University of Groningen

Ss-LrpB, a transcriptional regulator from *Sulfolobus solfataricus*, regulates a gene cluster with a pyruvate ferredoxin oxidoreductase-encoding operon and permease genes

Peeters, Eveline; Albers, Sonja-Verena; Vassart, Amelia; Driessen, Arnold J. M.; Charlier, Daniel

Published in:
Molecular Microbiology

DOI:
[10.1111/j.1365-2958.2008.06578.x](https://doi.org/10.1111/j.1365-2958.2008.06578.x)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2009

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Peeters, E., Albers, S.-V., Vassart, A., Driessen, A. J. M., & Charlier, D. (2009). Ss-LrpB, a transcriptional regulator from *Sulfolobus solfataricus*, regulates a gene cluster with a pyruvate ferredoxin oxidoreductase-encoding operon and permease genes. *Molecular Microbiology*, 71(4), 972-988.
<https://doi.org/10.1111/j.1365-2958.2008.06578.x>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Ss-LrpB, a transcriptional regulator from *Sulfolobus solfataricus*, regulates a gene cluster with a pyruvate ferredoxin oxidoreductase-encoding operon and permease genes

Eveline Peeters,¹ Sonja-Verena Albers,^{2†}
Amelia Vassart,¹ Arnold J. M. Driessen² and
Daniel Charlier^{1*}

¹Erfelijkheidsleer en Microbiologie, Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussel, Belgium.

²Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, the Netherlands.

Summary

Ss-LrpB is an Lrp-like transcriptional regulator from *Sulfolobus solfataricus*. Previously, *in vitro* binding of Ss-LrpB to the control region of its own gene has been extensively studied. However, nothing was known about the physiological role of this regulator yet. Here, using the knowledge of the DNA-binding sequence specificity of Ss-LrpB, several potential binding sites were predicted *in silico* in promoter regions of genes located adjacently to the *Ss-lrpB* gene. These genes include an operon encoding a pyruvate ferredoxin oxidoreductase (*porDAB*) and two genes encoding putative permeases. *In vitro* protein–DNA interaction studies allowed the identification of the Ss-LrpB binding sites in the cognate control regions. Intriguingly, the binding site organization in the *por* operator is identical to that in the *Ss-lrpB* control region. An *Ss-lrpB* gene disruption mutant was constructed and the gene expression of the above-mentioned targets in this mutant was analysed by qRT-PCR and compared with isogenic wild type. Our data demonstrate that *in vivo* Ss-LrpB acts as an activator at the promoters of the three predicted targets. Based on these results, it appears that not all regulators belonging to the archaeal Lrp

family perform a function related to the amino acid metabolism, unlike the bacterial Lrp-like regulators.

Introduction

It is well established that the archaeal basal transcription apparatus is a simplified version of the eukaryotic RNA polymerase (RNAP) II machinery (Bell and Jackson, 2001; Geiduschek and Ouhammouch, 2005). The core apparatus consists of RNAP, TATA-binding protein (TBP) and transcription factor B (TFB) which are homologues of their eukaryotic counterparts. An additional eukarya-like factor, transcription factor E (TFE), exerts a stimulatory effect on TBP binding and RNAP recruitment and stabilizes the transcription bubble during the early elongation phase (Grünberg *et al.*, 2007).

In contrast to the basal machinery, the regulation of transcription in archaea appears to be related to the bacterial system. In almost all sequenced archaeal genomes, several genes can be identified that are predicted to encode orthologues of bacterial transcription regulators (Aravind and Koonin, 1999; Kypides and Ouzounis, 1999). Most of these candidate regulators contain a helix–turn–helix (HTH) DNA-binding motif (Aravind and Koonin, 1999). Regulation performed by DNA-binding proteins occurs primarily at the level of transcription initiation and the effect can be either negative or positive (Ouhammouch, 2004; Geiduschek and Ouhammouch, 2005). Depending on the exact position of the regulator-binding site(s), negative control can be exerted by binding the promoter DNA thereby inhibiting either TBP/TFB binding or RNAP recruitment. Otherwise, positive control can be achieved by conveying stimulatory effects on the binding of TBP (Ouhammouch *et al.*, 2003). In the case of activation, binding occurs upstream of the TATA box promoter element.

The best-studied transcription regulator family in archaea is the Leucine-responsive Regulatory Protein (Lrp) family, also called Feast/Famine Regulatory Proteins (Brinkman *et al.*, 2003; Yokoyama *et al.*, 2006). The Lrp family is widely conserved among many bacteria and

Accepted 11 December, 2008. *For correspondence. E-mail dcharlie@vub.ac.be; Tel. (+32) 2 629 13 42; Fax (+32) 2 629 13 45.
†Present address: Max-Planck-Institute for terrestrial Microbiology, Karl-von-Frisch-Strasse, 35043 Marburg, Germany.

almost all archaea. The number of *lrp*-like genes in a genome can range from one up to more than 10 (Brinkman *et al.*, 2003). Lrp-like proteins have an N-terminal DNA-binding domain that contains a HTH motif. This domain is connected through a flexible hinge with the C-terminal oligomerization and cofactor response domain, also called Regulation of Amino acid Metabolism (RAM) domain (Ettema *et al.*, 2002). The latter exhibits an $\alpha\beta$ sandwich motif. For several archaeal and bacterial Lrp-like regulators, the crystal structure has been solved (for an overview, see Kawashima *et al.*, 2008). Recently, the first cocrystal structure of an Lrp-like dimer bound to a short DNA duplex was obtained for FL11 of *Pyrococcus* OT3 (Yokoyama *et al.*, 2007). The basic structural unit of Lrp-like proteins is always a dimer; usually they crystallize into octamers or higher association states demonstrating their ability to oligomerize. Many Lrp-like regulators interact with amino acids as cofactors (Brinkman *et al.*, 2003). Based on the several crystal structures, a structural code was developed to predict potentially interacting amino acids for a regulator (Okamura *et al.*, 2007). Cofactor binding plays an important role in the regulatory mechanism of Lrp-like regulators. It results in minor structural changes (allosteric control) or may lead to a change in the oligomeric state of the protein thereby affecting the DNA-binding affinity (Brinkman *et al.*, 2003; Okamura *et al.*, 2007).

As opposed to their structure, the regulation mechanisms and regulatory target genes of Lrp regulators are not as conserved. Lrp-like regulators can act as an activator or a repressor, or both. Their regulon can be either limited (one target gene, e.g. AsnC from *Escherichia coli*; Kölling and Lother, 1985) or very extended (e.g. the global regulator *E. coli* Lrp which regulates about 10% of all genes; Hung *et al.*, 2002; Tani *et al.*, 2002). Target genes of Lrp-like regulators often encode proteins performing functions in the amino acid or central metabolism of the cell.

In archaea, the physiological role of most identified Lrp-like regulators remains elusive. Archaeal Lrps with a known function include LysM from *Sulfolobus solfataricus* which positively regulates an operon-encoding enzymes catalysing lysine biosynthesis (Brinkman *et al.*, 2002) and Ptr2 from *Methanocaldococcus jannaschii* and its orthologues FL10/LrpA from *Pyrococcus* species activating the expression of rubredoxin and ferredoxin genes (Ouhammouch *et al.*, 2003; Ouhammouch and Geiduschek, 2005; Kawashima *et al.*, 2008). The only global archaeal Lrp-like regulator that has been studied is FL11 from *Pyrococcus* OT3 (Yokoyama *et al.*, 2007). Based on *in silico* analysis, it is predicted that this regulator represses the expression of about 20% of all genes directly or indirectly (genes involved in lysine biosynthesis, ATP synthesis, transmembrane transport and genes

encoding other regulators thereby leading to regulatory cascades).

The genome of *S. solfataricus*, a hyperthermoacidophilic crenarchaeon, is predicted to encode seven full-length Lrp-like proteins (Yokoyama *et al.*, 2006). To date, five of these regulators have been characterized, including Ss-LrpB (Napoli *et al.*, 1999; 2001; Brinkman *et al.*, 2002; Enoru-Eta *et al.*, 2002; Peeters *et al.*, 2004). Ss-LrpB binds the control region of the *Ss-lrpB* gene itself, which suggests that the protein regulates its own expression (Peeters *et al.*, 2004). Binding occurs at three specific binding sites (boxes) on the same face of the DNA helix with the two outer boxes being high-affinity targets. These two boxes need to be occupied by Ss-LrpB dimers before the middle low-affinity box is bound as well. Furthermore, Ss-LrpB binding induces strong DNA deformations and, upon occupation of all three binding sites, even DNA wrapping (Peeters *et al.*, 2006).

