

Archaeal Homolog of Bacterial Type IV Prepilin Signal Peptidases with Broad Substrate Specificity

Sonja-Verena Albers, Zalán Szabó, and Arnold J. M. Driessen*

Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute,
University of Groningen, 9751 NN Haren, The Netherlands

Received 4 November 2002/Accepted 21 April 2003

A large number of secretory proteins in the thermoacidophile *Sulfolobus solfataricus* are synthesized as a precursor with an unusual leader peptide that resembles bacterial type IV prepilin signal sequences. This set of proteins includes the flagellin subunit but also various solute binding proteins. Here we describe the identification of the *S. solfataricus* homolog of bacterial type IV prepilin peptidases, termed PibD. PibD is an integral membrane protein that is phylogenetically related to the bacterial enzymes. When heterologously expressed in *Escherichia coli*, PibD is capable of processing both the flagellin and glucose-binding protein (GlcS) precursors. Site-directed mutagenesis of the GlcS signal peptide shows that the substrate specificity of PibD is consistent with the variations found in proteins with type IV prepilin-like signal sequences of *S. solfataricus*. We conclude that PibD is responsible for the processing of these secretory proteins in *S. solfataricus*.

Sulfolobus solfataricus is an obligate aerobic thermoacidophilic crenarchaeon that can use a variety of sugars as a sole carbon source for heterotrophic growth (16, 43). Sugar uptake is mediated by high-affinity binding protein-coupled ABC transporters (2). Some of the binding proteins are synthesized as precursors with an unusual amino-terminal signal sequence for membrane targeting that consists of a short positively charged peptide followed by the cleavage site and a hydrophobic, putative membrane-spanning domain. This signal sequence resembles that of bacterial type IV prepilin proteins. In bacteria, type IV pilins are the structural subunits of pili, which are surface-associated structures involved in processes such as surface attachment, twitching motility, or uptake of extracellular DNA (11, 15, 22, 26). The discovery that the precursors of archaeal flagellins contain a signal sequence homologous to bacterial type IV pilin signal sequences has led to the hypothesis that the assembly and structure of archaeal flagella are similar to those of bacterial pili (5, 14). Indeed, the core structures of bacterial pili and the flagellum of the halophilic archaeon *Halobacterium salinarum* are similar (8). Also, the cleavage of flagellin precursors could be shown by purification and amino-terminal sequencing of flagellins from various archaea (40) and by in vitro cleavage of heterologously expressed preflagellin (10). However, in some archaea, type IV pilin signal sequences are not only confined to flagellin subunits but are also found in precursors of other extracellular proteins (3). This was first demonstrated for the glucose-binding protein from *S. solfataricus* (2). In addition, two other sugar-binding proteins, i.e., an arabinose- and a trehalose-binding protein, were found to be processed at their amino termini as expected for type IV prepilin-like signal sequences (2, 12). A recent

analysis of the genome sequence of *S. solfataricus* suggests that there are 10 proteins with a type IV pilin-like signal sequence (1, 32), 6 of which are (putative) solute binding proteins. This raises the question of whether these proteins utilize the same signal peptidase as the flagellin subunits (3) or if they are involved in a separate pathway for protein targeting and maturation, as suggested in another study (19).

One of the best-studied type IV prepilin peptidases is PilD from *Pseudomonas aeruginosa*, a bifunctional enzyme that both cleaves and N-methylates the pilin precursor PilA (36). Whereas methylation is not necessary for pilus function (29), cleavage is an essential activity, as PilD mutants are defective in pilus assembly and accumulate uncleaved PilA precursors in the cytoplasmic membrane (28). Because of the similarity between prepilin and archaeal preflagellin signal sequences, the presence of a gene coding for an enzyme similar to PilD was expected in archaeal genomes. This assumption was recently confirmed by the cloning of FlaK, the preflagellin peptidase from *Methanococcus maripaludis* (4). However, FlaK and its archaeal homologs show only a very low similarity to bacterial type IV prepilin peptidases, possibly because the bacterial enzyme is bifunctional, while there is no evidence for a modified amino acid at the +1 position of mature archaeal flagellins (12, 40).

In this study, we have employed an in vitro assay to monitor type IV prepilin-like signal peptidase activity in isolated *S. solfataricus* membranes. This assay allowed the biochemical demonstration of processing activity toward the precursor forms of flagellin (preFlaB) and glucose-binding protein (preGlcS). We identify a candidate gene in the *S. solfataricus* genome that encodes a membrane protein that has weak homology to PilD. When expressed in *E. coli*, processing of both preFlaB and preGlcS could be demonstrated. This finding and a further analysis of the substrate specificity of this enzyme, termed PibD (for peptidase involved in biogenesis of prepilin-like proteins), suggest that *S. solfataricus* employs a single type

* Corresponding author. Mailing address: Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kercklaan 30, 9751 NN Haren, The Netherlands. Phone: 31-50-3632164. Fax: 31-50-3632154. E-mail: a.j.m.driessen@biol.rug.nl.

TABLE 1. Plasmids used in this study

| Plasmid | Description | Source or reference |
|---------|---|---------------------|
| pET15b | Expression vector with T7 promoter | Novagen |
| pSA4 | Derivative of pET15b containing the multiple cloning site and C-terminal hexa-His tag of pSA5 | This study |
| pET324 | Derivative of pTRC99 | (42) |
| pSA5 | Insertion of a C-terminal hexa His tag in the <i>Bam</i> HI/ <i>Xba</i> I site of pET324 | This study |
| pET2120 | pSA5 carrying <i>glcS</i> as <i>Bsp</i> HI/ <i>Bam</i> HI fragment | This study |
| pET2152 | pET2120 derivative lacking the internal <i>Nco</i> I fragment of <i>glcS</i> | This study |
| pET2171 | pET2152 derivative with G12A mutation | This study |
| pET2172 | L13I | This study |
| pET2173 | G12A/L13I | This study |
| pET2174 | L13F | This study |
| pET2175 | G12L | This study |
| pET2176 | K11A | This study |
| pET2177 | A10K | This study |
| pET2178 | K11D | This study |
| pET2179 | G12R | This study |
| pET2180 | L13R | This study |
| pET2181 | L13D | This study |
| pZA1 | pSA4 carrying <i>flaB</i> Δ 1–4 <i>Nco</i> I/ <i>Bam</i> HI fragment | This study |
| pZA2 | pSA4 carrying <i>flaB</i> as <i>Bsp</i> HI/ <i>Bam</i> HI fragment | This study |
| pZA5 | pSA4 carrying <i>pibD</i> as <i>Nco</i> I/ <i>Bam</i> HI fragment | This study |

IV prepilin-like peptidase to process a broad range of precursor proteins.

