JOURNAL OF BACTERIOLOGY, June 2006, p. 4198–4206 Vol. 188, No. 12
0021-9193/06/$08.00+0 doi:10.1128/JB.00061-06
Copyright © 2006, American Society for Microbiology. All Rights Reserved.

Mutations and Rearrangements in the Genome of *Sulfolobus solfataricus* P2†

Peter Redder* and Roger A. Garrett

Danish Archaea Centre, Institute for Molecular Biology and Physiology, Copenhagen University, Sølvgade 83H, DK-1307 Copenhagen K, Denmark

Received 13 January 2006/Accepted 26 March 2006

The genome of *Sulfolobus solfataricus* P2 carries a larger number of transposable elements than any other sequenced genome from an archaeon or bacterium and, as a consequence, may be particularly susceptible to rearrangement and change. In order to gain more insight into the natures and frequencies of different types of mutation and possible rearrangements that can occur in the genome, the *pyrEF* locus was examined for mutations that were isolated after selection with 5-fluoroorotic acid. About two-thirds of the 130 mutations resulted from insertions of mobile elements, including insertion sequence (IS) elements and a single nonautonomous mobile element, SM2. For each of these, the element was identified and shown to be present at its original genomic position, consistent with a progressive increase in the copy numbers of the mobile elements. In addition, several base pair substitutions, as well as small deletions, insertions, and a duplication, were observed, and about one-fifth of the mutations occurred elsewhere in the genome, possibly in an orotate transporter gene. One mutant exhibited a 5-kb genomic rearrangement at the *pyrEF* locus involving a two-step IS element-dependent reaction, and its boundaries were defined using a specially developed "in vitro library" strategy. Moreover, while searching for the donor mobile elements, evidence was found for two major changes that had occurred in the genome of strain P2, one constituting a single deletion of about 4% of the total genome (124 kb), while the other involved the inversion of a 25-kb region. Both were bordered by IS elements and were inferred to have arisen through recombination events. The results underline the caution required in working experimentally with an organism such as *S. solfataricus* with a continually changing genome.

Studies of members of the genus *Sulfolobus*, which frequent terrestrial acidic hot springs (~80°C and pH ~2 to 3), have provided many new insights into the basic molecular processes of both crenarchaea and the *Archaea* in general. One of the organisms that has received particular attention is *Sulfolobus solfataricus* P2, for which the genome sequence is available (24). The genome is exceptional in that it contains a large number of potentially mobile elements, including over 200 copies of intact insertion sequence (IS) elements of at least 25 different types and more than 140 copies of miniature inverted-repeat transposable elements (MITEs), which have been grouped into four main classes (SM1 to SM4), as well as numerous fragmented elements (4, 5, 22, 24). Some types of IS elements and MITEs are present in multiple copies of identical or near-identical sequence, consistent with their having recently transposed within the genome (4, 22). In total, they constitute more than 11% of the 3.0-Mb *S. solfataricus* genome and are more prevalent than in any other sequenced archaeal or bacterial genome (4, 7).

Each of the IS elements of *S. solfataricus* P2 encodes a transposase, and many carry inverted terminal repeats to which the transposase attaches, while MITEs lack a transposase gene but exhibit inverted terminal repeats similar to those of an IS element (4, 22). The latter similarity is probably sufficient to ensure that the MITEs are mobilized by the corresponding IS element-encoded transposase. However, the internal sequences of the *S. solfataricus* P2 MITEs (type II) are different from those of the corresponding IS elements, suggesting that they evolved convergently (14, 21, 22).

The detailed transpositional mechanisms of the *Sulfolobus* elements, and those from other archaea, have received little attention experimentally, and they have generally been considered to be similar to those of their bacterial or eukaryal counterparts (4, 17). For the last two, transposition generally involves a double-strand excision from a given genomic position, with a subsequent nucleophilic attack by both 3′ ends at another target site, which results in the insertion of the element, all catalyzed by the encoded transposase, which binds as a multimer to the inverted terminal repeats. Attack on the two DNA strands often occurs a few base pairs apart and thereby produces a direct repeat (DR), constituting a short (2- to ~14-bp) duplication of the target site, which borders the newly inserted element (17).

Genome comparison studies have suggested that *S. solfataricus* P2 readily undergoes genomic rearrangements (5). Moreover, since it carries substantially more IS elements and MITEs than *Sulfolobus tokodaii*, or than *Sulfolobus acidocaldarius*, which probably lacks active mobile elements (4, 9, 11), it was proposed that the genomic rearrangements are driven primarily by mobile-element-dependent mechanisms (5). This proposal complements the earlier hypothesis that the megaplasmid pNRC100 of the euryarchaeon *Halobacterium halobium* arose via a multistep IS element-mediated process (10, 20).

The circumstantial evidence for high transpositional activity, and the possibility of mobile-element-induced genomic rear-

* Corresponding author. Present address: Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, Institut Pasteur, 25, rue Dr. Roux, 75724 Paris Cedex 15, France. Phone: (33)140613722. Fax: (33)145688834. E-mail: predder@pasteur.fr.

† Supplemental material for this article may be found at http://jb.asm.org/.

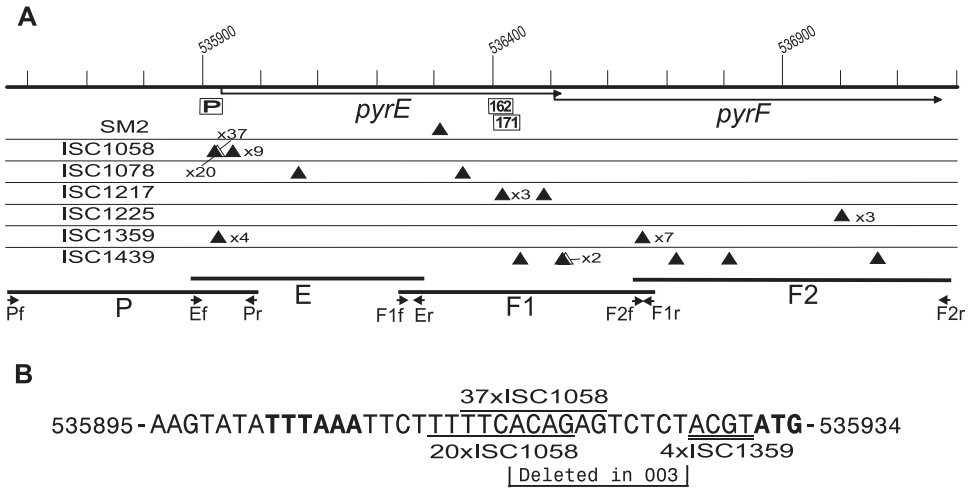

FIG. 1. (A) The 1,600-bp *pyrEF* locus, with *pyrE* and *pyrF* denoted by arrows. The PCR coverage (lower lines, P, E, F1, and F2) and mobile-element insertion sites (triangles) are shown. The numbers adjacent to the triangles indicate the total number of insertions observed at each site. The primer pairs Pf-Pr, Ef-Er, F1f-F1r, and F2f-F2r are indicated by small arrows. Boxed P, 162, and 171 mark the promoter region, the deletion in mutant P2A-162, and the duplication in P2A-171, respectively. (B) Sequence of the promoter region, with the TATA-like box and the *pyrE* start codon indicated in boldface; the bracket indicates the 9 bp deleted from strain P2A-003. Lines below and above the sequence denote the DR of the ISC*1058*s inserted at positions 535912 and 535914, respectively. The double line indicates the DR of ISC*1359* (P2A-003, -102, -150, and -177).