Besides the proposed autoregulatory role of Ss-LrpB, nothing was known yet about its physiological role. Here, we have determined its function by a combination of *in silico* binding site analysis, *in vitro* DNA-binding studies and *in vivo* gene expression analysis. Since most specific Lrp-like regulators target juxtaposed genes, we employed the knowledge of the genomic context of *Ss-lrpB* and its binding sequence specificity (Peeters *et al.*, 2007) to search for its potential target genes. *Ss-lrpB* is located adjacently to an operon predicted to encode subunits of a pyruvate ferredoxin oxidoreductase (POR) and preceded by two genes encoding potential permeases with an as yet unknown function (Fig. 1A). Ss-LrpB binding sites were identified in the promoter regions of all three targets (*por* operon and two permease genes) and protein-DNA binding was extensively studied. Furthermore, an *Ss-lrpB* gene disruption mutant was constructed and gene expression of the above-mentioned targets was analysed in isogenic *Ss-lrpB*⁺ and *Ss-lrpB::lacS* strains. Our data demonstrate that Ss-LrpB acts as a positive regulator of the POR-encoding operon and of the two permease genes.

Results

The genomic context of Ss-LrpB and binding to promoter regions of neighbouring genes

The *Ss-lrpB* gene (Sso2131) is encoded on the reverse genome strand and located downstream of a 170-amino-acid (aa)-long hypothetical protein-encoding gene (Sso2125), two permease genes (Sso2126 and Sso2127) and three pyruvate ferredoxin oxidoreductase (POR) subunit-encoding genes (Sso2128, Sso2129, Sso2130) respectively (Fig. 1A; She *et al.*, 2001). On the opposite side of *Ss-lrpB*, a gene encoding a glycerol kinase can

Fig. 1. A. Schematic representation of the genomic environment of *Ss-LrpB*. ORFs are represented by large open arrows. From left to right *Sso2125*, *Sso2126*, *Sso2127*, *Sso2128*, *Sso2129*, *Sso2130*, *Sso2131* and *Sso2133* are depicted with indication of the promoter regions (arrows).

B. EMSAs of binding of purified *Ss-LrpB* to the promoter/operator regions of *Sso2126*, *Sso2127* and *Sso2128* (*porDAB*). The positions of the wells of the gel (W), of the bound protein-DNA complexes (B) and of the free DNA (F) are indicated. *Ss-LrpB* concentrations (nM) are mentioned above each lane.

be found (*Sso2133*). All these genes are transcribed in a direction opposite to that of *Ss-LrpB*. Primer extension analysis allowed the mapping of the transcription start sites (see Fig. S1) and the prediction of the promoter elements of the *Sso2126*, *Sso2127* and *Sso2128* genes (Fig. 2). The *Sso2128*, *Sso2129* and *Sso2130* genes, annotated as *POR* δ , *POR* α and *POR* β subunits, respectively (She *et al.*, 2001), form an operon. The existence of a *porDAB* polycistronic messenger was experimentally demonstrated by performing reverse transcription polymerase chain reaction (RT-PCR) with total RNA extracted from *S. solfataricus* P2 cells (Fig. S1). The *porDAB* transcript is leaderless, but a Shine-Dalgarno motif is found 13 nt upstream of the ATG start codon of *porB*, 5'-GAGGTGA-3', which is the reverse complement of a part of the 16S rRNA 3' terminus, 5'-AUCACCUCC-3' (Torarinsson *et al.*, 2005). No apparent ribosome binding motif was found around the translational start of *porA*. The 4 nt overlap between the coding parts of the *porD* and *porA* genes suggests a mechanism of translational coupling and rephasing of the ribosomes. Analysis of the

genomic DNA of strains P2 and PBL2025 used in this work demonstrated the absence of a copy of the insertion sequence (IS) element ISC1078 in between *Ss-LrpB* and *Sso2133*. This is in contrast to the published genome sequence of *S. solfataricus* P2 (Fig. S1). However, PCR and sequencing analysis has shown that a copy of ISC1078 is inserted about 6 kbp further upstream in PBL2025 (between *Sso2126* and *Sso2127*), but not in P2 (Fig. S1). These observations pinpoint the high genomic plasticity of *S. solfataricus* (Brügger *et al.*, 2002; Redder and Garrett, 2006).

Specific regulators often control the expression of genes located adjacently or closely to the regulator gene itself. Therefore, because of the search for *Ss-LrpB* target genes, a survey of *in vitro* *Ss-LrpB* binding to the promoter regions of neighbouring genes was performed. The hypothesis of a potential regulation of adjacently located genes was further empowered by *in silico* binding site analyses of all *S. solfataricus* P2 promoter regions that revealed several theoretical *Ss-LrpB* binding sites in the promoter regions of *Sso2126*, *Sso2127* and *porD* (see

Fig. 2. A–C. Nucleotide sequences of the control regions of *porDAB* (A), *Sso2126* (B) and *Sso2127* (C). *In silico* predicted potential binding sites are indicated on top of the sequence. The actual *Ss-LrpB* binding sites are boxed. The protected regions against DNase I cleavage are indicated in a red letter type (protection at low and high *Ss-LrpB* concentration) and in an orange letter type (protection only at high *Ss-LrpB* concentration). DNase I hyperreactive sites are indicated by a red ball-and-stick symbol (hyperreactivity at low *Ss-LrpB* concentration) or by an orange ball-and-stick symbol. On the sequences of *P/O porDAB* and *P/O Sso2126*, depurination binding interference effects are represented by open light blue rectangles (weak effects), filled light blue rectangles (intermediate effects) and dark blue rectangles (strong effects). Depyrimidination binding interference effects are represented by open light green rectangles (weak effects), filled light green rectangles (intermediate effects) and dark green rectangles (strong effects). The promoter elements, transcription start sites and start codons (boxed) are displayed as well.

D. Cartoon of the *Ss-LrpB* consensus binding sequence (Peeters *et al.*, 2004) with indication of the centre of the inverted repeat (red dot), conserved palindromic half sites (arrow) with highly conserved bps in red and an A+T-rich centre (dashed line).

below). Electrophoretic mobility shift assays (EMSAs) were performed with purified Ss-LrpB and labelled DNA fragments containing the promoter regions of the above-mentioned genes (Fig. 1B). The results indicate that Ss-LrpB binds specifically and with a high affinity to the Sso2126, Sso2127 and *porD* promoter regions, suggesting that these genes belong to the Ss-LrpB regulon. At low protein concentrations, a highly specific and stable complex is formed which probably corresponds to binding of one Ss-LrpB dimer to a specific binding site. At higher protein concentrations, multiple Ss-LrpB molecules bind resulting in the formation of higher-order nucleoprotein complexes that appear less stable as indicated by the observed smearing. Supershifting of these higher-order complexes and the formation of complexes which do not enter the gel are most likely due to protein aggregation (Peeters *et al.*, 2006). The highest binding affinity was observed with a 307 bp fragment containing the Sso2126 promoter region (ranging from -263 to +44). The overall apparent equilibrium dissociation constant K_D of this interaction, 10 nM, is comparable with the K_D for binding to the control region of the *Ss-LrpB* gene itself (Peeters *et al.*, 2004). The apparent K_D s of binding to a fragment encompassing the promoter regions of Sso2127 (-209 to +59) or *porD* (-138 to +57) were slightly higher, about 30 and 40 nM respectively (Fig. 1B). No specific binding was observed to the promoter regions of Sso2125 and Sso2133 (data not shown).

High resolution contact mapping of Ss-LrpB-porDAB operator interactions

In the control region of the *por* operon, a cluster of several potential binding sites for Ss-LrpB was predicted (Fig. 2A): the partially overlapping pot Box 3 and pot Box 3* sites, the partially overlapping pot Box 2 and pot Box 2* sites and the most downstream pot Box 1. Based on the *in silico* predictions alone, the binding site which is actually recognized by the regulator is not evidently allocated, since both alternative binding sites share a highly conserved half site (Fig. 2D). Therefore, *in vitro* binding to this control region was further studied by employing a combination of footprinting, premodification binding interference and mutant studies. DNase I footprinting revealed three discrete regions of protection both for the top and bottom strand (Fig. 3A), indicating the existence of three separate binding sites (Fig. 2A). Binding to these sites is highly cooperative, as indicated by the fast formation of the slowest migrating specific complex in the EMSA experiment, which can be assumed to correspond to a complex having the three regions bound. The positions of the corresponding protected regions on the two strands are slightly shifted with respect to one another in the 3' direction due to the conformation of the DNA helix. DNase I

hyperreactivity was observed at several regularly spaced positions in the operator region (Figs 2A and 3A).