MATERIALS AND METHODS

Strains and growth conditions. *S. solfataricus* P2 (DSM 1617, obtained from the Deutsche Sammlung von Mikroorganismen und Zellkultur GmbH, Braunschweig, Germany) was grown aerobically at 80°C by using the medium described by Brock et al. (6) supplemented with 0.1% sucrose and 0.1% yeast extract. The medium was adjusted to pH 3 with sulfuric acid. *Escherichia coli* strain DH5 α (17) was used for all cloning steps. *E. coli* strains BL21(DE3) (Novagen, Madison, Wis.) and C43 (DE3) (27) were used for the overproduction of protein. Both strains carry the pACYC-RIL plasmid that encodes additional tRNAs for rare codons (Stratagene, La Jolla, Calif.).

Isolation of membranes. To prepare membranes of *E. coli*, cells were resuspended in 50 mM Tris-Cl (pH 7.5) and 1 mM EDTA containing a small amount of DNaseI. The suspension was subsequently passed through a French pressure cell at 800 lb/in². Unbroken cells were removed by centrifugation at 3,000 \times g for 15 min at 4°C, and membranes were collected by ultracentrifugation (90,000 \times g, 40 min at 4°C; Beckman TLA 100.4 rotor). Cytoplasmic and outer membranes were separated by sucrose gradient centrifugation (90,000 \times g, 40 min, 4°C) (20), subsequently washed with 20 mM morpholineethanesulfonic acid (MES; pH 6.5), and stored at –80°C. *S. solfataricus* membranes were prepared as described previously (12).

Cloning and plasmid construction. Chromosomal DNA from *S. solfataricus* was isolated by CsCl-buoyant density centrifugation (31). Plasmids and vectors described below are summarized in Table 1. Oligonucleotide primers for the PCR amplification of *flaB* (SSO2323), *glcS*, and *pibD* (SSO0131) were designed based on the genome sequence of *S. solfataricus* P2 (<http://www-archbac.u-psud.fr/projects/sulfolobus/>). The forward primer of *flaB* (5'-CCCCGAATTCATGAAGTCCAAAAAGATG for the version corresponding to the annotated open reading frame [ORF] and 5'-CCCCCATGGTAAAGGAATACAAC for the version lacking the first five amino acids [restriction sites are underlined]) contained a *Bsp*HI or *Nco*I restriction site, respectively. The reverse primer (5'-CCCCGGATCCCTATTACTGATACGCTACCC) contained a recognition sequence for *Bam*HI. Additionally, the native stop codon was deleted with the reverse primer to enable in-frame fusion to a C-terminal epitope tag. The digested PCR products of 924 and 908 bp, respectively, were ligated into the *Nco*I/*Bam*HI sites of the T7 promoter expression vector pSA4, yielding plasmids pZA1 and pZA2 containing C-terminally hexa-histidine-tagged short and long versions of *flaB*, respectively. The expression vector pSA4 was constructed based on the pET15b expression vector (Novagen). The unique *Eco*RI and *Hind*III in pET15b sites were removed by digestion with the corresponding enzymes followed by a fill-in reaction with *Pwo* DNA polymerase (Roche Diagnostics, Mannheim, Germany) and religation. Subsequently, the multiple cloning site

from pET400 (kindly provided by Karel van Wely) was inserted as a 46-bp *Nco*I-*Xho*I fragment, thereby introducing new *Eco*RI and *Hind*III sites downstream from the T7 promoter and removing the N-terminal epitope tag of the vector. To delete unwanted restriction sites, this vector was digested with *Bam*HI and *Xho*I, the cohesive ends of the vector were filled up with *Pwo* DNA polymerase, and the product was religated, yielding pSA3. Finally, a 392-bp *Eco*RI-*Hind*III fragment from pAMP42 (kindly provided by Antonia Picon) was integrated into the corresponding restriction sites of pSA3, thereby introducing a C-terminal hexa-His tag in frame with a *Bam*HI restriction site. This vector was designated pSA4 and is suitable for cloning and expression of ORFs provided as *Nco*I-*Bam*HI or *Bsp*HI-*Bam*HI fragments.

SSO0131 (*pibD*) was amplified with forward (5'-CCCCCATGGTCGTTATATATATATCCAAATTTTCC) and reverse (5'-CCCCGGATCCCAATTG GAAAGCCTATAATCAGCGAC) primers containing *Nco*I and *Bam*HI restriction sites, respectively, and deleting the native stop codon. The PCR product was cloned into the expression vector pSA4, yielding plasmid pZA5.

The gene for *glcS* was amplified with the primers 5'-CCCCCGAATTCATGAAAAGGAAGTACCCGTATAG (forward, containing *Bsp*HI site) and 5'-CCCCGGATCCCTTCAAGAGATAGTATTATTGTC (reverse, containing *Bam*HI). Again, the native stop codon was deleted. The PCR product was cloned into the vector pSA5, which contains a *trc* promoter and adds a C-terminal hexa-His tag in frame with the *Bam*HI restriction site, yielding pET2120. pSA5 was constructed by the ligation of the 392-bp *Bam*HI-*Hind*III fragment from pAMP42 into the expression vector pET324 (42), thereby allowing us to add a C-terminal hexa-His tag in frame with the *Bam*HI restriction site. pET2152 was obtained by removing an 822-bp *Nco*I fragment from pET2120 that is located in the ORF of *glcS* by restriction digest with *Nco*I. After religation of the plasmid, the ORF was restored, resulting in the truncated gene designated *glcS**. The mutations A10K, K11A, K11D, G12A, G12L, L13I, L13F, L13D, and G12A/L13I introduced in *glcS** were constructed in pET2152 by the Quickchange (Stratagene) method of mutagenesis, which utilizes inverse PCR primed by divergent overlapping primers containing the desired complementary nucleotide exchanges.

Expression of recombinant genes in *E. coli*. *E. coli* strain BL21(DE3)-RIL codon plus (Stratagene) carrying plasmid pZA1, pZA2, pET2120, or pET2152 was grown in dYT medium (16 g of trypton per liter, 10 g of yeast extract per liter, and 5 g of NaCl per liter) supplemented with antibiotics at 37°C until an optical density at 600 nm of 0.5 to 0.8 was reached. Expression was induced by the addition of 0.5 mM isopropyl-beta-D-thiogalactopyranoside (IPTG, Roche Diagnostics). Cells were grown for an additional 4 h and harvested by centrifugation (15 min, 5,000 \times g, 4°C). Overexpression of the hexa-His-tagged PibD (plasmid pZA5) occurred as described above, except that *E. coli* strain C43(DE3) carrying plasmid pACYC-RIL (Stratagene) was used. After induction with IPTG, cells were incubated at 30°C for 6 h.

