The seven cistrons involved in de novo synthesis of UMP are tightly packed, and only *pyrB* and *pyrI*, encoding the catalytic and regulatory subunits of ATCase, respectively, show no overlap at either the 5′ or 3′ end: *pyrB* and *pyrI* are separated by a single nucleotide, and *pyrI* and *orf1* are separated by 85 nt. The start points of the divergent transcripts initiated at the *pyrE* and *pyrB* promoters are separated by 52 nt only, and the corresponding initiation codons are separated by 71 nt. This short intergenic region appears to contain a pair of overlapping divergent promoters showing a good match to the consensus TATA and transcription factor *B* responsive element promoter elements. This peculiar situation constitutes an interesting model system for the study of basal and regulated transcription in hyperthermophilic archaea, still a poorly documented area of investigation. Indeed, the intimate intertwining of divergent promoters creates an interesting test case for the analysis of sequence determinants imposing transcription polarity, for the study of mutual influences of nearby promoters on binding of transcription factors (TBP, TFB, and TFE) and polymerase recruitment, and for the analysis of variations in template topology generated by the moving polymerases engaged in divergent transcription (positive supercoiling in front, negative ahead) on transcription initiation frequencies. This last aspect is particularly interesting, since DNA in hyperthermophiles is known to be relaxed to slightly positively supercoiled, in contrast to the negative supercoiling of mesophilic genomes.

We have shown that *pyr* gene expression in *S. acidocaldarius* is regulated at least in part at the transcriptional level. *pyrB* mRNA levels were found to differ by a factor of 2 in minimal medium devoid of or supplemented with uracil. On the other hand, ATCase enzyme activities measured in cultures of various pyrimidine auxotrophs grown on limiting amounts of uracil showed a 2- to 10-fold derepression (23). Taken together, these data suggest the existence of a several-fold regulation of transcription initiation that would provide *S. acidocaldarius* with the flexibility required to attune pyrimidine biosynthesis to the cellular needs and fluctuations in available nutrients. The mode of pyrimidine-specific repression of *pyrB* transcrip-

tion initiation in *S. acidocaldarius* is not known, but the mechanism appears to be different from the various types of attenuation-like mechanisms operative in mesophilic and thermophilic bacteria. The *pyrE* and *pyrB* leader sequences are too short to contain attenuator-antiattenuator structures. In contrast, the short *pyrE-pyrB* intergenic region exhibits features (direct repeats and imperfect palindromic sequences) characteristic of potential binding sites for a hypothetical transcriptional regulator(s), which are at least in part conserved in *S. solfataricus* and *S. tokodaii* (Fig. 5e).

The quasi-leaderless *pyrB* and *pyrE* messengers bear no SD site, whereas in contrast, the ORFs embedded in the polycistronic messengers show a rather extensive complementarity to the 3′ end of 16S rRNA (Table 3). Interestingly, it was previously shown that the disruption of the SD sequence abolished translation, but most importantly, the effect of this disruption could be suppressed by deleting the 5′-untranslated region, thus creating a leaderless messenger (9). In all organisms, bacteria, archaea, and eucarya, leaderless messengers exist, even for abundant proteins (28). Translation initiation of the archaeal leaderless mRNAs might require the formation of a complex equivalent to the bacterial 30S-fMet-tRNA$_f^{Met}$-IF2 intermediate implicated in translation of such mRNAs in *E. coli*, as is suggested by the faithful in vitro translation of leaderless λ *cI* mRNA with an archaeal translation system (21).

The bipolar *pyr* operon of *S. acidocaldarius* does not harbor the genes encoding carbamoylphosphate synthetase. Studies with mutants indicate that *S. acidocaldarius* has only one carbamoylphosphate synthetase (23), and on the *S. solfataricus* P2 genome the adjacent *carA* and *carB* genes encoding the sole CPSase of this organism are embedded in an arginine gene cluster (52). This peculiar situation raises interesting questions regarding the regulation of CPSase activity in relation to the cellular needs for arginine and pyrimidine biosynthesis. A comparable situation exists in *L. lactis*, where the *carA* and *carB* genes encoding the sole CPSase are dispersed; *carA* is part of a *pyr* gene cluster, whereas *pyrB* is monocistronic, but both are submitted to pyrimidine regulation (38, 39).