For the unambiguous identification of the binding sites, EMSAs were performed with mutated operator fragments (Fig. 4A). These operator mutants contained a single bp substitution predicted to either impair or improve the theoretical binding affinity to one of the potential Boxes. All these substitutions are targeted at position -5 or +5 of the corresponding potential site (Figs 2D and 4D), for which it is crucial that it contains a G-C or C-G bp respectively (Peeters *et al.*, 2007). Replacing G-C at position -5 of pot Box 3 (sub1) resulted in the formation of the specific complex with a much lower affinity while substituting C-G at position +5 of pot Box 3* (sub2) did not noticeably affect binding (Fig. 4A). For pot Box 2, the T-A bp at position -5 was substituted by a more optimal G-C bp (sub3); binding to this mutant operator occurred with a remarkably higher affinity as compared with binding to the wild-type fragment. In contrast, substitution of C-G at position +5 of pot Box 2* (sub4) had hardly any effect on complex formation. These results strongly indicate recognition by Ss-LrpB of the most upstream Box of either of the pairs of alternative sites. Furthermore, EMSAs with 45 bp oligonucleotide duplexes containing the isolated (potential) binding sites demonstrated specific binding to pot Box 1 and pot Box 3 (called Box 1 and Box 3 hereafter), but not to any of the other fragments (Fig. 4B). Both pot Box 2 and pot Box 2* exhibit no detectable specific binding but the previous observation of improved binding to the pot Box 2 operator mutant (sub3) proves the involvement of the prior (called Box 2 hereafter) in the context of the full-length operator. Box 3 is bound with a slightly higher affinity than Box 1 (Fig. 4B).

Depurination binding interference (Brunelle and Schleif, 1987) allowed a higher resolution analysis and screening of the purines which are important for recognition by Ss-LrpB (Fig. 3B). A DNA preparation of the *porD* operator fragment was obtained in which purines were sparingly removed, with an average of one modification per DNA molecule. This population of modified DNA molecules was then subjected to Ss-LrpB binding in an EMSA. The various populations of molecules [input (I) and free (F) DNA and bound complexes (B1 and B3)] were recovered from the gel, cleaved at the depurinated positions and analysed by denaturing gel electrophoresis. DNA molecules missing a purine at a position that is important for Ss-LrpB binding, will be under-represented in the bound population while molecules that miss a purine at a position which is irrelevant for binding, will be equally represented in the bound and free populations. The removal of purines at several positions had an effect on binding (Fig. 2A). Most of these effects were located in or at the border of the three previously identified binding sites. In between Box 3 and Box 2, a few weak effects were also observed

Fig. 3. A. DNase I footprinting of the binding of Ss-LrpB to the control region of *porDAB* with either top strand or bottom strand labelled. The applied Ss-LrpB concentrations are indicated. A+G and C+T represent the corresponding Maxam-Gilbert sequencing ladders. Regions which are protected against DNase I cleavage are indicated with a vertical line; hyperreactive sites are pointed out with a ball-and-stick symbol. The position of the identified Boxes is indicated.

B. Depurination binding interference experiment of Ss-LrpB binding to the control region of *porDAB* with either top strand or bottom strand labelled. To the left, the EMSA with depurinated DNA is shown, having the top strand labelled. Input DNA (I), free DNA (F) and bound complexes (B1 and B3) are indicated. The observed effects are indicated with a horizontal line and the position of the Boxes is shown to the left of the autoradiographs.

for A-96', A-95', G-91, A-90 and A-89. All effects were observed for bound population B3, corresponding to the complex that is assumed to have all three binding sites bound. Concerning B1, no clear-cut effects could be

observed (Fig. 3B). This might be explained by the heterogeneity of B1 complexes, containing DNA fragments bound each at a different site. Therefore, creating an abasic site in one binding site does not affect binding to

Fig. 4. A. EMSAs of the binding of Ss-LrpB to 195 bp *porDAB* operator fragments with single bp substitutions in the (potential) boxes. The position of the free DNA (F) and bound complexes (B) is indicated, as is the concentration of Ss-LrpB (nM). In the EMSAs with *P₀porDAB* pot Box 2 mutant and with *P₀porDAB* pot Box 3 mutant, an additional DNA population is apparent (migrating slightly slower than the fastest migrating complex), which is thought to correspond to single-strand DNA. This DNA is not depleted by Ss-LrpB however.

B. EMSAs of the binding of Ss-LrpB to 45 bp oligonucleotide duplexes containing separate (potential) Boxes with indication of the position of the free DNA (F) and bound complexes (B) and the used concentration of Ss-LrpB (nM).

C. EMSAs of the binding of Ss-LrpB to oligonucleotide duplexes containing truncated parts of the *porDAB* control region (combination of two adjacent Boxes). The length of these duplexes is 55 and 54 bp respectively. The position of the free DNA (F) and bound complexes (B) is indicated, as is the concentration of Ss-LrpB (nM).

D. Schematic representation of the *porDAB* and *Ss-LrpB* control regions, aligned to each other at the respective transcription start sites. The positions of the BRE and TATA box elements and the different (potential) Ss-LrpB binding sites (with sequences) are shown. On the *P₀porDAB* presentation, the different single bp substitutions are illustrated (sub1, sub2, sub3 and sub4).

the other binding sites. The depurination binding interference indicates that Box 2 (not Box 2*) and Box 3 (not Box 3*) are bound by Ss-LrpB and is therefore in full agreement with the mutant operator binding studies.

Intriguingly, the binding site organization in the control region of *porD* is almost exactly identical as the organization in the control region of *Ss-LrpB* itself (Fig. 4D; Peeters *et al.*, 2004). Box 2 is separated from Box 3 and Box 1 by a 16 bp and 17 bp linker respectively. Also common to both is that the central Box 2 has a very low affinity whereas the outer Boxes are high-affinity sites. Both outer binding sites are required to stimulate binding to the central site through cooperative actions. This is demonstrated by performing EMSAs with truncated fragments, containing either Box 2

+3 or Box 1 +2 (Fig. 4C). The presence of only one of the two high-affinity sites does not suffice to promote stable binding to Box 2. This is especially the case for the Box 1 +2 combination: there is no detectable formation of a complex with the two sites bound.

Ss-LrpB binding to the control regions of the two permease genes, *Sso2126* and *Sso2127*

An *in silico* approach combining different methods (see *Experimental procedures*) indicated the existence of two potential Ss-LrpB binding sites in the promoter/operator (*P₀*) region of *Sso2126*: one at position -61 (pot Box 1) and a second site much further upstream starting at posi-

tion -220 (pot Box 2) (Fig. 2B). Characterization of the recognized binding sites in this promoter region was done by footprinting and premodification binding interference experiments (Fig. 5). At low Ss-LrpB concentration, DNase I footprinting resulted in clear protection in a region corresponding to pot Box 2 (called Box 2 hereafter; Figs. 5A and 2B). At higher concentrations (> 336 nM), this protection extended further downstream, encompassing almost the entire control region including pot Box 1 and the promoter elements. DNase I hyperreactivity was observed at the borders of Box 2 and, at higher Ss-LrpB concentrations, between positions -134/-135 and in pot Box 1 (between positions -50/-51 and -60/-61). This indicates minor groove widening due to protein-induced DNA deformations. To confirm the identity of the recognized Box in the highly specific, fastest migrating complex in the EMSA, which is already formed at very low Ss-LrpB concentration (> 0.8 nM; Fig. 1B), in-gel Cu-phenantroline footprinting was performed (Fig. 5B). Indeed, protection is only apparent in the upstream region encompassing Box 2 indicating that the bound population is uniform and composed only of nucleoprotein complexes with Ss-LrpB bound specifically to this upstream Box. Furthermore, Ss-LrpB exhibits binding to a 45 bp oligonucleotide duplex containing the Box 2 sequence but not to a duplex containing the pot Box 1 sequence (Fig. 5C and D). DNase I and in-gel footprinting was also performed with a probe having the downstream end located closer to Box 2, allowing a higher resolution contact mapping (Fig. 5E). Whether the top strand or the bottom strand was labelled, the protection was situated at Box 2 (Fig. 2B, top strand shown only). Protection against DNase I was most apparent at the two half sites of the Box (Fig. 5E). Depurination and depyrimidation binding interference experiments revealed that almost all residues inside or at the borders of Box 2 are involved in the interaction between Ss-LrpB and this binding site (Fig. 5F).