In vitro processing assay. The in vitro preFlaB and preGlcS processing assay was based on the method developed for the *Methanococcus voltae* preflagellin peptidase (10). Isolated *E. coli* membranes (about 5 µg of total protein) either containing preFlaB or preGlcS* and *S. solfataricus* membranes (about 10 µg of total protein) were each preincubated for 10 min in 10 µl of assay buffer (25 mM MES [pH 6.5], 0.5% [vol/vol] Triton X-100, 150 mM KCl, and 1 mM EDTA) at room temperature. Both solubilized samples were mixed and incubated for 30 min (preFlaB) or 1 h (preGlcS) at 55°C, unless stated otherwise. The reaction was stopped by the addition of 5 µl of sodium dodecyl sulfate (SDS) sample buffer and analyzed by SDS-12% polyacrylamide gel electrophoresis (PAGE) and subsequent transferring to polyvinylidene difluoride membranes (Roche Diagnostics). Protein was detected by Western immunoblotting by using monoclonal anti-His antibodies (Dianova, Hamburg, Germany). The activity of cloned PibD was tested as described above except that *E. coli* inner membranes (about 5 µg of total protein) containing overexpressed PibD were used instead of the *S. solfataricus* membranes.

RESULTS

In vitro demonstration of type IV prepilin-like signal peptidase activity. To identify the *S. solfataricus* enzyme responsible for the processing of precursor proteins with a type IV prepilin-like signal sequence, an in vitro assay was established. For this purpose, two proteins with type IV prepilin-like signal sequences were chosen as substrates, flagellin and glucose-binding protein (preGlcS) (2). The flagellin gene was identified in the *S. solfataricus* genome by using the published amino-terminal sequence of the closely related *Sulfolobus shibatae* flagellin (13). We discovered one protein encoded by SSO2323, annotated as putative flagellin, and named it according to previously characterized archaeal flagellins FlaB. The genes coding for preGlcS and preFlaB were cloned by PCR into an expression vector suitable for heterologous expression in *E. coli*, adding a C-terminal hexa-His epitope tag for detection and purification. The annotated sequence of SSO2323 contains two amino-terminal methionine residues at positions 1 (as annotated) and 5, and it is possible that in vivo the second ATG sequence is used as a start codon for translation of *flaB*. Therefore, both the possible genes were cloned. Unless stated otherwise, the shorter version of preFlaB was used in the following experiments. Overexpressed preFlaB migrated in an SDS-polyacrylamide gel at an apparent molecular mass of about 29 kDa (Fig. 1), which deviates from the predicted mass of 34 kDa, probably due to tight folding of the hydrophobic flagellin. PreGlcS was also expressed but it was partially degraded during overexpression in *E. coli* (data not shown). However, removal of an internal *NcoI* fragment of the *glcS* gene resulted in a truncated version of preGlcS, termed preGlcS*, that could be stably expressed.

Since the precursor proteins expressed cofractionated with cytoplasmic membranes (data not shown), inner membrane vesicles isolated from *E. coli* cells expressing either preFlaB or preGlcS* were solubilized with Triton X-100 and mixed with *S. solfataricus* membranes. After incubation at 55°C, cleavage of the amino terminus of preFlaB and preGlcS* was detected by Western immunoblot analysis (Fig. 1B). Processing of preFlaB and preGlcS* resulted in the removal of peptides with lengths of 13 and 12 amino acids, respectively, while the carboxyl-terminal His tag remains intact and can be detected with a specific antibody (Fig. 1A). The precursors remained unprocessed when the incubation was performed at 4°C or when incubated in the absence of *S. solfataricus* membranes. For both precursors, the presence of detergent was required for

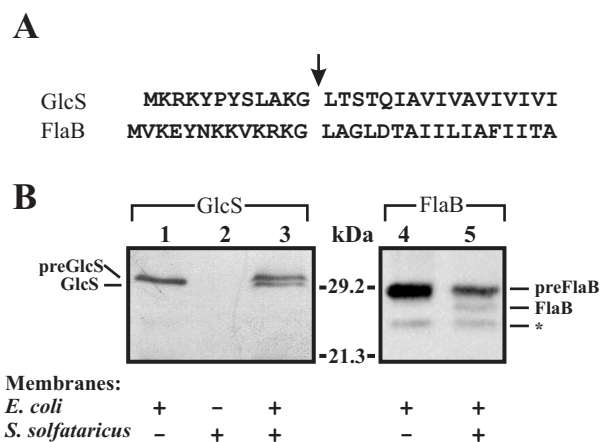


FIG. 1. In vitro cleavage of glucose-binding protein (GlcS*) and flagellin (FlaB) precursors. (A) For the amino-terminal sequences of preGlcS and preFlaB, the cleavage sites are indicated by an arrow. (B) Solubilized *E. coli* membranes containing preGlcS* or preFlaB and *S. solfataricus* membranes were mixed as indicated. No cross-reaction of *S. solfataricus* membrane proteins with the anti-His antibody was observed (lane 2). The asterisk indicates an unspecific degradation product of preFlaB.

cleavage to occur (data not shown). In the case of preFlaB, a second, weaker band appeared that was also present in the absence of *S. solfataricus* membranes. The intensity of this band varied with different preparations of *E. coli* membranes containing preflagellin, but the activity appeared to be due to degradation in *E. coli*. PreFlaB lacking the first four amino acids of the annotated sequence was processed more efficiently than the longer version (data not shown). Therefore, we decided to use the shorter protein in our experiments. These data demonstrate that membranes of *S. solfataricus* harbor a type IV prepilin-like signal peptidase activity.