Thermophilic OPRTases and DHOdecases deserve special attention. Mutants deficient in either one of these activities are resistant to 5-fluoroorotic acid and auxotrophic for uracil. Therefore, both the wild-type and mutant *pyrE* and *pyrF* alleles can be positively selected. These unique traits and the universal and ancient character of pyrimidine biosynthesis make the utilization of the *pyrE* and *pyrF* genes as genetic markers particularly attractive. *S. acidocaldarius pyrE* has proven to be particularly suited for studies on intragenic recombination, chromosomal marker exchange by archaeal conjugation, and genetic fidelity at high temperature (16, 22, 27, 46, 48). Furthermore, the gene has also been used for the construction of cloning vectors for the hyperthermophilic euryarchaeote *Pyrococcus abyssi*, where the *S. acidocaldarius* gene was chosen to avoid homologous recombination (S. Lucas et al., submitted).

ACKNOWLEDGMENTS

We are grateful to D. Grogan for the gift of strains and to J.-P. Ten Have for the artwork.

This work was supported by the Fund for Joint Basic Research-Flanders (FWO-Vlaanderen, contract no. G.0069.00) and the Flanders Interuniversity Institute for Biotechnology (VIB).

REFERENCES

1. **Andersen, P. S., J. Martinussen, and K. Hammer.** 1996. Sequence analysis and identification of the *pyrKDbF* operon from *Lactococcus lactis* including a novel gene, *pyrK*, involved in pyrimidine biosynthesis. J. Bacteriol. **178:** 5005–5012.
2. **Bell, S. D., P. L. Kosa, P. S. Sigler, and S. P. Jackson.** 1999. Orientation of the transcription preinitiation complex in *Archaea*. Proc. Natl. Acad. Sci. USA **96:**13662–13667.
3. **Bonner, E. R., J. N. D'Elia, B. K. Billips, and R. L. Switzer.** 2001. Molecular recognition of *pyr* mRNA by the *Bacillus subtilis* attenuation regulatory protein PyrR. Nucleic Acids Res. **29:**4851–4865.
4. **Brock, T. D., K. M. Brock, R. T. Belly, and R. L. Weiss.** 1972. *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch. Microbiol. **84:**54–68.
5. **Charlier, D., A. Kholti, N. Huysveld, D. Gigot, D. Maes, T.-L. Thia-Toong, and N. Glansdorff.** 2000. Mutational analysis of *Escherichia coli* PepA, a multifunctional DNA-binding aminopeptidase. J. Mol. Biol. **302:**411–426.
6. **Charlier, D., M. Roovers, D. Gigot, N. Huysveld, A. Piérard, and N. Glansdorff.** 1993. Integration host factor (IHF) modulates the expression of the pyrimidine-specific promoter of the *carAB* operons of *Escherichia coli* K12 and *Salmonella typhimurium* LT2. Mol. Gen. Genet. **237:**273–286.
7. **Charlier, D., M. Roovers, T.-L. Thia-Toong, V. Durbecq, and N. Glansdorff.** 1997. Cloning and identification of the *Sulfolobus solfataricus lrp* gene encoding an archaeal homologue of the eubacterial leucine-responsive global transcriptional regulator Lrp. Gene **201:**63–68.
8. **Chomczynski, P.** 1993. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cells and tissue samples. BioTechniques **15:**532–534.
9. **Condò, I., A. Ciammaruconi, D. Bellini, D. Ruggero, and P. Londei.** 1999. *Cis*-acting signals controlling translation initiation in the thermophilic archaeon *Sulfolobus solfataricus*. Mol. Microbiol. **34:**377–384.