rangements occurring in *S. solfataricus* P2, raises the question as to how the organism remains viable. The complexity of this issue is compounded by the observation that a large number of IS element fragments are present in the genome, including some element types which are not present as intact copies. This suggests that they are difficult to delete cleanly (3, 4). However, although they are hard to remove, it was recently proposed that expression of the transposases is rigorously regulated via antisense RNAs (26), and several small putative antisense RNAs were detected in cell extracts of *S. solfataricus* using different experimental approaches (26, 27). Such a regulatory mechanism has also been shown to operate for some bacterial IS elements of the IS*4* family, to which some of the *Sulfolobus* elements also belong (4, 25).

In order to gain a clearer picture of the transpositional events and mobile-element-induced genomic rearrangements, as well as other types of mutational change that occur in *S. solfataricus* P2, we performed a comprehensive mutational study of the *pyrEF* locus under selective conditions. Mutants were generated which were resistant to 5-fluoroorotic acid (FOA), which is converted into the toxic compound 5-fluoro-UMP by the gene products of *pyrE* and *pyrF* (16). Earlier studies, using a similar approach, detected only a few IS element insertions, mainly in the *pyrE* gene or the promoter region for the related strain *Sulfolobus solfataricus* P1 (18), while for the more distantly related *S. acidocaldarius*, only point mutations, small insertions, and deletions were observed at this locus (11).

The large number of spontaneous knockout mutants isolated for *S. solfataricus* P2 were analyzed using a combination of PCR amplification and sequencing. For those mutations which involved transpositions, the donor site of the transposed element was always examined. Furthermore, the locations and sizes of genomic rearrangements were identified using a method developed specifically for determining the sequences beyond the junctions of the rearranged region.

MATERIALS AND METHODS

Isolation of mutant strains. The *Sulfolobus* medium contained 25 mM $(NH_4)_2SO_4$, 3 mM K_2SO_4, 1.5 mM KCl, 20 mM glycine, 4.0 μM $MnCl_2$, 10.4 μM $Na_2B_4O_7$, 0.38 μM $ZnSO_4$, 0.13 μM $CuSO_4$, 62 nM Na_2MoO_4, 59 nM $VOSO_4$, 18 nM $CoSO_4$, 19 nM $NiSO_4$, 0.72 μM $FeSO_4$, and 0.1 mM HCl. It was adjusted to pH 3 to 3.5 with H_2SO_4.

S. solfataricus P2 was grown at 80°C in liquid culture for 2 days to an optical density at 600 nm (OD_{600}) of 0.19 to 0.67 in medium supplemented with 2 g/liter Tryptone Peptone (Difco, Sparks, MD). The culture was then diluted to an OD_{600} of 0.0004 to 0.005 with fresh liquid medium, and 400-μl aliquots were spread onto Gelrite plates (8 g/liter; Kelco, San Diego, CA) containing the *Sulfolobus* medium and 1 mM $MgCl_2$, 0.3 mM $Ca(NO_3)_2$, 50 mg/liter 5-fluoro-orotic acid (Apollo, Whaley Bridge, United Kingdom), and 20 mg/liter uracil (Sigma, St. Louis, MO). Control samples were diluted to an OD_{600} of 2×10^{-7} to 1×10^{-6} and plated onto nonselective Gelrite plates. After 6 to 8 days at 75°C, the colonies were counted, and from each selective plate, 30 to 60 were streaked onto selective solid medium. After 2 to 3 days, the surviving strains were transferred to 5 ml selective medium and grown for 2 days at 80°C. Strains that failed to grow were discarded. One milliliter of the final culture was stored at −80°C, and 20 μl of the liquid culture was added to 400 μl buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA) and mixed thoroughly on a vortex before being used as a template for PCR amplification.

PCR amplification. Primers were designed with pDRAW32 (ACACLONE Software) and Oligo Analyzer 1.0.2 (Teemu Kuulasma; see the primer list in the supplemental materials; software is available at http://molbiol-tools.ca/molecular_biology_freeware.htm). Fifteen microliters of reaction mixture contained 1 μM of each primer, 2 μl template, 1.5 μl 10× ThermoPol buffer (New England Biolabs), 1 mM deoxynucleoside triphosphates, 1.25 mM $MgCl_2$, and 1 unit *Taq* DNA polymerase (New England Biolabs). Reactions were generally performed in a TRIO-thermoblock (Biometra, Goettingen, Germany) for 5 min at 96°C; 45 s at 96°C, 45 s at 50°C, and 4 min at 72°C for 35 cycles; and 5 min at 72°C, and then maintained at 4°C. For *pyrEF* locus amplification, four sets of PCR primers were used: Pf and Pr, Ef and Er, F1f and F1r, and F2f and F2r (Fig. 1).

Three microliters of each PCR product was checked for homogeneity and size on an agarose gel; 12 μl of the products was purified with a QIAquick PCR Purification kit (QIAGEN, Westburg, Germany), according to the protocol but with an extra washing step with 500 μl PE buffer (QIAGEN), and then eluted

with 30 μl elution buffer after 1 min of incubation. Sometimes, when possible donor mobile-element copies were examined, multiple bands were visible on the agarose gel, and then each band was excised from the gel and extracted with the QIAquick gel extraction kit (QIAGEN) according to the protocol, but using the optional step of an extra wash with buffer QG (QIAGEN), an extra washing step with 700 μl buffer PE incubated for 2 min, and elution in 30 μl elution buffer after a 1-min incubation.