Identification and cloning of the *S. solfataricus* type IV prepilin-like signal peptidase. To identify the type IV prepilin-like signal peptidase, we first searched the *S. solfataricus* genomic database for a homolog of bacterial enzymes, including those from thermophilic bacteria. However, the BLAST search was unsuccessful. On the other hand, the Cluster of Orthologous Groups (COG) database (<http://www.ncbi.nlm.nih.gov/COG> [38, 39]) lists several archaeal proteins in the group designated COG1989, signal peptidase, cleaves prepilin like proteins. The COG database relies on a phylogenetic classification of proteins. Therefore, proteins with very little sequence identity are clustered in one group due to phylogenetic relationships (39). Subsequently, protein sequences of archaeal COG1989 members were used to perform BLAST searches in the *S. solfataricus* genome database. Single significant hits ($e < 10^{-3}$) were obtained only with MJ1282.1 from *Methanocaldococcus jannaschii* and AF0936 from *Archaeoglobus fulgidus* for the same putative membrane protein, SSO0131. The SSO0131 protein shares 19% identical amino acid sequences with AF0936 and 16.1% with MJ1282.1, respectively. However, higher identities are found with proteins that are not included in COG1989: ST2258 from *Sulfolobus tokodaii*, MA3102 from *Methanosarcina acetivorans*, and FlaK from *M. maripaludis* (Fig. 2A). The primary sequence of SSO0131 has a length of 236 amino acids and a predicted topology of five transmembrane segments

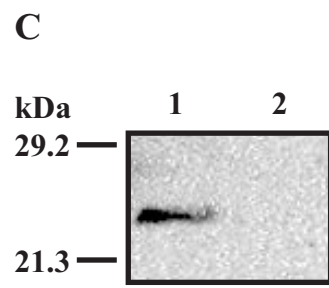
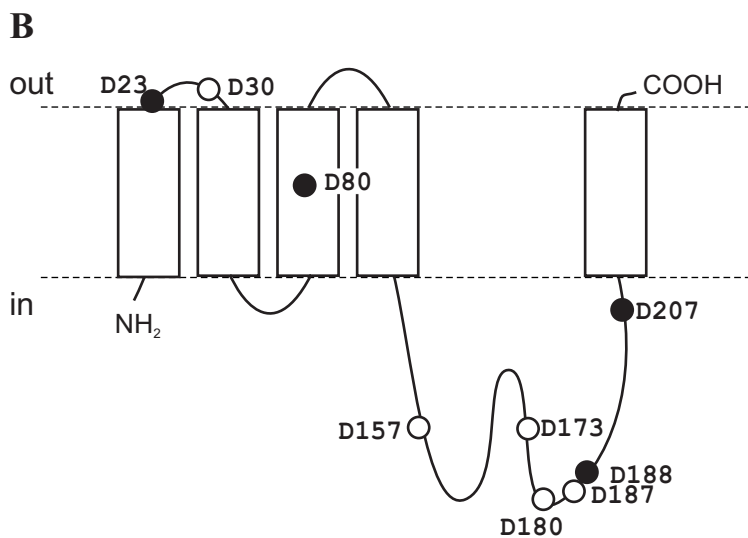
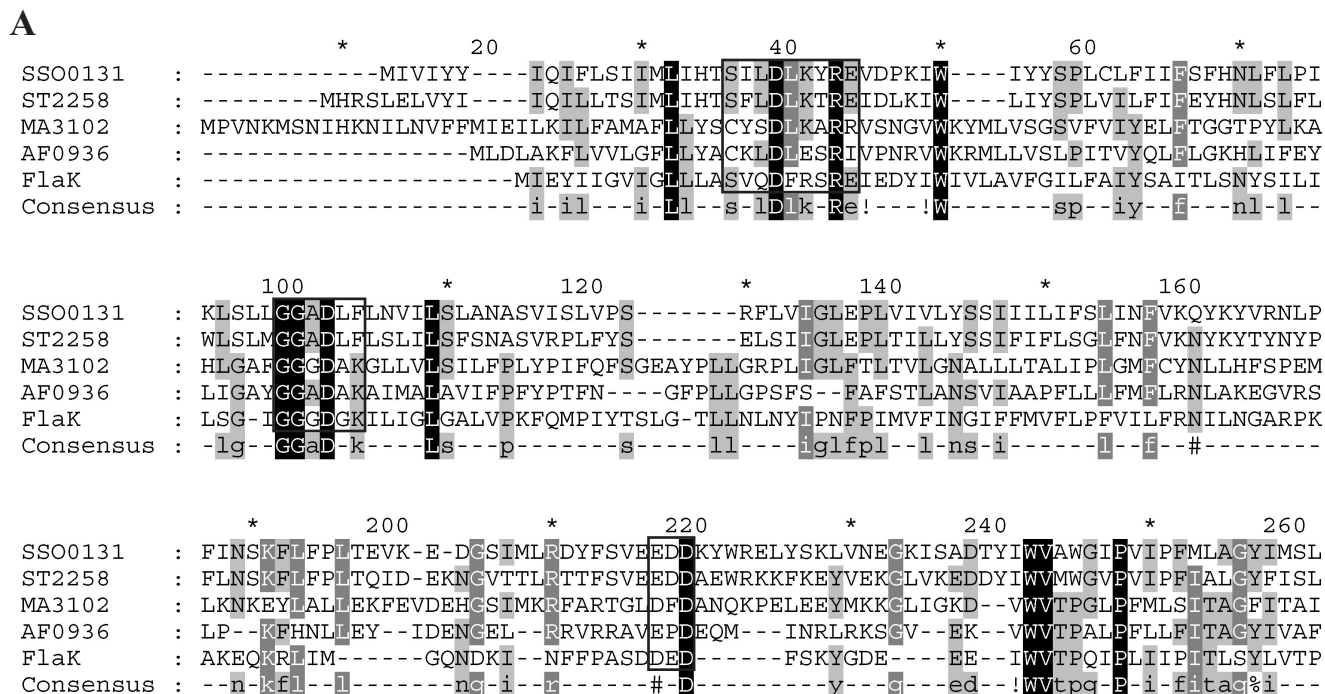


FIG. 2. Identification and heterologous expression of SSO0131 (*pibD*). (A) Multiple alignment carried out by using Multalin (9) of the PibD primary sequence with archaeal homologs from *S. tokodaii* (ST2258), *M. acetivorans* (MA3102), *A. fulgidus* (AF0936), and *M. maripaludis* (FlaK) reveals putatively conserved residues. The sequence motifs surrounding the conserved aspartate residues that might be involved in catalysis are boxed. Residues that are 100%, greater than 80%, or greater than 60% conserved are shaded black, dark gray, and light gray, respectively. (B) Membrane topology of PibD predicted with TMHMM2 (23, 33). Aspartate residues are highlighted and those conserved in at least three sequences are highlighted with filled circles. (C) Heterologous expression of *pibD* in *E. coli*. Inner membranes isolated from cells expressing the gene (lane 1) or harboring the empty vector (lane 2) were analyzed by SDS-PAGE and detected with an antibody directed against the C-terminal hexa-His epitope tag.

(TMS) with a large cytoplasmic loop between TMS 4 and 5 (Fig. 2B). This topology is similar to most putative archaeal type IV prepilin-like signal peptidases but is unlike that of bacterial enzymes such as *Vibrio cholerae* TcpJ with eight TMS and various large cytoplasmic loops (24, 25). The gene coding for the SSO0131 protein is not contained in an apparent operon structure, but remnants of an insertion element are

located upstream of the gene, suggesting a possible rearrangement of the surrounding genome sequence.