10. **Dagert, M., and S. D. Ehrlich.** 1979. Prolonged incubation in calcium chloride improves the competence of *E. coli* cells. Gene **6:**23–28.
11. **Durbecq, V., T.-L. Thia-Toong, D. Charlier, V. Villeret, M. Roovers, R. Wattiez, C. Legrain, and N. Glansdorff.** 1999. Aspartate carbamoyltransferase from the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. Eur. J. Biochem. **264:**233–241.
12. **Elagöz, A., A. Abdi, J.-C. Hubert, and B. Kammerer.** 1996. Structure and organisation of the pyrimidine biosynthesis pathway genes in *Lactobacillus plantarum*: a PCR strategy for sequencing without cloning. Gene **182:**37–43.
13. **Enoru-Eta, J., D. Gigot, T.-L. Thia-Toong, N. Glansdorff, and D. Charlier.** 2000. Purification and characterization of Sa-Lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon *Sulfolobus acidocaldarius* homologous to the bacterial global transcription regulator Lrp. J. Bacteriol. **182:**3661–3672.
14. **Fields, C., D. Brichta, M. Shepherdson, M. Farinha, and G. O'Donovan.** 1999. Phylogenetic analysis and classification of dihydroorotases: a complex history for a complex enzyme. Paths Pyrimidines **7:**49–63.
15. **Fitz-Gibbon, S. T., H. Ladner, U.-J. Kim, K. O. Stetter, M. I. Simons, and J. H. Miller.** 2002. Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. Proc. Natl. Acad. Sci. USA **99:**984–989.
16. **Ghané, F., and D. W. Grogan.** 1998. Chromosomal marker exchange in the thermophilic archaeon *Sulfolobus acidocaldarius*: physiological and cellular aspects. Microbiology **144:**1649–1657.
17. **Ghim, S.-Y., and J. Neuhard.** 1994. The pyrimidine biosynthesis operon of the thermophile *Bacillus caldolyticus* includes genes for uracil phosphoribosyltransferase and uracil permease. J. Bacteriol. **176:**3698–3707.
18. **Ghim, S.-Y., P. Nielsen, and J. Neuhard.** 1994. Molecular characterization of pyrimidine biosynthesis genes from the hyperthermophile *Bacillus caldolyticus*. Microbiology **140:**479–491.
19. **Glansdorff, N.** 1965. Topography of cotransducible arginine mutations in *Escherichia coli* K12. Genetics **51:**167–179.
20. **Glansdorff, N., D. Charlier, P. Chen, A. Kholti, and Y. Xu.** 1998. Genetic studies on the pyrimidine paradigm in prokaryotes: new insights in regulatory mechanisms. Paths Pyrimidines **6:**53–63.
21. **Grill, S., C. O. Gualerzi, P. Londei, and U. Bläsi.** 2000. Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. EMBO J. **19:**4101–4110.
22. **Grogan, D. W., G. T. Carver, and J. W. Drake.** 2001. Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. Proc. Natl. Acad. Sci. USA **98:**7928–7933.
23. **Grogan, D. W., and R. P. Gunsalus.** 1993. *Sulfolobus acidocaldarius* synthesizes UMP via a standard *de novo* pathway: results of a biochemical-genetic study. J. Bacteriol. **175:**1500–1507.
24. **Grogan, D. W., P. Palm, and W. Zillig.** 1990. Isolate B12, which harbours a virus-like element, represents a new species of the archaebacterial genus *Sulfolobus*. Sulfolobus shibatae, sp. nov. Arch. Microbiol. **154:**594–599.
25. **Hain, J., W.-D. Reiter, U. Hüdepohl, and W. Zillig.** 1992. Elements of an archaeal promoter defined by mutational analysis. Nucleic Acids Res. **20:** 5423–5428.