The *in silico* approach also indicated the presence of two potential binding sites in the promoter/operator region of Sso2127: one at position -106 (pot Box 1) and a second at position -235 (pot Box 2) (Fig. 2C). The sequence of pot Box 1, 5'-TTGCATTTTATGCAA-3', corresponds well to the Ss-LrpB consensus site (Fig. 2D; Peeters *et al.*, 2004), having two perfect half sites and an A+T-rich centre. At the lower protein concentrations used, DNase I footprinting of Ss-LrpB:Sso2127 complexes showed one discrete region of protection extending from position -126 to -85, indicating specific binding of pot Box 1 (called Box 1 hereafter; Fig. 6A). Two sites of DNase I hyperreactivity border Box 1. The hypothesis that recognition of Box 1 is responsible for complex formation, is empowered by an EMSA with a 45 bp oligonucleotide duplex containing the Box 1 site (Fig. 6B). The K_D of this interaction is comparable to the K_D of binding to the entire operator fragment. At

higher Ss-LrpB concentrations, the DNase I protection zone is enlarged by extending both upstream and downstream and several additional sites of hyperreactivity are generated indicating conformational changes in the DNA (Figs. 6A and 2C). The downstream protection encompasses both the BRE and TATA box promoter elements. These data cannot exclude the involvement of pot Box 2 in the regulation of Sso2127, since all binding experiments were performed with a DNA fragment starting at position -206 of the Sso2127 control region. However, no binding of Ss-LrpB to this isolated pot Box 2 present on a 45 duplex could be detected (data not shown).

Gene expression analysis of target genes in an Ss-LrpB gene disruption mutant

To study the effects of the Ss-LrpB regulator *in vivo*, a gene disruption mutant was constructed (Fig. 7), as described (Albers and Driessen, 2007). Targeted gene disruption of *Ss-lrpB* was achieved by electroporation of strain PBL2025, which is lacking the *lacS* gene and is unable to catabolize lactose (Schelert *et al.*, 2004), with the pET2268*Ss-lrpB* knockout plasmid. This plasmid contains the flanking regions of *Ss-lrpB* and a *lacS* cassette in between. Homologous recombination resulted in the integration of the *lacS* cassette at the *Ss-lrpB* locus after selection on lactose minimal medium (Fig. 7A). Three rounds of single colony picking, culturing in liquid medium and replating, were necessary before the *Ss-lrpB::lacS* mutant was obtained with complete absence of the wild-type allele. This was demonstrated by PCR analysis (data not shown) and Southern blotting using *NdeI*-restricted genomic DNA (Fig. 7B). The growth of the *Ss-lrpB::lacS* strain is not affected in medium supplemented with either tryptone or glucose as compared with the growth of PBL2025 (data not shown).

Expression of Sso2126, Sso2127 and the *por* operon was notably lower in the *Ss-lrpB::lacS* strain, as observed by RT-PCR (Fig. 8A). This was confirmed and quantified by real-time quantitative PCR (Table S1; Fig. 8B), indicating a 3.1- and 5.9-fold lower expression for Sso2127 and *porDAB* respectively. The effect was less pronounced for Sso2126, for which the relative expression level is only 0.70. These results lead us to conclude that Ss-LrpB acts as an activator in the transcriptional regulation of its three neighbouring target genes under the given growth conditions.

Discussion

Overall, we show that Ss-LrpB activates the expression of a neighbouring *por* operon and two putative permease genes by binding to (multiple) binding sites in the corresponding control regions. Target genes of only a few

archaeal Lrp-like regulators have been identified. Ss-LrpB appears to share its physiological role with Ptr2 from *M. jannaschii*. We noticed that the first gene of the *porDAB* operon regulated by Ss-LrpB, annotated to encode a POR delta subunit, is highly similar to *fdxA*, one of the identified targets of Ptr2 (Ouhammouch *et al.*, 2003). The 61 aa long C-terminal part of *porD* has an aa sequence identity of 41% with *fdxA*, which codes for a ferredoxin. POR catalyses the oxidative carboxylation of pyruvate into acetyl-coA, thereby reducing ferredoxin. This enzyme is active in all archaea, whether they grow under anaerobic or aerobic conditions. It belongs to the

family of two oxoacid:ferredoxin oxidoreductases (OR), that can recognize several different substrates, all performing functions in the central or amino acid metabolism (besides POR, 2-ketoglutarate OR catalyses a citric acid cycle reaction and indolepyruvate OR oxidizes 2-keto acids generated from aromatic amino acids; Schut *et al.*, 2001). Furthermore, POR is also important in performing the reversed reaction of acetyl-coA reduction to pyruvate during autotrophic growth with fixation of CO₂ via a reductive citric acid cycle or during growth on acetate for the functioning of the gluconeogenesis. Some OR enzymes have a broad substrate specificity, not limited to one of the

Fig. 5. A. DNase I footprinting of the binding of Ss-LrpB to the control regions of Sso2126 with a fragment prepared with oligonucleotides DC348(f) and DC349(r) having the bottom strand labelled. The used Ss-LrpB concentrations are mentioned above each lane. A+G and C+T represent the Maxam-Gilbert sequencing ladders. Regions which are protected against DNase I cleavage are indicated with a vertical line; hyperreactive sites are pointed out with a ball-and-stick symbol. Additional protection regions and hyperreactive sites that are observed at higher protein concentrations are indicated with a dashed line and a dashed symbol respectively. The position of the identified Box 2 is indicated to the right of the autoradiograph.

B. In-gel Cu-phenantroline footprinting of Ss-LrpB binding to % Sso2126 [using the same DNA fragment as used in the DNase I footprinting experiment shown in (A)]. On top, the EMSA is shown which was used for the in-gel footprinting experiment. The populations of input (I), free (F) and bound (B) DNA are boxed. To the left of the footprint autoradiograph, the protected region is indicated with a vertical line. The position of the identified Box 2 is indicated to the right of the autoradiograph. A+G and C+T represent the Maxam-Gilbert sequencing ladders.

C. EMSA of binding of Ss-LrpB to a 45 bp oligonucleotide duplex with potential Box 1 of % Sso2126. The applied Ss-LrpB concentrations are mentioned, as is the position of free DNA (F).

D. EMSA of binding of Ss-LrpB to a 45 bp oligonucleotide duplex with Box 2 of % Sso2126. The applied Ss-LrpB concentrations are mentioned, as are the positions of free (F) and bound (B) DNA molecules.

E. Footprinting experiments of the binding of Ss-LrpB to % Sso2126 using a fragment prepared with oligonucleotides DC348(f) and DC707(r) having the bottom strand labelled. Protected regions are indicated with a vertical line, hyperreactivity sites with a ball-and-stick symbol and the position of Box 2 is shown as well. The first five lanes represent the DNase I footprint with the used Ss-LrpB concentrations indicated on top. A+G and C+T represent the Maxam-Gilbert sequencing ladders. F and B represent the free and bound populations of the in-gel footprinting experiment. The EMSA used for the in-gel footprinting is comparable to that shown in the top panel of (B).

F. Depurination and depyrimidation binding interference experiments of the binding of Ss-LrpB to % Sso2126 with a fragment prepared with oligonucleotides DC348(f) and DC707(r) having the top strand labelled. In the top panel, an EMSA with depyrimidated DNA is shown. The EMSA with depurinated DNA is similar. Populations of input DNA (I), free DNA (F) and the bound complex (B) are shown. The observed effects are indicated with a horizontal line and the position of Box 2 is shown to the right of the autoradiograph.

three classes of the above-mentioned substrates. This is also the case for the only characterized OR of *Sulfolobus* thus far (an $\alpha\beta$ heterodimer, encoded by Sso2815 and Sso2816), which has been shown to catalyse all three key reactions (Zhang *et al.*, 1996; Fukuda and Wakagi, 2002). Based on the fact that the expression of the POR enzyme studied here seems to be tightly regulated, it can be speculated that the enzyme is responsible for catalysis of a reaction that is involved in a more specific growth condition, rather than a central metabolic reaction. Previously, it has been shown that central metabolic enzymes in *S. solfataricus* undergo very little transcriptional and translational regulation (Snijders *et al.*, 2006). The POR enzyme subject of this study is annotated in the genome sequence as being specific for pyruvate (She *et al.*, 2001), although it has never been characterized biochemically. The subunit composition of this putative enzyme (composed of δ , α and β) appears to be novel, although there is a large variation in the subunit compositions for archaeal OR enzymes while the catalytic core is highly conserved.