To establish whether SSO0131 has cleavage activity *in vitro*, we PCR amplified and cloned the corresponding gene into an expression vector with a carboxyl-terminal hexa-His tag. The gene was overexpressed in *E. coli*, and the recombinant protein could be detected in membrane preparations by SDS-PAGE

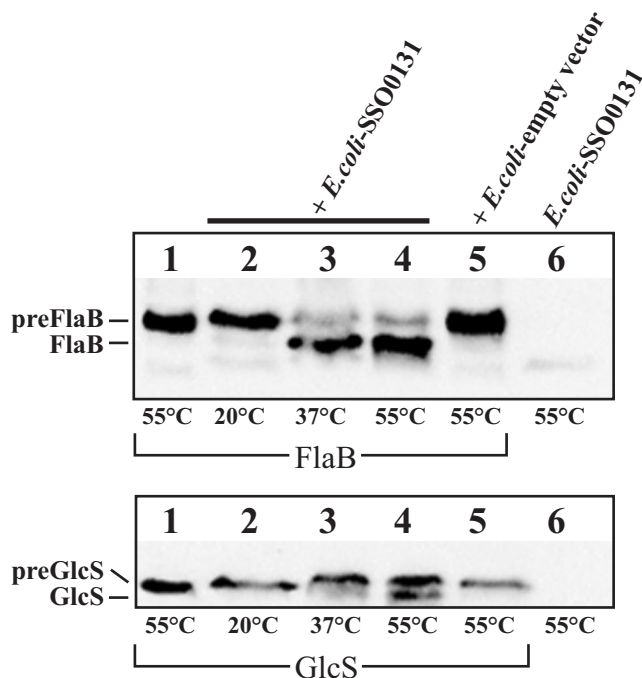


FIG. 3. Temperature dependent *in vitro* cleavage of preFlaB (top panel) and preGlcS* (lower panel) by heterologously expressed PibD. Solubilized inner membranes from *E. coli* containing preFlaB or preGlcS* were incubated with *E. coli* membranes containing overexpressed PibD protein and incubated at the temperatures indicated. No processing was observed at 20°C or when membranes isolated from cells harboring the empty vector were used.

and immunoblotting (Fig. 2). Recombinant SSO0131 was tested for *in vitro* cleavage activity as described above. Both precursor proteins were cleaved in the same way as found with *S. solfataricus* membranes (Fig. 3, lane 4) but remained unprocessed when incubated with membranes isolated from *E. coli* cells harboring the empty vector (Fig. 3, lane 5). The reaction is temperature dependent, as the cleavage reaction was less efficient at 20 and 37°C (Fig. 3, lanes 2 and 3). In order to confirm the *in vitro* cleavage site of FlaB, we made an attempt to purify cleaved FlaB and determine the N-terminal sequence of the protein. However, when cleaved at 55°C, flagellin was insoluble in various buffers, including those containing 2% Triton X-100 or 1% *N*-lauroylsarcosine in combination with 8 M urea or 6 M guanidine hydrochloride, respectively. When the cleavage reaction was performed at 37°C, the protein could be partially solubilized with 1.5% Triton X-100, 8 M urea, and 20 mM β -mercaptoethanol. The solubilized portion of flagellin was partially purified on a Ni-nitrilotriacetic acid Sepharose column, but the eluted protein resisted N-terminal sequencing, although no N-terminal blockage was detected (data not shown).

In conclusion, we demonstrated that the SSO0131 protein is a thermostable type IV prepilin-like signal peptidase of *S. solfataricus*. We therefore name the corresponding gene *pibD* (protein involved in biogenesis of prepilin-like proteins). PibD cleaves both the preFlaB and preGlcS and is therefore potentially involved in the secretion of two proteins with completely unrelated functions.

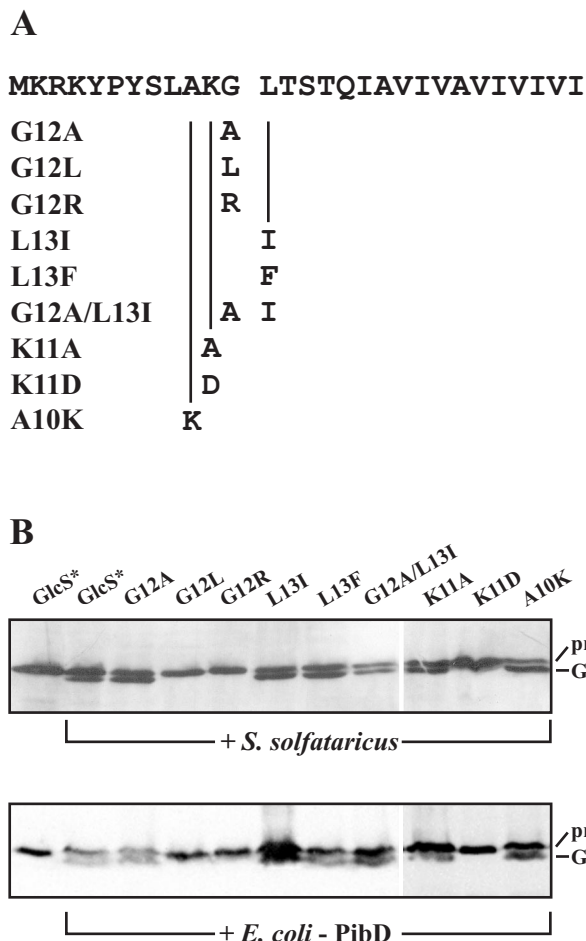


FIG. 4. *In vitro* processing of preGlcS* signal peptide mutants. (A) The signal sequence of GlcS is shown at the top; the cleavage site is indicated by a gap. The mutations introduced are listed with the nomenclature used in panel B. (B) GlcS* precursors with altered signal sequences were incubated with membranes from either *S. solfataricus* membranes (top panel) or *E. coli* overexpressing recombinant *pibD* (lower panel). As controls, preGlcS* was incubated in absence or presence of cleavage activity, respectively (first two lanes in each panel).