26. **Han, X., and C. L. Turnbough, Jr.** 1998. Regulation of *carAB* expression in *Escherichia coli* occurs in part through UTP-sensitive reiterative transcription. J. Bacteriol. **180:**705–713.
27. **Jacobs, K. L., and D. W. Grogan.** 1998. Spontaneous mutation in a thermoacidophilic archaeon: evaluation of genetic and physiological factors. Arch. Microbiol. **169:**81–83.
28. **Janssen, G. R.** 1993. Eubacterial, archaebacterial and eukaryotic genes that encode leaderless mRNA, p. 59–67. *In* R. H. Baltz, G. D. Hegeman, and P. L. Skatrud (ed.), Industrial microorganisms: basic and applied molecular genetics. American Society for Microbiology, Washington, D.C.
29. **Jensen, K. F., and O. Bjornberg.** 1998. Evolutionary and functional families of dihydroorotate dehydrogenases. Paths Pyrimidines **6:**20–28.
30. **Kahler, A. E., and R. L. Switzer.** 1996. Identification of a novel gene of pyrimidine nucleotide biosynthesis, *pyrDII*, that is required for dihydroorotate dehydrogenase activity in *Bacillus subtilis*. J. Bacteriol. **178:**5013–5016.
31. **Kawarabayasi, Y., Y. Hino, H. Horikawa, S. Yamazaki, Y. Haikawa, K. Jin-No, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, H. Nakazawa, M. Takamiya, S. Masuda, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Ogushi, K. Aoki, K. Kubota, Y. Nakamura, N. NOmura, Y. Sako, and H. Kikuchi.** 1999. Complete genome sequence of an aerobic hyperthermophilic crenarchaeon, Aeropyrum pernix K1. DNA Res. **6:**83–101.
32. **Kawarabayasi, Y., Y. Hino, H. Horikawa, K. Jin-No, M. Takahashi, M. Sekine, S.-I. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishiijima, R. Otsuka, H. Nakasawa, M. Takamiya, Y. Kato, T. Yoshizawa, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K.-I. Aoki, S. Masuda, M. Yanagii, M. Nishimura, A. Yamagishi, T. Oshimla, and H. Kikuchi.** 2001. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. DNA Res. **8:**123–140.
33. **Kholti, A., D. Charlier, D. Gigot, N. Huysveld, and N. Glansdorff.** 1998. *pyrH*-encoded UMP-kinase directly participates in pyrimidine-specific modulation of promoter activity in *Escherichia coli*. J. Mol. Biol. **280:**571–582.
34. **Labedan, B., A. Boyen, M. Baetens, D. Charlier, P. Chen, R. Cunin, V. Durbecq, N. Glansdorff, G. Hervé, C. Legrain, Z. Liang, C. Purcarea, M. Roovers, R. Sanchez, T.-L. Thia-Toong, M. Van de Casteele, F. Van Vliet, Y. Xu, and Y.-F. Zhang.** 1999. The evolutionary history of carbamoyltransferases: a complex set of paralogous genes was already present in the last universal common ancestor. J. Mol. Evol. **49:**461–473.
35. **Legrain, C., M. Demarez, N. Glansdorff, and A. Piérard.** 1995. Ammonia-dependent synthesis and metabolic channeling of carbamoyl phosphate in the hyperthermophilic archaeon *Pyrococcus furiosus*. Microbiology **141:**1093–1099.
36. **Li, X., G. M. Weinstock, and B. E. Murray.** 1995. Generation of auxotrophic mutants of *Enterococcus faecalis*. J. Bacteriol. **177:**6866–6873.
37. **Lowry, O. H., N. J. Rosenbrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurements with folin phenol reagent. J. Biol. Chem. **193:**265–275.
38. **Martinussen, J., and K. Hammer.** 1998. The *carB* gene encoding the large subunit of carbamoylphosphate synthetase from *Lactococcus lactis* is transcribed monocistronically. J. Bacteriol. **180:**4380–4386.
39. **Martinussen, J., J. Schallert, B. Andersen, and K. Hammer.** 2001. The pyrimidine operon *pyrRPB-carA* from *Lactococcus lactis*. J. Bacteriol. **183:** 2758–2794.
40. **Massant, J., P. Verstreken, V. Durbecq, A. Kholti, C. Legrain, S. Beeckmans, P. Cornelis, and N. Glansdorff.** 2002. Metabolic channeling of carbamoyl phosphate, a thermolabile intermediate: evidence for physical interaction between carbamate kinase-like carbamoyl phosphate synthetase and ornithine carbamoyltransferase from the hyperthermophile *Pyrococcus furiosus*. J. Biol. Chem. **277:**18517–18522.
41. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101:**499–560.
42. **Neuhard, J., and R. A. Kelln.** 1996. Biosynthesis and conversion of pyrimidines, p. 580–599. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
43. **Purcarea, C., D. R. Evans, and G. Hervé.** 1999. Channeling of carbamoyl phosphate to the pyrimidine and arginine biosynthetic pathways in the deep sea hyperthermophilic archaeon *Pyrococcus abyssi*. J. Biol. Chem. **274:**6122–6129.