DNA sequencing. Most sequencing reactions consisted of 4 μl DYEnamic ET Dye Terminator reagent premix (Amersham Biosciences, Little Chalfont, United Kingdom), 1 μl 10 pM primer, and 5 μl purified PCR product. The Peltier Thermal Cycler PTC-225 (MJ Research) was set to 2 min at 94°C; 30 s at 93°C, 30 s at 50°C, 1 min at 60°C for 30 cycles; then 30 s at 93°C, 30 min at 50°C, and 2 s at 60°C; and 10 min at 60°C, and then maintained at 4°C. The sequence reaction products were then purified with Sephadex G50 and sequenced with a MegaBACE 1000 (Amersham Biosciences). The remaining sequences were performed on an ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA), where 10 μl of sequencing reaction mixture consisted of 3 to 5 μl purified PCR product, 1.6 pM primer, and a 2-μl BigDye Terminator v. 1.1 cycle-sequencing kit (Applied Biosystems). The reaction was run on a TRIO-thermoblock (Biometra) (30 s at 96°C, 15 s at 50°C, 4 min at 60°C for 25 cycles, and then maintained at 4°C), after which it was ethanol precipitated and redissolved in 12.5 μl Template Suppressing Reagent (Applied Biosystems). The sequences were analyzed with Sequencher (Gene Codes, Ann Arbor, MI); BLAST searches used the Sulfolobus Database (http://dac.molbio.ku.dk/dbs/Sulfolobus/cbin/mutagen.pl), and sequence visualizations and manipulations were carried out with pDRAW32 (ACACLONE Software).

"In vitro library" with NlaIII. Genomic DNA was isolated with a DNeasy Tissue Kit (QIAGEN) using the bacterial protocol. A palindromic 42-bp library template oligonucleotide (300 pmol) in 20 μl H_2O was heated to 98°C for 5 min and then cooled slowly to room temperature. To this was added 20 units NlaIII, 1× NEB4 restriction buffer, 0.01% bovine serum albumin (all from New England Biolabs), and 2^{-16} mol (400 ng) genomic DNA (corresponding to approximately 2.6×10^{-12} mol cut 3′ overhangs) in a 60-μl final volume, which was maintained for 3 h at 37°C. NlaIII was heat inactivated for 20 min at 65°C; 3 μl 20 mM ATP and 1,000 units T4 DNA ligase (New England Biolabs) were added and incubated at 16°C overnight. The ligase was heat inactivated for 25 min at 65°C, and a QIAquick PCR purification kit (QIAGEN) (wash, two times with 700 μl; elution, 50 μl elution buffer after 1 min of incubation) was used to remove excess cut (and uncut) oligonucleotides, as well as ligations to small NlaIII fragments. The library was used as a template in 30 cycles of PCR (50°C annealing) with 1 μM library primer, together with 1 μM homing primer, and the relevant bands were excised from an agarose gel and sequenced using the homing primer. The library template was as follows (the NlaIII site is underlined): CGCCCGTCCG CTCCTGTCCCATGGGACAGGAGCGGACGGGCG; the library primer was as follows: CGCCCGTCCGCTCCTGTCC.

RESULTS

Characterization of mutations. Mutants of *S. solfataricus* P2 that were resistant to FOA were isolated as colonies on Gelrite plates containing FOA and uracil. In order to limit overrepresentation of mutations that may have occurred in liquid culture, before plating the bacteria onto selective medium, three independent experiments were performed. The experiments yielded on average 0.00013 mutants per plated CFU.

The *pyrEF* locus was examined by PCR for 130 FOA-resistant mutants. Four primer sets were designed to cover a region extending from 371 bp upstream of the *pyrE* start codon to 14 bp downstream from the *pyrF* stop codon (genome positions 535561 to 537188) (Fig. 1). If any of the E, F1, or F2 PCR products (Fig. 1) migrated differently from the control PCR product on an agarose gel, they were sequenced. When no size change was detected, all three PCR products were sequenced, and if no mutation was detected, then the P PCR product, covering the sequence upstream of the promoter region, was generated and sequenced (Fig. 1). The identified mutations are presented schematically (Fig. 1) and summarized (Tables 1 and 2).

TABLE 1. Overview of mutations

Mutation	No.	Location	Type
Rearrangement	1 (2)[a]	Promotor, SSO0620, (SSO1367, SSO1466, SSO6050, SSO0549)[a]	Transposase catalyzed (homologous recombination)[a]
MITE	1	*pyrE*	SM2
IS element	93	Promotor, *pyrE*, *pyrF*	ISC*1058*, ISC*1078*, ISC*1217*[b], ISC*1225*[b], ISC*1359*[b], ISC*1439*[b]
Base substitution	9	*pyrE*, *pyrF*	All G-C to A-T
Insertion	2	*pyrF*	1 bp
Deletion	4	*pyrE*	1, 2, and 44 bp
Duplication	1	*pyrE*	62 bp

[a] Parentheses indicate rearrangements that had occurred prior to FOA selection.
[b] IS elements in this study for which antisense RNAs were detected (26, 27).

(Details of each mutant are given in the supplemental materials.)

Frameshifts, deletions, and point mutations. Eight mutant strains carried a frameshift or a stop codon in either *pyrE* or *pyrF*, and of these, two exhibited deletions of 2 bp and 44 bp and one showed a 62-bp sequence duplication. Eight others exhibited an altered amino acid in one of the protein products, which because of the nature of the assay should have resulted in reduced, or no, enzymatic activity (Table 2). Each base pair substitution was G-C to A-T, and a further bias was observed for the five identical Gly117Arg mutations, at least two of which arose independently in two mutant isolation experiments (Fig. 2A and Table 2).

MITE-based mutations. A single SM2 MITE insertion was detected at position 536303 in *pyrE* (Fig. 1). Although each of the four MITE classes (SM1 to -4) exists in multiple copies in the *S. solfataricus* P2 genome, there are minor sequence differences between most copies, so that the origin of the transposed element could be located. The inserted SM2 MITE was shown by PCR to be still present at the original genomic site (position 1980139) in the mutant (data not shown).

IS element-based mutations. About two-thirds of all the mutations were caused by IS element insertions, and their locations and properties are summarized in Fig. 1 and Table 3. Many of them were caused by the IS element ISC*1058*, and the insertions of this element were concentrated at three different sites, 9 in *pyrE* and 20 and 37 insertions at sites separated by only 2 bp within the promoter region (Fig. 1). At least 24 of these insertions did not correspond to preselection mutations because they (i) were isolated in different experiments, (ii) occurred in both orientations, and (iii) derived from different donor IS element copies (see below and Table 3). Similarly, for ISC*1225*, all three observed insertions occurred at one position, but they (i) originated from different donor copies, (ii) were present in different orientations, and (iii) exhibited two different DR sizes (Table 3). Since this target site shows sequence similarity to the inverted repeats of ISC*1225* (Fig. 2B), the conserved sequence probably facilitates transposase recognition. A broader target specificity was observed for the seven ISC*1439* mutants, but the six different target sites (Fig. 1) still exhibited high pairwise sequence similarity (Table 3).