The function of the two permeases, encoded by Sso2126 and Sso2127, is as yet unknown. Blast analyses indicate that Sso2126 exhibits homology with bacterial L-lactate permeases and that Sso2127 is related to halophilic oxalate/formate antiporters. The link between the POR enzyme and the two permeases, explaining their coregulation by a specific transcriptional regulator, and the exact functions of these proteins in the cell is currently not clear. A very tentative link can be made with the metabolism of the sulphate-reducing anaerobic euryarchaeote *Archaeoglobus fulgidus* (Klenk *et al.*, 1997). This organism is able to metabolize lactate during sulphate reduction. Part of the lactate oxidation pathway is

the conversion of pyruvate in acetyl-coA and CO₂, catalysed by POR (Labes and Schönheit, 2001); incomplete lactate oxidation can result in the production of formate (Habicht *et al.*, 2005). To the best of our knowledge, growth of *S. solfataricus* on lactate has not been reported.

Many Lrp-like regulators bind amino acids allowing a response in function of the *in vivo* concentration of the involved amino acid(s). Depending on the regulator, ligand specificity can vary notably. Some regulators have a broad aa specificity range while other regulators are restricted to binding to only one type of amino acid (e.g. LysM which can only bind lysine; Brinkman *et al.*, 2002). Several ligand-bound cocrystal structures of Lrp-like regulators have recently been determined (Thaw *et al.*, 2006; Okamura *et al.*, 2007; Ren *et al.*, 2007; Shrivastava and Ramachandran, 2007; Kumarevel *et al.*, 2008; Reddy *et al.*, 2008). These structures allowed the identification of residues that are important in the formation of the ligand-binding pocket and the establishment of specific interactions with the ligand. The aa conservation of Ss-LrpB of these residues is very low (Fig. 9), and this is also observed for Ptr2 and LrpA. The binding pocket is formed between the loop connecting two β strands of one monomer (β 3 and β 4) and β 5 of another subunit that forms a dimer with the former. The weak conservation is especially observed for the loop, in which only a glycine is conserved that is present in all Lrp-like regulators studied so far (position 37 of the C-terminal domain; Thaw *et al.*, 2006). This suggests that if Ss-LrpB is bound by an effector molecule, this might be another small molecule apart from an amino acid.

The *porDAB* operator exhibits a binding site organization identical to that in the control region of the *Ss-lrpB* gene itself (Fig. 5D; Peeters *et al.*, 2004). It is highly

plausible that the formation of the different complexes follow the same sequential assembly as shown for the autoregulation. In contrast, the architecture of the Sso2126 and Sso2127 operators is fundamentally different with only one binding site that is bound specifically with a high affinity. At a higher protein concentration, this binding extends further downstream (and in the case of Sso2127, somewhat upstream as well) until the protection reaches and overlaps the promoter elements. Suppos-

Fig. 6. A. DNase I footprinting of the binding of Ss-LrpB to the control region of Sso2127 having the top strand labelled. The used Ss-LrpB concentrations are mentioned above. A+G and C+T represent the Maxam-Gilbert sequencing ladders. Regions which are protected against DNase I cleavage are indicated with a vertical line; hyperreactive sites are pointed out with a ball-and-stick symbol. Additional protection regions and hyperreactive sites that are observed at higher protein concentrations are indicated with a dashed line and a dashed symbol respectively. The position of Box 1 is indicated to the right of the autoradiograph. B. EMSA of binding of Ss-LrpB to a 45 bp oligonucleotide duplex with Box 1 of % Sso2127. The applied Ss-LrpB concentrations are mentioned, as are the positions of free DNA (F) and bound complexes (B).

edly, one Ss-LrpB dimer which is bound to the core high-affinity Box triggers the cooperative binding of additional dimers. It can be noted that Ss-LrpB binding to the operator fragments of all three regulatory targets results in the formation of rather unstable higher order complexes and that the K_D s of these interactions are somewhat higher as compared with binding to the *Ss-lrpB* control region (overall K_D of about 10 nM; Peeters *et al.*, 2004). This might be explained by the lack of an effector molecule in the *in vitro* binding experiments, which is supposedly not required for autoregulation but would enhance complex formation in the control regions of the target genes. It is worth noting that most Lrp-like regulators perform an effector-independent autoregulation (Kölling and Lother, 1985; Madhusudhan *et al.*, 1993; 1995; Keuntje *et al.*, 1995; Jafri *et al.*, 1999).

The observation of DNase I hyperreactivity at several sites in the three target operator regions (mainly at the borders of the binding sites and in between two juxtaposed Boxes) points out that binding of Ss-LrpB notably induces DNA deformations. Previously we have shown that at each specifically bound Box in the *Ss-lrpB* operator a local bending is induced with an average bending angle of about 50° (Peeters *et al.*, 2004). In the case of % *porDAB*, the organization of the binding sites ensures that these local bending angles are additive upon formation of the higher-order nucleoprotein complex and additional architectural deformations such as DNA wrapping could occur (similar to the situation in autoregulation; Peeters *et al.*, 2006). In the case of % Sso2126 and % Sso2127, additional less-specific binding can also lead to increased DNA bending, and even DNA wrapping and/or looping. DNA wrapping is a basic propensity of Lrp-like proteins. Based on octameric crystal structures, several structural models of Lrp-DNA complexes in which the DNA is wrapped around an octameric molecule have been proposed (Leonard *et al.*, 2001; Koike *et al.*, 2004; Thaw *et al.*, 2006; de los Rios and Perona, 2007; Yokoyama *et al.*, 2007). DNA wrapping is not only important when performing regulation of gene expression, but also when playing a more general role in DNA compaction,

Fig. 7. A. Schematic overview of the genotype of PBL2025 and the *Ss-LrpB::lacS* gene disruption strain (with indication of the position of the NdeI restriction site in the ORF of *Ss-LrpB*). Arrows indicate start points of transcription.

B. Southern blot experiment with NdeI-restricted gDNA of PBL2025 and *Ss-LrpB::lacS* and a combination of *Ss-LrpB* and *lacS* probes. *Ss-LrpB* contains an internal NdeI site in contrast to the *lacS* cassette; therefore, the labelled *Ss-LrpB* probe hybridizes to a 1945 bp fragment and the *lacS* probe to a 6139 bp fragment.

Fig. 8. A. Gene expression analysis by means of RT-PCR comparing the expression of the target genes and 23S rRNA in either the *Ss-LrpB::lacS* gene disruption strain or PBL2025. Products were analyzed on gel after a PCR with either 15 or 30 cycles, as indicated.

B. Scheme of the average relative expression levels of Sso2126, Sso2127 and *porDAB* in either the *Ss-LrpB::lacS* gene disruption strain or PBL2025, based on triplicate real-time qPCR experiments. These data, including standard deviations, are shown in Table S1.

Fig. 9. Multiple amino acid sequence alignment of the C-terminal domains of Lrp-like proteins (constructed with ClustalW). The sequences shown here are: Ss-LrpB from *S. solfataricus* (SsLrpB), LrpA from *P. furiosus* (LrpA; Leonard *et al.*, 2001), Ptr2 from *M. jannaschii* (Ptr2; Ouhammouch *et al.*, 2003), DM1, a stand-alone RAM domain, from *Pyrococcus* (DM1; Okamura *et al.*, 2007), AsnC from *E. coli* (AsnC; Thaw *et al.*, 2006), LrpA from *Mycobacterium tuberculosis* (MtbLrp; Shrivastava and Ramachandran, 2007; Reddy *et al.*, 2008), Grp from *Sulfolobus tokodaii* (Grp; Kumarevel *et al.*, 2008), Lrp from *Neisseria meningitidis* (NmLrp; Ren *et al.*, 2007). Residues that are identified to play an important role in the interaction with the ligand, and are printed bold. These observations are made based on ligand-bound cocrystal structures, in the references mentioned above. Below the alignment, the secondary structure elements are illustrated.

chromosome structure and organization. This has been shown for several bacterial homologues (*E. coli* Lrp, LrpC from *Bacillus subtilis*; d'Ari *et al.*, 1993; Tapias *et al.*, 2000; Beloin *et al.*, 2003), and also for Smj12 from *S. solfataricus* (Napoli *et al.*, 2001).