Signal sequence specificity of the *S. solfataricus* PibD protein. To determine the substrate specificity of PibD and to confirm the exact site of processing, we constructed signal peptide mutants of preGlcS* by site-directed mutagenesis (Fig. 4). We were particularly interested in the conserved residues around the cleavage site (−1 and +1 residues) and the positively charged −2 residue, as these residues vary among the *S. solfataricus* proteins with predicted type IV prepilin-like signal sequences (Table 2) (1). All mutants could be overexpressed in *E. coli* as unprocessed precursor proteins (data not shown). The *in vitro* cleavage of the preGlcS* signal peptide mutants, either by native *S. solfataricus* membranes (Fig. 4A) or by *E. coli* membranes harboring the recombinant PibD (Fig. 4B), was determined. For all mutants, both types of membranes yielded identical results, suggesting that PibD is the sole enzyme responsible for the type IV prepilin-like signal peptidase activity in *S. solfataricus* membranes. Replacing glycine with

TABLE 2. Alignment of the N-terminal parts of *S. solfataricus* secretory proteins with a putative type IV prepilin-like signal sequence^a

| ORF no. | Signal sequence | Function |
|----------------|---|----------|
| SSO0118 | M KGGY K L K RRKGLSSILGTVIVLAIITLVLGGLLYAYSNGLFSSLTQNAS | Hypo |
| SSO0999 | M SRSD K FSN K E K MRRLSTTT I IGIVVAIV I IVIGAVAAV T LLSHKPSQVVST | TreS |
| SSO2146 | MDMAS R R K NARGLSGAVTALILVIVASV I I A L V VVGF A FG L FG A FTGQGT | Hypo |
| SSO2152 | MLNIY M R K GLSDSVTMMIVLLASVILAI T VV S IL F TYLGYFGSNYG | Hypo |
| SSO2323 | M L K EY N K V K R K GLAGLDT A I I L I A F I I TASV L AYVA I N M GL F V T Q K A K S | FlaB |
| SSO2681 | M Q K Y R K G LE N AL V T V LL L LV A IA A V S L I S Y Y F FG V LR H SM I TT G | Hypo |
| SSO2847 | M K R K Y P Y S L A K L T S T Q I A V I V A V I V I I I GV V AG V L T K G P S T T A V T | GlcS |
| SSO0037 | M K G I S S I L G A I I L I Q I V V S V G L L I L Y L T S L N A K M S N I A Y S | Hypo |
| SSO2684 | M R G I S E A I T V V F L L V L T L I A I A I V T I Y Y L H I V N A N Q Y G L Y | Hypo |
| SSO0489 | M K G F S T L A V V I I I I V V I A V A G I F F V I N S Q G H N T T T T S T | PBP |
| SSO0117 | M M W L K A I S S F S T L I V V M I T L S L I V P L Y L F F T Q T Y T N S S I Q A N | Hypo |
| SSO1171 | M G R K G K I D Y K A I S K T L V A V I V V V I V I A I G G V Y A F L S S Q H S P A P S T | sugar |
| SSO2846 | M E G K Y K R A I S T S T A I I I A V V V I L I V V G V V A Y F Q M G S H A P T S S S | Hypo |
| SSO3066 | M S R R R L Y K A I S R T A I I I I V V V I I A A I A G G L A A Y S S S K P P A T S T S | AraS |
| SSO2712 | M K A L S T L A M A V I V V I A V V A A A Y L I T S S S H P S I S T T T | Sugar |
| SSO3140 | M K K A L S S A I F L I I I T L I L L S V L I P A L L I F N S T P I Y S S Q Q | Hypo |
| Consensus | KGLS RAIT FA | |

^a Positively charged residues in the signal peptide are shown in boldface while hydrophobic amino acids in the core of the signal sequence are underlined. The N-terminal sequence of proteins with boldface numbers has been determined experimentally. Abbreviations used: Hypo, hypothetical protein; PBP, putative phosphate-binding protein.

alanine at the -1 position relative to the cleavage site resulted in a slight improvement in the processing of preGlcS*. However, introduction of an isoleucine or arginine residue at -1 completely abolishes the processing reaction. Leucine at the $+1$ position can be readily exchanged for isoleucine or phenylalanine without interfering with the processing, but no cleavage was observed with arginine or aspartate at this position. With a double mutant in which glycine and leucine at the -1 and $+1$ positions were replaced with an alanine and isoleucine, respectively, significant processing could still be observed. The -2 position is highly conserved among the signal sequences of archaeal flagellins and *S. solfataricus* proteins with a type IV prepilin-like signal sequence and contains a positively charged residue (K or R). When lysine at -2 of preGlcS* was altered to an aspartate, which results in a charge change, no processing could be detected. However, an alanine at the -2 position had no effect on the processing. Finally, when the alanine at the -3 position was changed to lysine, cleavage occurred normally. Hence, a number of mutations close to the cleavage site in preGlcS* can be introduced without interfering with the in vitro cleavage reaction. This shows that PibD is able to cleave type IV prepilin-like signal sequences with all amino acid combinations around the cleavage site that have been found in (putative) prepilin-like precursor proteins encoded by the *S. solfataricus* genome sequence.

DISCUSSION

Here we have identified the *S. solfataricus* type IV prepilin peptidase, PibD, by a combinatorial genomic approach. We used sequences of putative archaeal type IV prepilin peptidases listed in the COG database in order to conduct a BLAST search of the *S. solfataricus* genome. This identified SSO0131 as a possible candidate gene. Importantly, processing activity of native *S. solfataricus* membranes and recombinant SSO0131

(PibD) overexpressed in *E. coli* could be demonstrated by using an assay previously used for the identification and characterization of the peptidases from various bacteria (37) and methanogenic archaea (4, 10). This assay also allowed for the detection of the processing of both the precursors of the flagellin and the glucose-binding protein. Although the optimal growth temperature of *S. solfataricus* is 80°C (43), we could not detect cleavage activity when the assay was performed at temperatures above 60°C (data not shown). This phenomenon is due to the clouding point of the Triton X-100 detergent that was used for the membrane solubilization. At temperatures above 63°C , Triton X-100 loses its solubilizing properties.

Since the recombinant PibD in isolated *E. coli* inner membranes showed processing activity for both precursor proteins tested, our results indicate that a single enzyme of *S. solfataricus* is capable of processing these distinct secretory proteins (3). FlaB ultimately assembles into the flagellum structure, whereas GlcS has to function as an extracellular sugar-binding subunit of an ABC transporter.