Each potential donor element identified in the genome sequence was examined by PCR, even when several identical potential donor copies were identified (Table 3). The terminal

TABLE 2. Nontranspositional mutations

Position	Type	Effect	Strain(s)
pyrE (535932–536519)			
536116	G to A	Gly66Glu	159
536265	G to A	Asp112Asn	006
536280	G to A	Gly117Arg	125, 132, 163, 166, 176
536281	G to A	Gly117Glu	158
536282	ΔA	Frameshift CAGGAGTAT	157
536392–536435	Δ44 bp	Frameshift and deletion of amino acids TTGGAAAAAC-Δ44bp-GATGAATTGT	162
536402–536463	+62 bp	Frameshift and addition of amino acids	171
pyrF (536506–537174)			
536838	+T	Frameshift ACTTAGTT(+T)GCCGTAA	045
536898–536899	ΔAA	Frameshift GTTATAAGAGAG	161
536956	C to T	Gln151Stop(UAG)	019
536974	+G	Frameshift GACTTCG(+G)AAAAAA	156
536976	ΔA	Frameshift GACTTCGAAAAAAT	139
537044	G to A	Gly180Glu	160

region of one of each type of PCR product was sequenced in order to establish that the PCR products indeed contained the putative donor and that the junctions remained unaltered. All were located in the parent P2 strain in accordance with the published sequence, except for two sets of mutants with inserted ISC*1359* and ISC*1439* elements, and they are considered separately below.

First, the ISC*1359* copy at position 1300529 was the only exact match to elements inserted in two of the mutants (strains P2A-003 and -102). However, this putative donor was not PCR amplifiable from either the mutant strains or the parent P2 strain. Further investigation showed that a large region, constituting about 4% of the published genome and extending from positions 1201319 to 1325406, was absent. In the published genome sequence, the region is bordered by two almost identical copies of ISC*1439*, which were reduced to one copy in the parent strain. Therefore, it was inferred that a deletion had occurred as a result of an intramolecular recombination between the two ISC*1439*s (Fig. 3A). Since the large deletion demonstrated a radical difference between the *S. solfataricus* P2 stock used for genome sequencing and that used for mutant isolation, we refer subsequently to the latter as strain P2A.

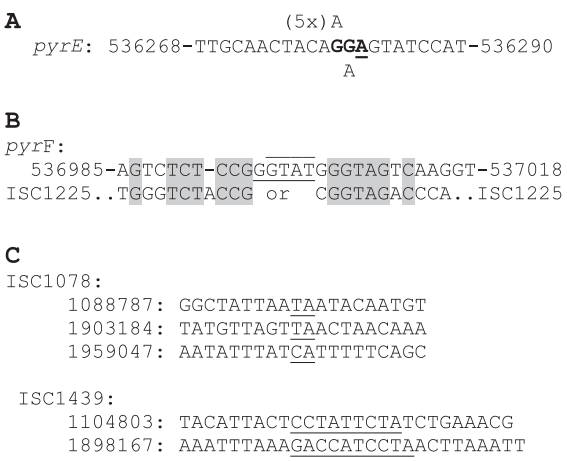

FIG. 2. (A) The sequence bordering the six mutations in codon 117 of *pyrE* (in boldface). Five G-to-A mutations are indicated, which yielded four Gly117Arg and one Gly117Glu change. A single-base-pair deletion is also underlined. (B) Sequence alignment showing the similarity between the inverted terminal repeats of ISC*1225* and its preferred target site in *pyrF*. Gray shading denotes identical bases, and the 4-bp and 5-bp DRs are indicated by lines above and below the sequence, respectively. (C) Empty target sites for ISC*1078* at three positions and for ISC*1439* at two positions. Nucleotides that produce the DR in the published genome sequence are underlined.

In order to identify the actual donor ISC*1359* copy that had transposed into the two mutants (strains P2A-003 and -102), an "in vitro library" strategy (see Materials and Methods and below for details) was employed using a homing primer with a 3′ base unique to ISC*1359* at position 1300529. This enabled us to locate the donor copy at position 1671315, and subsequent PCR amplification and sequencing revealed that it differed at eight positions from the published genome sequence but was instead identical to the transposed elements in the two mutant strains.

Second, the ISC*1439* copies inserted into mutant strains P2A–121, –147, and –165 were identical, but this sequence was absent from the published genome sequence. Therefore, we assumed that either a more recent mutation had occurred in a genomic ISC*1439* copy or there was an error in the genome sequence. In order to locate the donor element, one end (about 500 bp) was PCR amplified and sequenced for 22 of the 33 genomic copies of ISC*1439* that exhibited the sequences most similar to the element inserted in the three mutants. None of the sequences obtained matched perfectly, and moreover, two of the 22 copies and their DRs were absent from the mutant and parent P2A strains (Fig. 2C). It was then discovered that the interrupted ISC*1439* copy at position 454065 and the full-length copy at 480919 had apparently recombined (or had possibly been misassembled in the genome sequence) to produce two different ISC*1439* copies, one of which was identical to the donor copy for all three mutants. This was confirmed by sequencing PCR products generated across the junctions of the rearranged region (Fig. 3B).

The dynamic complexity of the *S. solfataricus* P2 genome was further emphasized by the finding that one of the three iden-

TABLE 3. Properties of mobile-element insertions

Mobile element	Copy no. in published sequence[a]	No. of mutants	Direct repeats (bp)[a]	Aligned target sequences[b]	Possible donating copies
SM2	25	1	6	148: ATAGAGAAAGCAACATTGGAGATTCTTAACGGTG	1980139
ISC*1058*	14	66	9	20 (9) × 535912: T**AAATT**C**TTTTTCA**C**AG**AG**TCTCT**A	A[h]/B[h]/2811580
				37 (9) × 535914: **AA**TTC**TTTTT**CA**CAGAG**T**CTCT**ACG	A/B/2811580
				9 (6) × 535952: **A**T**AATTT**CC**TTTC-GAGT**AAGA**CTT**C ms	A/B/2811580
ISC*1078*	8[c]	2	2	009: TT**T**CAG**ACG**CAC**TAG-TAATCA**T**AGATAGAC**AAG**	35310/1088786[i]/2734629
				110: AA**T**TTT**ACG**ATA**TAGT**T**AATCA-AGCTATA-AAG**A	1711152/1903184[i]/1960125[i]
ISC*1217*	11	4	6 or 7	105: GTC**A**A**TTCT**A**G-AT**TA**T**T**TGG	1892826
				3 (1) × 536422: AGA**A**T**TTCT-GAA**ATT**TT**TAAA ms	838162
				P1: ATT**A**GAA**CTCG-AT**GTAAAAG[d]	—[d]
				SM2: AAG**A**-A**TCT**CCAA**T**G**TT**GCTT ms	1980139
ISC*1225*	11	3	4 or 5	101: **CATAGTCTCTCCGGGTATGGGTAGTCAAGGTGGA**[e]	1362900
				104: **CATAGTCTCTCCGGGTATGGGTAGTCAAGGTGGA**[e]	2476985
				154: **CATAGTCTCTCCGGGTATGGGTAGTCAAGGTGGA**[e]	1362900
ISC*1359*	10[c]	11	4	4 (4) × 535928: C**A**G**AG**TC**TC**T**ACGT***ATG***AATTTCG**[f]	445747/1300529[i]/1671315
				7 (5) × 536659: A**AT**TTCC**TC**C**ACGT**C**TAAA**CCAAT ms	445747
				003 at 539558: A**ATAG**CT**TC**A**ACGT-TAAA**ACTTGA ms	See the text and Fig. 3
				P1 at 536969: T**ATAG**GAGAG**ACT**TCG**AAA**AAATG[d]	—[d]
ISC*1439*	32[c]	7	9	121: TAC**ATA**A**TGAAAAGCATAGT**TGA**AA**GA**TTATC**T**	454065-480919[j]
				129: A**GA***ATG*T**TGAAAAG**TA**GAGT**---**AATATTA**G**CAA**T[g]	2595886
				170: T**GAA**GA**GTGA**TA**AA**CTT**AA**GGATA**ATT**GAAA**AG**AA**G	2595886
				2 (2) × 536523: G**AAA**A--G**TAGAGT**AA**T**ATTA**TT**AG**C**AA**TGG**AT**AAACC	670424/1418678/2903467
				147: T**AAA**TTTT**TAGAGT-TT**GC-**AT**C**AA**GAT**AT**CTTTT ms	454065-480919[j]
				165: C**T**G**CAT**TA**TA**AAT**A**C**TTCT**TCC**AAT**GA**TTT**C-AT ms	454065-480919[j]
				P1: A**TT**C**AT**C-**TA**-GA**ATTTCT**G-A**AAT**--**TTT**AA**AT**A[d] ms	—[d]