Despite the dissimilar operator organizations, the regulatory outcome for the different target genes is identical in the given growth conditions: Ss-LrpB exerts an activation at all three promoters. The difference in expression between WT and an *Ss-lrpB* gene disruption mutant is highest for *porDAB* (5.9-fold upregulation in the WT), followed by Sso2127 (3.1-fold) and Sso2126 (1.4-fold). Although the latter value is too low to be significant when determined by real-time qPCR, taken together with the fact that Ss-LrpB binds this % region *in vitro*, it might still indicate the possibility of an increased Ss-LrpB activation effect in other growth conditions. The high-affinity Box in % Sso2126 is located more upstream than in the other two operators (position -220) and possibly the *in vivo* concentration of Ss-LrpB in the applied growth conditions (exponential growth in rich medium) is too low to promote sequential binding all the way down to the promoter region. The precise mechanism of Ss-LrpB activation at the *porDAB* promoter remains to be elucidated, but might be similar to the mechanism employed by Ptr2. Ptr2 activates transcription by direct recruitment of TBP to the TATA box (Ouhammouch *et al.*, 2003). The minimal required upstream activating sequence for Ptr2 regulation is a binding site about 6–8 bp upstream of the TATA box. Therefore, it is very likely that Ss-LrpB activation is exerted by the establishment of protein–protein contacts between Ss-LrpB bound at the most downstream Box 1 and TBP. Box 1 is separated by 7 bp of the TATA box and only by 1 bp of the BRE element (Fig. 3A). Despite this very close spacing, it might be hypothesized that simultaneous binding of Ss-LrpB to Box 1 and TBP/TFB

to the promoter elements is possible. Simultaneous binding of the regulator and the transcription factors has previously been shown to occur in the % *Ss-lrpB*, in which Box 1 and BRE overlap with 1 bp. Furthermore, based on the crystal structure of the TBP/TFB/DNA complex (Littlefield *et al.*, 1999), it might be hypothesized that Ss-LrpB and TBP/TFB bind opposing faces of the DNA helix thereby avoiding a steric clash. It is difficult to predict the effect of the protein-induced DNA conformational changes upon the regulatory outcome, although it is likely that DNA bending and wrapping affect the initiation frequency. Together with the fact that different complexes are formed in a concentration-dependent manner, this suggests that different regulatory responses might be generated depending on the Ss-LrpB concentration. For the autoregulation, a hypothetical regulatory switch model has been proposed in which it is hypothesized that at low concentrations, upon occupation of only one Box, an activation occurs while at higher concentrations, when all three Boxes are bound, DNA deformations lead to repression (Peeters *et al.*, 2004; 2006).

Here, we describe the first construction of a gene disruption mutant of an archaeal *lrp* homologue. In the future, this mutant will allow an even more thorough genome-wide investigation of the complete Ss-LrpB regulon. Bacterial Lrp-like regulators, having either a global or a specific function, are always involved in modulating the expression of genes that are linked to the amino acid metabolism. The as yet identified ligands that bind these regulators are without exception amino acids. Now, it is becoming clear that the archaeal branch of the Lrp family contains regulators that can fulfil a wider variation of physiological tasks in the cell that are not confined to the amino acid metabolism. Based on their target genes that have been identified so far, and the proposed lack of an amino acid binding pocket in the C-terminal domain,

Ss-LrpB seems to constitute a separate functional Lrp subclass together with the euryarchaeotic orthologues Ptr2 and LrpA (which have 33% and 32% aa sequence identity with Ss-LrpB respectively; Peeters *et al.*, 2004). Their physiological role is related to regulating redox or energy processes in the cell. Furthermore, it is very likely that Ss-LrpB employs the same molecular mechanism of activation as Ptr2, at least for one of the promoters. Here, activation might occur through Ss-LrpB-mediated recruitment of TBP to the TATA box. It is obvious that we are only beginning to understand the enormous complexity of this bacterial/archaeal regulator family with regard to their physiological roles, regulatory mechanisms and ligand responses.

Experimental procedures

In silico binding site identification

A genome-wide *in silico* binding site prediction was done with the Ss-LrpB binding energy position weight matrix (Peeters *et al.*, 2007), taking cooperativity into account (van Oeffelen *et al.*, 2008). This analysis will be published elsewhere. In addition, based on the sequences of known Ss-LrpB binding sites prediction of sites was performed by employing the web server PredictRegulon (<http://www.cdfd.org.in/predictregulon/>; Yellaboina *et al.*, 2004). Finally, some potential binding sites were found by visual inspection.

Strains and growth conditions

Sulfolobus solfataricus P2 (DSM1617) and PBL2025 (Schertl *et al.*, 2004) were grown aerobically at 80°C in medium described by Brock, supplemented with either 0.1% tryptone or 0.4% glucose as carbon and energy source (Brock *et al.*, 1972). Prior to transformation, *S. solfataricus* PBL2025 was grown in medium supplemented with both 0.1% tryptone and 0.2% sucrose. For selection of transformants, 0.4% lactose was added to the basic mineral medium. Growth of the cultures was followed by measuring the optical density at 660 nm.

Sulfolobus solfataricus was grown on plates containing Brock medium with 0.64% Gelrite and with 0.1% tryptone as a sole carbon and energy source (Grogan, 1989). Colonies were obtained after incubation at 78°C during 5–7 days.

Escherichia coli MC1061 or TOP10 (Invitrogen) was used for propagation and production of plasmid DNA and grown in rich medium (853; Glansdorff, 1965) with 50 µg ml⁻¹ ampicillin. Growth was also followed by measuring the optical density at 660 nm.

DNA and RNA extractions

Genomic DNA was extracted from 4 ml of *S. solfataricus* culture with a QuickPick SML gDNA kit (magnetic bead purification; BioNobile). Plasmid DNA was extracted from *E. coli* by means of the alkaline lysis method with a Miniprep kit

(Qiagen; Birnboim and Doily, 1979). RNA extraction from *S. solfataricus* was done with an RNeasy mini kit (Qiagen), after performing cell lysis with proteinase K treatment. To prevent genomic DNA contamination, RNA samples were treated on-column with DNase I following the manufacturer's instructions (Qiagen).

Plasmid constructions

The promoter regions of Sso2126, Sso2127 and *porD* were amplified from genomic DNA [primer pairs DC348(f)/DC349(r), DC350(f)/DC351(r) and DC104(f)/DC178] and cloned into the pCR2.1.TOPO vector using a TOPO TA cloning kit (Invitrogen). Site-directed mutagenesis of the *porDAB* control region was performed with the overlap PCR method (Higuchi *et al.*, 1988) with inside primers DC518(f)/DC519(r) (Box 3 mutant), DC520(f)/DC521(r) (Box 3* mutant), DC522(f)/DC523(r) (Box 2 mutant) or DC524(f)/DC525(r) (Box 2* mutant) and outside primers DC104(f)/DC178.

The pET2268*Ss-lrpB* knockout plasmid was constructed by cloning the 909 bp *Ss-lrpB* upstream region, amplified with DC445(f) and DC446(r), in pET2268 (Szabo *et al.*, 2007) using the KpnI and NcoI restriction sites. The downstream region was amplified with DC449(f) and DC448(r), producing a 1068 bp fragment, and cloned into the BamHI and NotI restriction sites. A list of all the oligonucleotides used in this work is provided in the *Supplementary material* (Table S2).

EMSAs, footprinting and binding interference

Recombinant Ss-LrpB was overexpressed in *E. coli* and purified by a combination of heat treatment and ion exchange chromatography, as described (Peeters *et al.*, 2006). Protein : DNA interaction experiments were performed with purified Ss-LrpB and DNA fragments generated either by PCR by using ReadyMix Taq PCR Mix (Sigma-Aldrich) or by complementary oligonucleotide annealing. The following oligonucleotides were used for annealing: DC283(f)/DC284(r) (‰ *porDAB*Box 1 + 2 truncated fragment), DC285(f)/DC286(r) (‰ *porDAB*Box 2 + 3 truncated fragment), DC533(f)/DC534(r) (‰ *porDAB* pot Box 1), DC535(f)/DC536(r) (‰ *porDAB* pot Box 2), DC539(f)/DC540(r) (‰ *porDAB* pot Box 3), DC537(f)/DC538(r) (‰ *porDAB* pot Box 2*), DC541(f)/DC542(r) (‰ *porDAB* pot Box 3*), DC384(f)/DC385(r) (‰ Sso2126 pot Box 1), DC703(f)/DC704(r) (‰ Sso2126 Box 2) and DC386(f)/DC387(r) (‰ Sso2127 Box 1). All primer sequences are listed in Table S2. For fragments containing the promoter regions of Sso2126, Sso2127 and *porDAB* (including *porDAB* operator mutants), the same primer pairs were used as for the cloning. A shorter ‰ Sso2126 fragment, with the downstream end located closer to Box 2, was prepared using DC348(f) and DC707(r). All fragments or oligonucleotide duplexes were labelled by having one primer 5' end labelled with [γ -³²P]-ATP (GE Healthcare). Prior to EMSA, footprinting or binding interference analysis, the DNA fragments were purified by polyacrylamide gel electrophoresis.