44. **Purcarea, C., G. Hervé, R. Cunin, and D. R. Evans.** 2001. Cloning, expression, and structure analysis of carbamate kinase-like carbamoyl phosphate synthetase from *Pyrococcus abyssi*. Extremophiles **5:**229–239.
45. **Quinn, C. L., B. T. Stephenson, and R. L. Switzer.** 1991. Functional organization and nucleotide sequence of the *Bacillus subtilis* pyrimidine biosynthetic operon. J. Biol. Chem. **266:**9113–9127.
46. **Reilly, M. S., and D. W. Grogan.** 2001. Characterization of intragenic recombination in a hyperthermophilic archaeon via conjugational DNA exchange. J. Bacteriol. **183:**2943–2946.
47. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5469.
48. **Schmidt, K. J., K. E. Beck, and D. W. Grogan.** 1999. UV stimulation of chromosomal marker exchange in *Sulfolobus acidocaldarius*: implications for DNA repair, conjugation and homologous recombination at extremely high temperatures. Genetics **152:**1407–1415.
49. **Schurr, M. J., J. F. Vickrey, A. P. Kumar, A. L. Campbell, R. Cunin, R. C. Benjamin, M. S. Shanley, and G. A. O'Donovan.** 1995. Aspartate transcarbamoylase genes of *Pseudomonas putida*: requirement for an inactive dihydroorotase for assembly into the dodecameric holoenzyme. J. Bacteriol. **177:**1751–1759.
50. **Sensen, C. W., R. L. Charlebois, C. Chow, Ib G. Clausen, B. Curtis, W. F. Doolittle, M. Duguet, G. Erauso, T. Gaasterland, R. A. Garrett, P. Gordon, I. Heikamp de Jong, A. C. Jeffries, C. Kozera, N. Medina, A. De Moors, J. van der Oost, H. Phan, M. A. Ragan, M. E. Schenk, Q. She, R. K. Singh, and N. Tolstrup.** 1998. Completing the sequence of the *Sulfolobus solfataricus* P2 genome. Extremophiles **2:**305–312.
51. **Sensen, C. W., H.-P. Klenk, R. K. Singh, G. Allard, C. C.-Y. Chan, Q. Y. Liu, S. L. Penny, F. Young, M. E. Schenk, T. Gaasterland, W. F. Doolittle, M. A. Ragan, and R. L. Charlebois.** 1996. Organizational characteristics and information content of an archaeal genome: 156 kb of sequence from *Sulfolobus solfataricus* P2. Mol. Microbiol. **22:**175–191.
52. **She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C.-Y. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. K. Gordon, I. Heinkamp-de Jong, A. C. Jeffries, C. J. Kozeka, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. A. Ragan, C. W. Sensen, and J. Van der Oost.** 2001. The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc. Natl. Acad. Sci. USA **98:**7835–7840.
53. **Sørensen, P. G.** 2001. A new type of dihydroorotate dehydrogenase-type 1S-from the thermoacidophilic archaeon *Sulfolobus solfataricus*. Ph.D. thesis. University of Copenhagen, Copenhagen, Denmark.
54. **Switzer, R. L., R. J. Turner, and Y. Lu.** 1999. Regulation of the *Bacillus subtilis* pyrimidine biosynthetic operon by transcriptional attenuation: control of gene expression by a mRNA-binding protein. Prog. Nucleic Acid Res. Mol. Biol. **62:**329–367.
55. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
56. **Turner, R. J., Y. Lu, and R. L. Switzer.** 1994. Regulation of the *Bacillus subtilis* pyrimidine biosynthetic (*pyr*) gene cluster by an autogenous transcriptional attenuation mechanism. J. Bacteriol. **176:**3708–3722.
57. **Van de Casteele, M., C. Legrain, L. Desmarez, P. G. Chen, A. Piérard, and N. Glansdorff.** 1997. Molecular physiology and carbamoylation under extreme conditions: what can we learn from extreme thermophilic microorganisms? Comp. Biochem. Physiol. **118A:**463–473.
58. **Van de Casteele, M., P. Chen, M. Roovers, C. Legrain, and N. Glansdorff.** 1997. Structure and expression of a pyrimidine gene cluster from the extreme thermophile *Thermus* strain ZO5. J. Bacteriol. **179:**3470–3481.
59. **Van de Casteele, M., L. Desmarez, C. Legrain, P. G. Chen, K. Van Lierde, A. Piérard, and N. Glansdorff.** 1994. Genes encoding aspartate carbamoyltransferase of *Thermus aquaticus ZO5* and *Thermotoga maritima* MSB8: modes of expression in *E. coli* and properties of their products. Biocatalysis **112:**135–179.
60. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertional mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.
61. **Zillig, W., K. O. Stetter, S. Wunderl, W. Schulz, H. Priess, and I. Scholz.** 1980. The *Sulfolobus-Caldariella* group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. Arch. Microbiol. **125:**259–269.