[a] This study and reference 3.
[b] The sequences duplicated to form direct repeats are underlined. With single insertions at a site, the mutant strain number is shown, and with multiple insertions, the number of insertions at each position is given, with the minimum number of independent insertions in parentheses. Bases that are identical in more than 66% of the target sites are in italics, except for ISC*1439*, where the italics are for pairwise identity. ms, the minus strand is used for the alignment.
[c] An IS element copy was missing from one or more of the sites (see the text).
[d] Insertions in *pyrEF* of *S. solfataricus* P1, where donor information is unavailable (from reference 16).
[e] All three insertions at the same position, and the 4-bp and 5-bp DRs, are shown with double and single underlining, respectively.
[f] The *pyrE* start codon is marked in boldface.
[g] The *pyrF* start codon is marked in boldface.
[h] Groups of identical ISC*1058* copies: A, 1663758/1703044/1847493/2968643; B, 1042703/1050863.
[i] This IS element copy is not present in the *S. solfataricus* P2A strain (see the text).
[j] The donating copy is probably a recombination product of ISC*1439* at positions 454065 and 480919 (see the text).

tical copies of the potential donor for the inserted ISC*1078* in strain P2A-009 was absent from *S. solfataricus* P2A, as were two of the three possible donors for ISC*1078* in strain P2A-110. Sequencing revealed the absence of 2-bp DRs at the three expected ISC*1078* sites (positions 1088787, 1903184, 1959047), and we infer, therefore, that the P2A strain never carried ISC*1078* copies at these positions (Fig. 2C). Thus, the inserted copies of ISC*1078* must have originated from the copies at position 35310 or 2734629 for the P2A-009 strain and at 1711152 for the P2A-110 strain (Table 3).

In conclusion, the donor for each IS element inserted into the *pyrEF* region could be accounted for, and each one was maintained at its original position in the strain P2A genome.

Genomic rearrangement at the *pyrEF locus*. In a single strain, P2A-003, no PCR products were obtained across the *pyrEF* region with either Pf plus Pr or Ef plus Er primer pairs (Fig. 1). PCR products of parental-strain length were obtained when primer pairs were annealed upstream of position 535728 or downstream of position 536105, but a combination of these primers yielded no product, and therefore, one could exclude a simple deletion of the priming sites (Fig. 1 and 3C). It was inferred that a genomic rearrangement had occurred, and by using the novel "in vitro library" strategy, sequences across the rearrangement junctions were obtained (see Materials and Methods). The method was developed to determine the sequence of an unknown region of a genome, with the only requirement being that it lie adjacent to a known sequence (in this case, the upstream region of *pyrE*). The genome was digested with NlaIII, and a 19-bp synthetic "library template" was then ligated to each cohesive end, thereby tying a defined 19 bp to the restriction site (Fig. 4). PCR with a "library primer" annealed to the library template and a "homing primer" annealed (homed in) to the known sequence then amplified the DNA between the homing primer and the closest

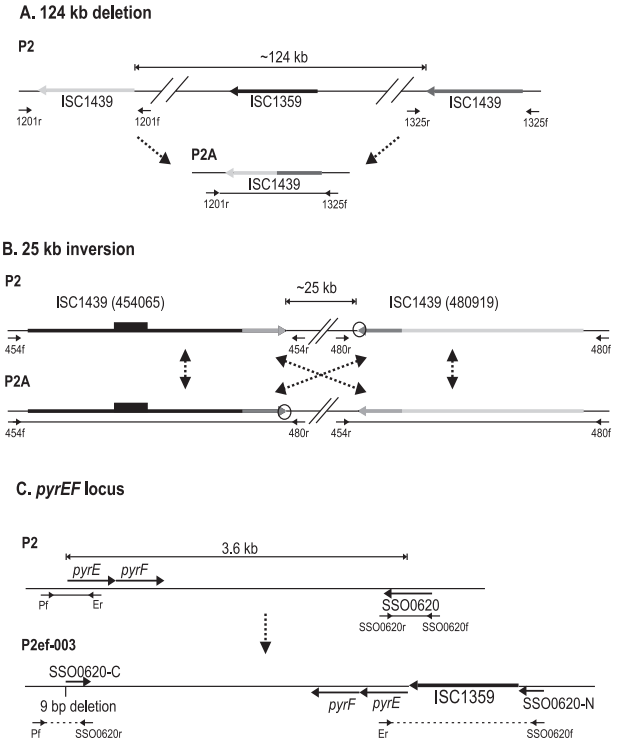

FIG. 3. Genome rearrangements. Lines indicate PCR products obtained using strain P2A DNA as a template, and small arrows with labels indicate annealing sites for PCR primers. (A) The 124-kb region of *S. solfataricus* P2, bordered by ISC*1439* elements, which has been deleted in strain P2A as a result of recombination at the two tandemly arranged elements. The black arrow indicates the deleted copy of ISC*1359* at position 1300529. (B) Putative rearrangement that has occurred between positions 454065 and 480919. The upper line corresponds to the published genome sequence (8), while the lower line corresponds to strain P2A. The two ISC*1439* copies are indicated by horizontal lines (shown in black and light gray). The lower-right copy (480919) is identical to the one inserted in *pyrEF* in strains P2A-121, -147, and -165. This copy exhibits only 2 bp difference (encircled) compared to the published 480919 copy (upper right), whereas the other ISC*1439* (in black) contains numerous mismatches and a 208-bp insertion (black box) compared with this copy. The two 277-bp regions in which recombination (or misassembly) occurred are indicated by intermediate shading. (C) Rearrangement at the *pyrEF* locus in strain P2A-003 with the relevant genes indicated by arrows. The upper line shows the organization of the parent P2A strain (identical to the published sequence), and the lower part shows the changes found in strain P2A-003, where SSO0620 is partitioned and the arrow on the boxed ISC*1359* indicates the direction of the transpose gene. The structure was confirmed by generating PCR products using primer pairs Pf-SSO0620r and Er-SSO0620f (dotted lines) on P2A-003 template DNA.