Cleavage of preGlcS*, an internally truncated version of full-length preGlcS, appeared less efficient than that of pre-FlaB. This could be due to incorrect folding, which possibly interferes with its interaction with PibD. However, we favor the explanation that the difference in cleavage efficiency is due to intrinsic properties of the signal peptides, such as the number of positively charged residues in the signal sequence, which is six and four for preFlaB and preGlcS, respectively. Indeed, a signal peptide mutant of preGlcS* which contained an additional lysine residue in the signal peptide was more efficiently processed than the wild-type protein (Fig. 4). Jarrell and colleagues (19) suggested that the archaeal type IV prepilin-like signal peptidase homologs are in fact preflagellin peptidases, i.e., enzymes with a dedicated role in the processing of flagellin precursors. According to these authors' hypothesis, other pro-

teins with type IV prepilin-like signal sequences (such as the sugar-binding proteins GlcS, TreS, and AraS from *S. solfataricus*) would require another processing enzyme. It was argued that the highly conserved amino acids in the signal peptide of all archaeal flagellins, i.e., glycine at -1 and lysine or arginine at positions -2 and -3 , are essential for recognition by the preflagellin peptidase and thus deviation from this consensus sequence would require a second peptidase with altered substrate specificity. Processing of the flagellin precursor by the *M. voltae* peptidase indeed appears very sensitive toward alterations of the signal sequence. Most mutations result in partial or complete loss of cleavage activity (41). However, *S. solfataricus* PibD seems to be equipped with a much broader specificity. It processes preGlcS*, even though the signal sequence lacks a lysine or arginine at position -3 , whereas an *M. voltae* flagellin mutant with alanine or glutamate at the same position was cleaved only partially (41). Moreover, a lysine-to-alanine mutation of the GlcS* signal sequence at position -2 did not affect the cleavage reaction, and glycine at position -1 could be changed to alanine (Fig. 4) without loss of cleavage activity. Recently, we identified a number of putative precursor proteins with type IV prepilin-like signal sequences (1). Rescanning the genome with the current consensus sequence revealed a number of new candidate proteins (Table 2). These signal sequences have a relatively conserved core of four amino acids around the cleavage site with the consensus sequence [K/R][G/A][L/I/F][S/T/A] (-1 and $+1$ positions relative to the processing site shown in boldface). We have previously shown that the sugar-binding proteins GlcS and TreS are processed between a glycine and a leucine residue, while AraS is processed between alanine and isoleucine (2). Taken together, these data suggest that PibD is capable of processing a wide variety of proteins with a type IV prepilin-like signal sequence, including binding proteins (12) and preflagellin. Intriguingly, the genome sequence of another species from the genus *Sulfolobus*, *S. tokodaii* (21), does not code for homologs of the *S. solfataricus* sugar-binding proteins that contain a type IV prepilin-like signal sequence (unpublished data). Hence, *S. solfataricus* seems to have undergone a specialization concerning the variety of solute binding proteins and also the mechanism of secretion of these proteins.

In the gram-negative bacterium *P. aeruginosa*, a number of possible alterations at the $+1$ position (phenylalanine) of the prepilin PilA are tolerated in vivo, whereas replacement of the -1 glycine for any other amino acid except alanine resulted in complete loss of a functional pilus structure (35). With alanine at position -1 , PilA was partially processed in vivo. In contrast, the pseudopilin PulG from *Klebsiella oxytoca* remained unprocessed in vivo when the -1 residue (glycine) was changed to alanine, valine, or glutamate (30). Replacement of the glycine at position -1 of the *M. voltae* preflagellin signal sequence to alanine resulted in reduced processing activity, and other amino acids were not tolerated at all at this position, resulting in a preflagellin mutant that cannot be cleaved (41). Also, mutants with alterations at position -2 (lysine to alanine) or -3 (lysine to alanine or glutamate) were not cleaved or partially cleaved, respectively (41). Strikingly, no loss of cleavage activity was observed when these mutations were introduced into the signal sequence of preGlcS*. Hence, the *S. solfataricus*

PibD seems to be functionally more similar to PilD and seems to be equipped with a relatively broad substrate specificity.

In *M. jannaschii*, the putative preflagellin peptidase gene (designated *flaK* [4]) is downstream of the flagellin gene cluster and may be cotranscribed with other *fla* genes (4). The genome of *M. jannaschii* (7) codes for a paralog of FlaK, MJ1282.1, which is located elsewhere in the genome. The gene seems to be part of an operon that also codes for two homologs of bacterial type II secretion proteins. Although MJ1282.1 might represent a pseudogene without a function to the cell, it is also possible that it is involved in the processing of other proteins synthesized with a type IV prepilin-like signal sequence. So far the only candidate is the S-layer protein of *M. jannaschii* (3), but experimental evidence is required to confirm this hypothesis.

Although the primary sequence conservation between archaeal and bacterial type IV prepilin-like signal peptidases is very low, an evolutionary relationship between these proteins is evident and therefore the catalytic mechanism of signal peptide cleavage reaction may be identical for these enzymes. In the past, cysteines in PilD have been implicated in catalysis (34), although XpsO from *Xanthomonas campestris* can functionally replace PilD despite the fact that it does not contain cysteine residues (18). Also, a recent study on TcpJ of *V. cholerae* suggests that in fact two aspartate residues are essential for processing (24). Unlike the cysteine, the two aspartate residues are highly conserved in all described bacterial type IV prepilin-like signal peptidases. An alignment of the archaeal enzymes reveals the presence of three conserved aspartate residues (Fig. 2). Future studies should be directed at the identification of the catalytic residues by site-directed mutagenesis.

ACKNOWLEDGMENTS

This work was supported by a VENI-grant (project code 863.02.001) of The Netherlands Organization for Scientific Research (NWO) to S.-V.A.

We thank Michael Galperin for suggesting the use of the COG database for identification of the *S. solfataricus* type IV prepilin-like signal peptidase. S.-V. Albers and Z. Szabó contributed equally to this work.

REFERENCES

- Albers, S. V., and A. J. M. Driessen. 2002. Signal peptides of secreted proteins of the archaeon *Sulfolobus solfataricus*: a genomic survey. *Arch. Microbiol.* **177**:209–216.
- Albers, S. V., M. G. Elferink, R. L. Charlebois, C. W. Sensen, A. J. M. Driessen, and W. N. Konings. 1999. Glucose transport in the extremely thermoacidophilic *Sulfolobus solfataricus* involves a high-affinity membrane-integrated binding protein. *J. Bacteriol.* **181**:4285–4291.
- Albers, S. V., W. N. Konings, and A. J. M. Driessen. 1999. A unique short signal sequence in membrane-anchored proteins of Archaea. *Mol. Microbiol.* **31**:1595–1596.
- Bardy, S. L., and K. F. Jarrell. 2002. FlaK of the archaeon *Methanococcus maripaludis* possesses preflagellin peptidase activity. *FEMS Microbiol. Lett.* **208**:53–59.
- Bayley, D. P., and K. F. Jarrell. 1998. Further evidence to suggest that archaeal flagella are related to bacterial type IV pili. *J. Mol. Evol.* **46**:370–373.
- 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.
- Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. Geoghegan, and J. C. Venter. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**:1058–1073.