NlaIII restriction site (Fig. 4). Sequencing of the product using the homing primer then yielded the sequence of the unknown region (Fig. 4).

The sequence obtained using Pf and Ef as homing primers corresponded to that of the published genome sequence until position 535919, and further along, it was identical to the sequence on the opposite strand from position 539561. This indicated that an inversion event had occurred. Subsequent PCR amplification and sequencing confirmed the occurrence of an ~5-kb inversion that exhibited a copy of ISC*1359* at the

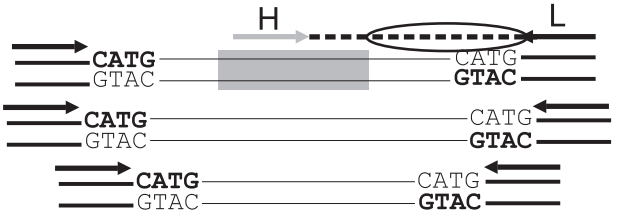

FIG. 4. The "in vitro library" strategy. The restriction-digested genome (thin lines) is ligated to a synthetic library template (thick lines and boldfac restriction sequence). The library primer (black arrows; L) anneals to all of the ligation products, but only one type of fragment contains the known sequence (gray shading) to which the homing primer (gray arrow; H) can anneal. The desired fragment was amplified using primers H and L (dotted line) and yielded the unknown sequence (encircled) when sequenced using the homing primer.

"downstream" end and a 9-bp deletion (positions 535919 to 535927) in the "upstream" region (Fig. 3C).

Undetected mutations. Nineteen mutant strains showed no sequence changes within the *pyrEF* locus for any of the PCR products (Fig. 1 and Table 1). Thus, the phenotype must result from a mutation in another genomic region. Each of these strains was streaked onto selective medium to confirm its mutant phenotype. In contrast to the strain P2A control, each of them grew and exhibited an FOA-resistant phenotype (data not shown). Since all 19 strains were from the same isolation experiment, one cannot readily estimate the mutation frequency. However, if they are independent, then the high number suggests a gene-size target, possibly that of an unidentified orotate transporter.

DISCUSSION

The assay used in this study produces a bias toward transpositional mutations because it identifies only knockout mutants in *pyrE* and/or *pyrF* and it would fail to pick up, for example, most base pair substitutions occurring in the codons of nonessential amino acids. Nevertheless, single-base-pair changes, including 10 substitutions, 2 insertions, and 2 deletions, were observed, as well as a sequence duplication and two larger deletions (Table 2). The deletion mutants, in particular, could prove useful for developing genetic systems for *S. solfataricus* P2 (15). Furthermore, about one-seventh of the mutations occurred elsewhere in the genome. These results contrast with earlier studies on the *lacS* and *pyrEF* loci in *S. solfataricus* P1, which exclusively detected IS element insertions (18, 23).

The identified nontranspositional mutations show some similarities to mutations observed in *S. acidocaldarius*, which probably lacks active mobile elements (9). For example, the 44-bp deletion had no recognizable direct repeats, inverted repeats, or secondary DNA structure at the endpoints, consistent with the deletions observed in *S. acidocaldarius* (12). Moreover, we observed two ±1-bp mutations in runs of single bases (T_3 and A_6 in strains P2A-045 and -139, respectively) (Table 2), although there was no strong preference for −1 bp mutations in long runs of the same base, as occurs in *S. acidocaldarius*, which is richer in homopolynucleotide sequences (11). However, in contrast to the *pyrE* bias observed for mutations in the *pyrEF* region of *S. acidocaldarius* (11), mutations were distributed fairly evenly throughout the region (Table 2).

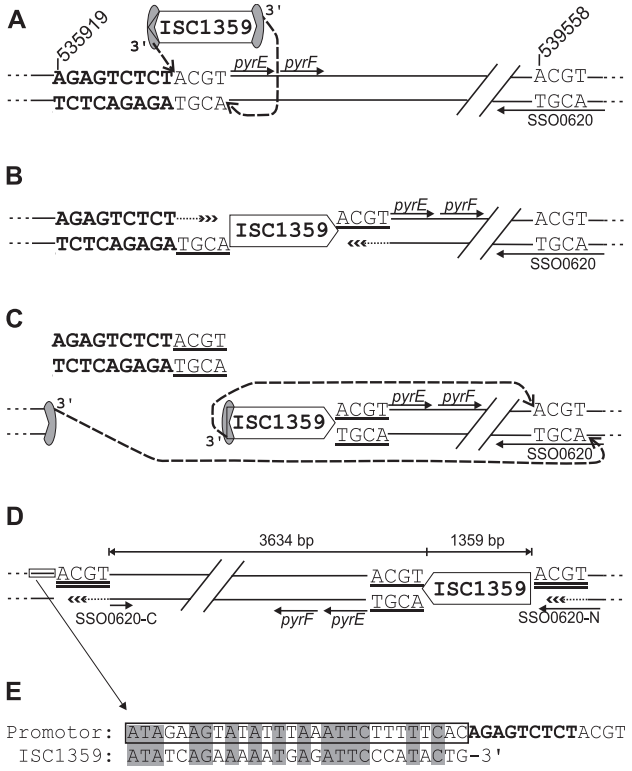

FIG. 5. Proposed model for the genomic rearrangement in strain P2A-003 mediated by the ISC*1359*-encoded transposase. The scheme shows a 3,643-bp section of the *S. solfataricus* P2 genome with the *pyrE*, *pyrF*, and SSO0620 genes represented by arrows and the 9-bp segment (boldface) deleted from strain P2A-003. (A) The multimeric transposase (gray shaded) binds to the inverted terminal repeats of ISC*1359* and catalyzes the nucleophillic attack of the 3′ ends of ISC*1359* on the 5′ ends of the ACGT target site (dashed arrows), located between the 9-bp segment (boldface) and the *pyrE* gene. (B) The tetranucleotides of the target site (underlined) are localized at each end of ISC*1359*, and the resulting single-strand gaps are repaired by host enzymes (small dotted arrows). (C) The transposase binds to one end of ISC*1359* and the sequence immediately adjacent to the 9-bp segment. It then induces double-stranded cuts, which remove the 9-bp segment and one copy of the duplicated 4-bp target site. The transposase then catalyzes nucleophilic attack (dashed arrows) of the free 3′ ends on the ACGT target site in the SSO0620 locus. In effect, this turns the entire genome (minus the excised 9 + 4 bp) into a transposon, which inserts into itself. (D) The previous reaction produced (i) a second target site duplication (double underlining), (ii) partitioning of the SSO0620 gene into SSO0620-N and -C, and (iii) inversion of the 3,634-bp sequence (plus the inserted ISC*1359*) between *pyrE* and the SSO0620 target site. The resulting single-strand gaps are repaired by host enzymes (small dotted arrows), leading to the sequence found in strain P2A-003. (E) Sequence similarity between the proposed transposase binding site in the *pyrEF* locus (boxed) and the 3′ end of ISC*1359*. Identical bases are shaded. The position of the 9-bp segment (boldface) that is lost from strain P2A-003 is also shown.

Transposition of mobile elements. Although there is some bias, especially for ISC*1058*, in the frequency of transposition, the low transpositional activities observed for some classes of IS element may partly reflect the limited availability of sequence-specific sites for transposition in the *pyrEF* region. However, the discovery of putative antisense RNAs complementary to transposase mRNAs in *S. solfataricus* (26, 27) may also contribute to the relatively low number of transpositions observed for ISC*1217*, ISC*1225*, ISC*1359*, and ISC*1439* (Table 1).

Surprisingly, given the large number of MITEs present in the strain P2 chromosome (140 copies), only one MITE transposition was observed. This MITE was predicted to be mobilized by the transposase of ISC*1217*, which was also transposed in this study, and both recognized the shared target site aanH HcgDDwtt (capital letters indicate a higher degree of conservation) (Table 3) (22). Although half of the mutants carried ISC*1058* insertions, none were observed for the multiple-copy SM3 MITEs, which are considered to be mobilized by the ISC*1058*-encoded transposase. Possibly, as has been proposed for SM4, SM3 is an older, more degenerate element in strain P2, at least within its inverted terminal repeat, in contrast to SM1 and SM2 (22). Nevertheless, a *Sulfolobus* MITE which was previously shown to be mobile in *Sulfolobus* strain Y99 9-19 was proposed to be mobilized by the transposase of ISC*1057*, a close relative of ISC*1058* (3). Although no genome data are available for this strain, it reinforces the fact that the SM3-like elements can be active.

Each of the mobile-element insertions observed in the *pyrEF* locus could be correlated with a donor element within the genome of *S. solfataricus* P2 (24). However, during our sequence analyses, we discovered examples of multicopy elements that are present in the published strain P2 genome sequence but were absent from our parent strain P2A stock. These included three copies of ISC*1078* and two copies of ISC*1439* elements and their DRs. The most likely explanation is that during assembly of the *S. solfataricus* P2 genome, occasionally overlapping clones occurred, with or without an IS element, and if most clones contained the element, it was generally assembled into the genome sequence (24; Q. She, personal communication). Thus, strain P2A corresponds to the clones that lacked these IS elements.

Each transposition event observed in this study, except the second transposition in strain P2A-003 (see below), resulted in an increased copy number of the element. The mechanism remains unclear. Bacterial-type replicative transposition can be ruled out, as it involves the donor-carrying replicon inserting between the two daughter IS elements (17), and this was not observed in the present study, where each inserted element was flanked by *pyrEF* sequences. Moreover, an earlier bacterial study demonstrated that persistent genomic proliferation of nonreplicative IS*10* elements did not occur by "cut and paste" from one replicon to another with the subsequent loss of the donor replicon (1). They demonstrated instead that proliferation was most likely to occur by one of two possible mechanisms involving homologous repair: (i) when transposition occurs coincidentally with DNA replication and formation of hemimethylated DNA or (ii) when two or more sister chromosomes coexist postreplication (1, 19).

For *Sulfolobus* species, which exhibit a G_2 phase constituting about 60% of the cell cycle (2), solution ii is the more likely, involving homologous repair of the donor chromosome using the sister chromosome as a template. However, to distinguish definitively between these mechanisms would require more sophisticated genetic tools than are currently available for *Sulfolobus*.

Genomic rearrangements. A model is proposed for the mechanism of genomic rearrangement observed in strain P2A-

003 (Fig. 3C and 5). It involves two transposition steps. In the first, ISC*1359* is inserted at position 535928 just upstream from *pyrE*, as also occurs in strains P2A-102 and -150 (Fig. 5A and B). In the second step, the multimeric ISC*1359*-encoded transposase binds to one terminal repeat of the inserted ISC*1359* and to another similar sequence located just upstream of the 9-bp segment that is absent from strain P2A-003 (Fig. 5C). It then catalyzes excision of the genome as a giant mobile element. This results in the loss of the 9-bp segment, as well as one of the 4-bp DRs, from the first ISC*1359* insertion. The genome-size mobile element then inserts into itself at position 539558, just 5 kb from the excision site, creating a second DR (Fig. 5C). This mechanism produces the sequence found in strain P2A-003 and accounts for the location of ISC*1359* and the inversion event, as well as the absence of the 9-bp segment (Fig. 5D).

The model requires that the ISC*1359* transposase be able bind to the sequence adjacent to the deleted base pairs, which resembles the transposase binding site (Fig. 5E). Support for the occurrence of such an event derives from a study of the bacterial IS*10* transposase, also of the IS*4* family, which was shown to utilize element termini in any orientation, as well as so-called pseudoends, for transposition even when the normal inverted terminal repeat was present (6). Furthermore, the effect of base substitutions within the inverted terminal repeats on transposition activity for IS*10* elements showed that (i) mutations at both terminal repeats depress activity more than mutations at one end and (ii) the most important base pairs are located at positions 6 to 13 of the 22-bp inverted terminal repeat (13). In the model, only one transposon binding site is different from the normal site, and the most conserved region of that site is in the central region (Fig. 5E) which corresponds approximately to base pairs 6 to 13 in the IS*10* elements.

Two other genomic rearrangements were discovered during this study. One was probably caused by homologous recombination between two almost identical copies of ISC*1439* in the same orientation, at positions 1201319 and 1325406, that produced a 124-kb deletion that included the ISC*1359* copy at position 1300529 (Fig. 3A). The deletion was traced back to the plating of a culture of *S. solfataricus* P2 from which two single colonies, A and B, were grown for retention as stock at −80°C (Q. She, personal communication). Colony A had this deletion, whereas colony B did not, indicating that rearrangements occurred under normal laboratory conditions with no known selection. The other 25-kb inversion (Fig. 3B) was present in all *S. solfataricus* P2 strains available in our laboratory, all of which originated from the Deutsche Sammlung von Mikroorganismen und Zellkulturen, in contrast to the strain P2 used for genome sequencing, which was obtained from Wolfram Zillig.