8. Cohen-Krausz, S., and S. Trachtenberg. 2002. The structure of the archaebacterial flagellar filament of the extreme halophile *Halobacterium salinarum* R1M1 and its relation to eubacterial flagellar filaments and type IV pili. *J. Mol. Biol.* **321**:383–395.
9. Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**:10881–10890.
10. Correia, J. D., and K. F. Jarrell. 2000. Posttranslational processing of *Methanococcus voltae* preflagellin by preflagellin peptidases of *M. voltae* and other methanogens. *J. Bacteriol.* **182**:855–858.
11. Dubnau, D. 1997. Binding and transport of transforming DNA by *Bacillus subtilis*: the role of type-IV pilin-like proteins—a review. *Gene* **192**:191–198.
12. Elferink, M. G., S. V. Albers, W. N. Konings, and A. J. Driessen. 2001. Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. *Mol. Microbiol.* **39**:1494–1503.
13. Faguy, D. M., D. P. Bayley, A. S. Kostyukova, N. A. Thomas, and K. F. Jarrell. 1996. Isolation and characterization of flagella and flagellin proteins from the thermoacidophilic archaea *Thermoplasma volcanium* and *Sulfolobus shibatae*. *J. Bacteriol.* **178**:902–905.
14. Faguy, D. M., K. F. Jarrell, J. Kuzio, and M. L. Kalmokoff. 1994. Molecular analysis of archaeal flagellins: similarity to the type IV pilin-transport superfamily widespread in bacteria. *Can. J. Microbiol.* **40**:67–71.
15. Fernandez, L. A., and J. Berenguer. 2000. Secretion and assembly of regular surface structures in Gram-negative bacteria. *FEMS Microbiol. Rev.* **24**:21–44.
16. Grogan, D. W. 1989. Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. *J. Bacteriol.* **171**:6710–6719.
17. Hanahan, D. 1985. Techniques for transformation in *E. coli*, p. 109–135. In D. Rickwood and B. D. Hames (ed.), *DNA cloning, a practical approach*. IRL Press, Oxford, England.
18. Hu, N. T., P. F. Lee, and C. Chen. 1995. The type IV pre-pilin leader peptidase of *Xanthomonas campestris* pv. *campestris* is functional without conserved cysteine residues. *Mol. Microbiol.* **18**:769–777.
19. Jarrell, K. F., J. D. Correia, and N. A. Thomas. 1999. Is the processing and translocation system used by flagellins also used by membrane-anchored secretory proteins in archaea? *Mol. Microbiol.* **34**:395–398.
20. Kaufmann, A., E. H. Manting, A. K. Veenendaal, A. J. Driessen, and C. van der Does. 1999. Cysteine-directed cross-linking demonstrates that helix 3 of SecE is close to helix 2 of SecY and helix 3 of a neighboring SecE. *Biochemistry* **38**:9115–9125.
21. Kawarabayasi, Y., Y. Hino, H. Horikawa, K. Jin-no, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Kato, T. Yoshizawa, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, S. Masuda, M. Yanagii, M. Nishimura, A. Yamagishi, T. Oshima, and H. Kikuchi. 2001. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res.* **8**:123–140.
22. Koebnik, R. 2001. The role of bacterial pili in protein and DNA translocation. *Trends Microbiol.* **9**:586–590.
23. Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**:567–580.
24. LaPointe, C. F., and R. K. Taylor. 2000. The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases. *J. Biol. Chem.* **275**:1502–1510.
25. Lory, S., and M. S. Strom. 1997. Structure-function relationship of type-IV prepilin peptidase of *Pseudomonas aeruginosa*—a review. *Gene* **192**:117–121.
26. Mattick, J. S. 2002. Type IV pili and twitching motility. *Annu. Rev. Microbiol.* **56**:289–314.
27. Miroux, B., and J. E. Walker. 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**:289–298.
28. Nunn, D., S. Bergman, and S. Lory. 1990. Products of three accessory genes, *pilB*, *pilC*, and *pilD*, are required for biogenesis of *Pseudomonas aeruginosa* pili. *J. Bacteriol.* **172**:2911–2919.
29. Pepe, J. C., and S. Lory. 1998. Amino acid substitutions in PilD, a bifunctional enzyme of *Pseudomonas aeruginosa*. Effect on leader peptidase and N-methyltransferase activities *in vitro* and *in vivo*. *J. Biol. Chem.* **273**:19120–19129.
30. Pugsley, A. P. 1993. Processing and methylation of PuIG, a pilin-like component of the general secretory pathway of *Klebsiella oxytoca*. *Mol. Microbiol.* **9**:295–308.
31. Schleper, C., I. Holz, D. Janekovic, J. Murphy, and W. Zillig. 1995. A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. *J. Bacteriol.* **177**:4417–4426.
32. She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, 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.
33. Sonnhammer, E. L., G. von Heijne, and A. Rogh. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences, p. 175–182. In J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, and C. Sensen (ed.), *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, Menlo Park, Calif.
34. Strom, M. S., P. Bergman, and S. Lory. 1993. Identification of active-site cysteines in the conserved domain of PilD, the bifunctional type IV pilin leader peptidase/N-methyltransferase of *Pseudomonas aeruginosa*. *J. Biol. Chem.* **268**:15788–15794.
35. Strom, M. S., and S. Lory. 1991. Amino acid substitutions in pilin of *Pseudomonas aeruginosa*. Effect on leader peptide cleavage, amino-terminal methylation, and pilus assembly. *J. Biol. Chem.* **266**:1656–1664.
36. Strom, M. S., D. N. Nunn, and S. Lory. 1993. A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family. *Proc. Natl. Acad. Sci. USA* **90**:2404–2408.
37. Strom, M. S., D. N. Nunn, and S. Lory. 1994. Posttranslational processing of type IV prepilin and homologs by PilD of *Pseudomonas aeruginosa*. *Methods Enzymol.* **235**:527–540.
38. Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**:33–36.
39. Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. *Science* **278**:631–637.
40. Thomas, N. A., S. L. Bardy, and K. F. Jarrell. 2001. The archaeal flagellum: a different kind of prokaryotic motility structure. *FEMS Microbiol. Rev.* **25**:147–174.
41. Thomas, N. A., E. D. Chao, and K. F. Jarrell. 2001. Identification of amino acids in the leader peptide of *Methanococcus voltae* preflagellin that are important in posttranslational processing. *Arch. Microbiol.* **175**:263–269.
42. van der Does, C., T. den Blaauwen, J. G. de Wit, E. H. Manting, N. A. Groot, P. Fekkes, and A. J. Driessen. 1996. SecA is an intrinsic subunit of the *Escherichia coli* preprotein translocase and exposes its carboxyl terminus to the periplasm. *Mol. Microbiol.* **22**:619–629.
43. 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.