Conclusion. The results show that IS elements can facilitate large genomic changes in *Sulfolobus* both by transpositional mechanisms catalyzed by their transposases and by providing similar sequences for homologous recombination. Moreover, the copy number increase by transposition observed for both IS elements and MITEs continually creates new sites for potential rearrangement and change. Thus, it is extremely important to use mutation-minimizing strategies, such as purifying strains through plating and avoiding long-term cultivation, when working with *S. solfataricus*. Although in a natural environment this genomic variability may be a disadvantage for an individual cell, it does not seem to be a disadvantage for the cellular population, albeit a continually changing one. This is illustrated by the fact that *S. solfataricus* strains, and the closely related *Sulfolobus islandicus*, which is also rich in mobile elements (K. Brügger, personal communication), occur widely in natural environments.

ACKNOWLEDGMENTS

Q. She is thanked for discussions and for his critical comments on the manuscript. K. Brügger is thanked for granting access to the partially sequenced genome of *S. islandicus* HVE10/4 and for help with some bioinformatic analyses, and M. Awayez's help with DNA sequencing is appreciated.

P.R. was funded by a Ph.D. scholarship from the Faculty of Science, Copenhagen University. The research was supported by a grant from the Danish Research Council for Natural Science.

REFERENCES

1. **Bender, J., J. Kuo, and N. Kleckner.** 1991. Genetic evidence against intramolecular rejoining of the donor DNA molecule following IS10 transposition. Genetics **128:**687–694.
2. **Bernander, R., and A. Poplawski.** 1997. Cell cycle characteristics of thermophilic archaea. J. Bacteriol. **179:**4963–4969.
3. **Blount, Z. D., and D. W. Grogan.** 2005. New insertion sequences of *Sulfolobus*: functional properties and implications for genome evolution in hyperthermophilic archaea. Mol. Microbiol. **55:**312–325.
4. **Brügger, K., P. Redder, Q. She, F. Confalonieri, Y. Zivanovic, and R. A. Garrett.** 2002. Mobile elements in archaeal genomes. FEMS Microbiol. Lett. **206:**131–141.
5. **Brügger, K., E. Torarinsson, P. Redder, L. Chen, and R. A. Garrett.** 2004. Shuffling of *Sulfolobus* genomes by autonomous and non-autonomous mobile elements. Biochem. Soc. Trans. **32:**179–183.
6. **Chalmers, R. M., and N. Kleckner.** 1996. IS10/Tn10 transposition efficiently accommodates diverse transposon end configurations. EMBO J. **15:**5112–5122.
7. **Chandler, M., and J. Mahillon.** 2002. Insertion sequences revisited, p. 305–366. *In* N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, D.C.
8. **Charlebois, R. L., R. K. Singh, C. C. Chan-Weiher, G. Allard, C. Chow, F. Confalonieri, B. Curtis, M. Duguet, G. Erauso, D. Faguy, T. Gaasterland, R. A. Garrett, P. Gordon, A. C. Jeffries, C. Kozera, N. Kushwaha, E. Lafleur, N. Medina, X. Peng, S. L. Penny, Q. She, A. St. Jean, J. van der Oost, F. Young, Y. Zivanovic, W. F. Doolittle, M. A. Ragan, and C. W. Sensen.** 2000. Gene content and organization of a 281-kbp contig from the genome of the extremely thermophilic archaeon, *Sulfolobus solfataricus* P2. Genome **43:** 116–136.
9. **Chen, L., K. Brügger, M. Skovgaard, P. Redder, Q. She, E. Torarinsson, B. Greve, M. Awayez, A. Zibat, H.-P. Klenk, and R. A. Garrett.** 2005. The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. J. Bacteriol. **187:**4992–4999.
10. **DasSarma, S.** 1989. Mechanisms of genetic variability in *Halobacterium halobium*: the purple membrane and gas vesicle mutations. Can. J. Microbiol. **35:**65–72.
11. **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.
12. **Grogan, D. W., and J. E. Hansen.** 2003. Molecular characteristics of spontaneous deletions in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*. J. Bacteriol. **185:**1266–1272.
13. **Huisman, O., P. R. Errada, L. Signon, and N. Kleckner.** 1989. Mutational analysis of IS10's outside end. EMBO J. **8:**2101–2109.
14. **Jones, R. N.** 2005. McClintock's controlling elements: the full story. Cytogenet. Genome Res. **109:**90–103.
15. **Jonuscheit, M., E. Martusewitsch, K. M. Stedman, and C. Schleper.** 2003. A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. Mol. Microbiol. **48:**1241–1252.
16. **Kondo, S., A. Yamagishi, and T. Oshima.** 1991. Positive selection for uracil auxotrophs of the sulfur-dependent thermophilic archaebacterium *Sulfolobus acidocaldarius* by use of 5-fluoroorotic acid. J. Bacteriol. **173:**7698–7700.
17. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
18. **Martusewitsch, E., C. W. Sensen, and C. Schleper.** 2000. High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. J. Bacteriol. **182:**2574–2581.

19. **Nagy, Z., and M. Chandler.** 2004. Regulation of transposition in bacteria. Res. Microbiol. **155:**387–398.
20. **Ng, W. V., S. A. Ciufo, T. M. Smith, R. E. Bumgarner, D. Baskin, J. Faust, B. Hall, C. Loretz, J. Seto, J. Slagel, L. Hood, and S. DasSarma.** 1998. Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? Genome Res. **8:**1131–1141.
21. **Oosumi, T., B. Garlick, and W. R. Belknap.** 1996. Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. J. Mol. Evol. **43:**11–18.
22. **Redder, P., Q. She, and R. A. Garrett.** 2001. Non-autonomous elements in the crenarchaeon *Sulfolobus solfataricus*. J. Mol. Biol. **306:**1–6.
23. **Schleper, C., R. Roder, T. Singer, and W. Zillig.** 1994. An insertion element of the extremely thermophilic archaeon *Sulfolobus solfataricus* transposes into the endogenous beta-galactosidase gene. Mol. Gen. Genet. **243:**91–96.
24. **She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, P. Gordon, G. Allard, M. J. Awayez, C.-Y. Chan-Weiher, I. G. Clausen, B. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. K. Gordon, I. Heidekamp de Jong, A. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. Phan Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. M. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. 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.
25. **Simons, R. W., and N. Kleckner.** 1983. Translational control of IS10 transposition. Cell **34:**683–691.
26. **Tang, T.-H., N. Polacek, M. Zywicki, H. Huber, K. Brügger, R. Garrett, J. P. Bachellerie, and A. Hüttenhofer.** 2005. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. Mol. Microbiol. **55:**469–481.
27. **Zago, M. A., P. P. Dennis, and A. D. Omer.** 2005. The expanding world of small RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Mol. Microbiol. **55:**1812–1828.