Electrophoretic mobility shift assays were performed as described previously (Enoru-Eta *et al.*, 2000). All binding reactions were done at 37°C in LrpB binding buffer (Peeters

et al., 2007) and in the presence of 25 µg ml⁻¹ sonicated herring sperm DNA as non-specific competitor. The binding dissociation constants K_D were determined by estimating the half-saturation point. DNase I footprinting (Enoru-Eta *et al.*, 2000), in-gel Cu-phenantroline footprinting (Peeters *et al.*, 2004) and depurination/depyrimidation binding interference experiments (Wang *et al.*, 1998) were performed as described previously, using LrpB binding buffer to perform the binding reactions. Reference ladders were generated by chemical sequencing methods (Maxam and Gilbert, 1980).

Gene disruption

For inactivation of *Ss-lrpB* in PBL2025, the *S. solfataricus* gene disruption procedure was employed as described (Albers and Driessen, 2007).

Southern blotting

Southern blotting was performed as described (Zolghadr *et al.*, 2007), with the following modifications: 5 µg of genomic DNA (extracted from *S. solfataricus* PBL2025 or PBL2025*Ss-lrpB::lacS*) was digested overnight with NdeI and run on a 0.8% agarose gel. For the blotting, nylon membrane Hybond-NX from GE Healthcare was used. Probes representing *lacS* and *Ss-lrpB* were PCR-amplified by using PBL2025 genomic DNA and primers DC559(f)/DC560(r) and DC497(f)/DC498(r) respectively. Subsequently, these fragments were digoxigenin-labelled by using hexanucleotide mix, dNTP labelling mix and Klenow enzyme (Roche Applied Science). Detection was done with a GeneGnome chemiluminescence detector (Syngene).

Primer extension analysis, cDNA synthesis and RT-PCR

Primer extension analysis of *S. solfataricus* P2 RNA was performed as described previously (Enoru-Eta *et al.*, 2002) by using AMV Reverse Transcriptase (Roche Applied Science). In each experiment, 100 µg RNA was used. For Sso2126 transcription start determination, primer DC349(r) was used, for Sso2127 primer DC351(r) and for *porDAB* primer DC178.

cDNA was prepared from total RNA extracted from *S. solfataricus* PBL2025 or *Ss-lrpB::lacS* with Superscript II Reverse Transcriptase (Invitrogen). About 2 µg of total RNA was incubated with 200 ng random primers and the reaction was performed following the manufacturer's instructions. This was followed by RNase H treatment (Fermentas). cDNA was then used as a template in PCR reactions to analyse *porDAB* operon structure, by using primers DC569(f), DC570(r), DC571(f) and DC572(r) and to analyse gene expression of Sso2126, Sso2127, *porD* and 23S rRNA using primer pairs DC573(f)/DC574(r), DC575(f)/DC576(r), DC577(f)/DC578(r) and DC579(f)/DC580(r) respectively. All primer sequences are listed in Table S2.

Real-time quantitative PCR

First-strand cDNA synthesis (using 1 µg of total RNA) and preparation of the PCR reactions was done with the SYBR

GreenER two-step qRT-PCR kit for iCycler (Invitrogen). The amplicon sizes were between 100 and 300 bp. For each reaction, 1 µl of a 10⁻¹ dilution was used as a template. The primer pairs were the same as used in the RT-PCR analysis described above. Primer concentrations were optimized for each gene by calibration with genomic DNA. Real-time qPCR was carried out in an iCycler IQ (Bio-Rad) using the following protocol: 50°C for 2 min, 95°C for 8.5 min and 40 cycles of 95°C for 15 s and 60°C for 60 s. This was followed by a melting curve analysis. Three replicates per primer pair were performed. The threshold and threshold cycles (C_T) were determined for each reaction and the ratio of gene expression in the *Ss-lrpB::lacS* versus PBL2025 was quantified with the method of Pfaffl (2001). The results are normalized to the 23S rRNA gene, of which the expression was comparable in the two strains.

Acknowledgements

We acknowledge Carine Wartel and Merel Sanctorum for their technical assistance and we are grateful to Phu Nguyen Le Minh for performing the primer extension analyses. This work was supported by the Research Foundation Flanders (FWO-Vlaanderen, grant G.0015.04 and a grant for a stay abroad by E.P. in the laboratory of S.V.A. and A.D.), the Research Council of the Vrije Universiteit Brussel (OZR) and the Vlaamse Gemeenschapscommissie. E.P. is a postdoctoral fellow of FWO-Vlaanderen. S.V.A. was supported by a VIDI grant by the Dutch Science Organization (NWO).

References

- Albers, S.-V., and Driessen, A.J.M. (2007) Conditions for gene disruption by homologous recombination of exogenous DNA into the *Sulfolobus solfataricus* genome. *Archaea* **2**: 145–149.
- Aravind, L., and Koonin, E.V. (1999) DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res* **27**: 4658–4670.
- d'Ari, R., Lin, R.T., and Newman, E.B. (1993) The leucine-responsive regulatory protein: more than a regulator? *Trends Biochem Sci* **18**: 260–263.
- Bell, S.D., and Jackson, S.P. (2001) Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* **4**: 208–213.
- Beloin, C., Jussset, J., Révet, B., Mirambaut, G., Le Hégarat, F., and Le Cam, E. (2003) Contribution of DNA conformation and topology in right-handed DNA wrapping by the *Bacillus subtilis* LrpC protein. *J Biol Chem* **278**: 5333–5342.
- Birnboim, H.C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**: 1513–1523.
- Brinkman, A.B., Bell, S.D., Lebbink, R.J., de Vos, W.M., and van der Oost, J. (2002) The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J Biol Chem* **277**: 29537–29549.
- Brinkman, A.B., Ettema, T.J.G., de Vos, W.M., and van der Oost, J. (2003) The Lrp family of transcriptional regulators. *Mol Microbiol* **48**: 287–294.
- Brock, T.D., Brock, K.M., Belly, R.T., and Weiss, R.L. (1972)

- Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch Mikrobiol* **84**: 54–68.
- Brügger, K., Redder, P., She, Q., Confalonieri, F., Zivanovic, Y., and Garrett, R.A. (2002) Mobile elements in archaeal genomes. *FEMS Microbiol Lett* **206**: 131–141.
- Brunelle, A., and Schleif, R.F. (1987) Missing contact probing of DNA–protein interactions. *Proc Natl Acad Sci USA* **84**: 6673–6676.
- Enoru-Eta, J., Gigot, D., Glansdorff, N., and Charlier, D. (2002) High resolution contact probing of the Lrp-like DNA-binding protein Ss-Lrp from the hyperthermoacidophilic crenarchaeote *Sulfolobus solfataricus* P2. *Mol Microbiol* **45**: 1541–1555.
- Enoru-Eta, J., Gigot, D., Thia-Toong, T.-L., Glansdorff, N., and Charlier, D. (2000) Purification and characterization of Sa-Lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon *Sulfolobus acidocaldarius* homologous to the bacterial global transcriptional regulator Lrp. *J Bacteriol* **182**: 3661–3672.
- Ettema, T.J.G., Brinkman, A.B., Tani, T.H., Rafferty, J.B., and van der Oost, J. (2002) A novel ligand-binding domain involved in regulation of amino acid metabolism in prokaryotes. *J Biol Chem* **277**: 37464–37468.
- Fukuda, E., and Wakagi, T. (2002) Substrate recognition by 2-oxoacid: ferredoxin oxidoreductase from *Sulfolobus* sp. strain 7. *Biochim Biophys Acta* **1597**: 74–80.
- Geiduschek, E.P., and Ouhammouch, M. (2005) Archaeal transcription and its regulators. *Mol Microbiol* **56**: 1397–1407.
- Glansdorff, N. (1965) Topography of cotransducible arginine mutations in *Escherichia coli* K12. *Genetics* **51**: 167–179.
- Grogan, D.W. (1989) Phenotypic characterization of the archaeobacterial genus *Sulfolobus*: comparison of five wild-type strains. *J Bacteriol* **171**: 6710–6719.
- Grünberg, S., Bartlett, M.S., Naji, S., and Thomm, M. (2007) Transcription factor E is a part of transcription elongation complexes. *J Biol Chem* **282**: 35482–35490.
- Habicht, K.S., Salling, L., Thamdrup, B., and Canfield, D.E. (2005) Effect of low sulfate concentrations on lactate oxidation and isotope fractionation during sulfate reduction by *Archaeoglobus fulgidus* strain. *Z Appl Environ Microbiol* **71**: 3770–3777.
- Higuchi, R., Krummel, B., and Saiki, R.K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res* **16**: 7351–7367.
- Hung, S., Baldi, P., and Hatfield, G.W. (2002) Global gene expression profiling in *Escherichia coli* K12. *J Biol Chem* **277**: 40309–40323.
- Jafri, S., Evoy, S., Cho, K., Craighead, H.G., and Winans, S.C. (1999) An Lrp-type transcriptional regulator from *Agrobacterium tumefaciens* condenses more than 100 nucleotides of DNA into globular nucleoprotein complexes. *J Mol Biol* **288**: 811–824.
- Kawashima, T., Aramaki, H., Oyamada, T., Makino, K., Yamada, M., Okamura, H., *et al.* (2008) Transcription regulation by feast/famine regulatory proteins, FFRPs, in archaea and eubacteria. *Biol Pharm Bull* **31**: 173–186.
- Keuntje, B., Masepohl, B., and Klipp, W. (1995) Expression of the *putA* gene encoding proline dehydrogenase from *Rhodobacter capsulatus* is independent of NtrC regulation but requires an Lrp-like activator protein. *J Bacteriol* **177**: 6432–6439.
- Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., *et al.* (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**: 364–370.
- Koike, H., Ishijima, S.A., Clowney, L., and Suzuki, M. (2004) The archaeal feast/famine regulatory protein: potential roles of its assembly forms for regulating transcription. *Proc Natl Acad Sci USA* **101**: 2840–2845.
- Kölling, R., and Lothar, H. (1985) AsnC: an autogenously regulated activator of asparagine synthetase A transcription in *Escherichia coli*. *J Bacteriol* **164**: 310–315.
- Kumarevel, T., Nakano, N., Ponnuraj, K., Gopinath, S.C.B., Sakamoto, K., Shinkai, A., *et al.* (2008) Crystal structure of glutamine receptor protein from *Sulfolobus tokodaii* strain 7 in complex with its effector 1-glutamine: implications of effector binding in molecular association and DNA binding. *Nucleic Acids Res* **36**: 4808–4820.
- Kyrpides, N.C., and Ouzounis, C.A. (1999) Transcription in Archaea. *Proc Natl Acad Sci USA* **96**: 8545–8550.
- Labes, A., and Schönheit, P. (2001) Sugar utilization in the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus* strain 7324: starch degradation to acetate and CO₂ via a modified Embden-Meyerhof pathway and acetyl-CoA synthetase (ADP-forming). *Arch Microbiol* **176**: 329–338.
- Leonard, P.M., Smits, S.H.J., Sedelnikova, S.E., Brinkman, A.B., de Vos, W.M., van der Oost, J., *et al.* (2001) Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. *EMBO J* **20**: 990–997.
- Littlefield, O., Korkhin, Y., and Sigler, P.B. (1999) The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci USA* **96**: 13668–13673.
- Madhusudhan, K.T., Huang, N., and Sokatch, J.R. (1995) Characterization of BkdR-DNA binding in the expression of the *bkd* operon of *Pseudomonas putida*. *J Bacteriol* **177**: 636–641.
- Madhusudhan, K.T., Lorenz, D., and Sokatch, J.R. (1993) The *bkdR* gene of *Pseudomonas putida* is required for expression of the *bkd* operon and encodes a protein related to Lrp of *Escherichia coli*. *J Bacteriol* **175**: 3934–3940.
- Maxam, A.M., and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* **65**: 499–560.
- Napoli, A., van der Oost, J., Sensen, C., Charlebois, R., Rossi, M., and Ciaramella, M. (1999) An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds its own promoter. *J Bacteriol* **181**: 1474–1480.
- Napoli, A., Kvaratskeli, M., White, M.F., Rossi, M., and Ciaramella, M. (2001) A novel member of the Bacterial-Archaeal regulator family is a nonspecific DNA-binding protein and induces positive supercoiling. *J Biol Chem* **276**: 10745–10752.
- van Oeffelen, L., Cornelis, P., Van Delm, W., De Ridder, F., De Moor, B., and Moreau, Y. (2008) Detecting cis-regulatory binding sites for cooperatively binding proteins. *Nucleic Acids Res* **36**: e46.
- Okamura, H., Yokoyama, K., Koike, H., Yamada, M., Shimowasa, A., Kabasawa, M., *et al.* (2007) A structural code for

- discriminating between transcription signals revealed by the feast/famine regulatory protein DM1 in complex with ligands. *Structure* **15**: 1325–1338.
- Ouhammouch, M. (2004) Transcriptional regulation in Archaea. *Curr Opin Genet Dev* **14**: 133–138.
- Ouhammouch, M., and Geiduschek, E.P. (2005) An expanding family of archaeal transcriptional activators. *Proc Natl Acad Sci USA* **102**: 15423–15428.
- Ouhammouch, M., Dewhurst, R.E., Hausner, W., Thomm, M., and Geiduschek, E.P. (2003) Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proc Natl Acad Sci USA* **100**: 5097–5102.
- Peeters, E., Thia-Toong, T.L., Gigot, D., Maes, D., and Charlier, D. (2004) Ss-LrpB, a novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. *Mol Microbiol* **54**: 321–336.
- Peeters, E., Willaert, R., Maes, D., and Charlier, D. (2006) Ss-LrpB from *Sulfolobus solfataricus* condenses about 100 base pairs of its own operator DNA into globular nucleoprotein complexes. *J Biol Chem* **281**: 11721–11728.
- Peeters, E., Wartel, C., Maes, D., and Charlier, D. (2007) Analysis of the DNA-binding sequence specificity of the archaeal transcriptional regulator Ss-LrpB from *Sulfolobus solfataricus* by systematic mutagenesis and high resolution contact probing. *Nucleic Acids Res* **35**: 623–633.
- Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**: e45.
- Redder, P., and Garrett, R.A. (2006) Mutations and rearrangements in the genome of *Sulfolobus solfataricus* P2. *J Bacteriol* **188**: 4198–4206.
- Reddy, M.C., Gokulan, K., Jacobs, W.R., Jr, Ioerger, T.R., and Sacchettini, J.C. (2008) Crystal structure of *Mycobacterium tuberculosis* LrpA, a leucine-responsive global regulator associated with starvation response. *Protein Sci* **17**: 159–170.
- Ren, J., Sainsbury, S., Combs, S.E., Capper, R.G., Jordan, P.W., Berrow, N.S., et al. (2007) The structure and transcriptional analysis of a global regulator from *Neisseria meningitidis*. *J Biol Chem* **282**: 14655–14664.
- de los Rios, S., and Perona, J.J. (2007) Structure of the *Escherichia coli* leucine-responsive regulatory protein Lrp reveals a novel octameric assembly. *J Mol Biol* **366**: 1589–1602.
- Schelert, J., Dixit, V., Hoang, V., Simbahan, J., Drozda, M., and Blum, P. (2004) Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. *J Bacteriol* **186**: 427–437.
- Schut, G.J., Menon, A.L., and Adams, M.W. (2001) 2-keto acid oxidoreductases from *Pyrococcus furiosus* and *Thermococcus litoralis*. *Methods Enzymol* **331**: 144–158.
- She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., et al. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci USA* **98**: 7835–7840.
- Shrivastava, T., and Ramachandran, R. (2007) Mechanistic insights from the crystal structures of a feast/famine regulatory protein from *Mycobacterium tuberculosis* H37Rv. *Nucleic Acids Res* **35**: 7324–7335.
- Snijders, A.P., Walther, J., Peter, S., Kinnman, I., de Vos, M.G., van de Werken, H.J., et al. (2006) Reconstruction of central carbon metabolism in *Sulfolobus solfataricus* using a two-dimensional gel electrophoresis map, stable isotope labelling and DNA microarray analysis. *Proteomics* **6**: 1518–1529.
- Szabo, Z., Sani, M., Groeneveld, M., Zolghadr, B., Schelert, J., Albers, S.-V., et al. (2007) Flagellar motility and structure in the hyperthermoacidophilic archaeon *Sulfolobus solfataricus*. *J Bacteriol* **189**: 4305–4309.
- Tani, T.H., Khodursky, A., Blumenthal, R.M., Brown, P.O., and Matthews, R.G. (2002) Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proc Natl Acad Sci USA* **99**: 13471–13476.
- Tapias, A., Lopez, G., and Ayora, S. (2000) *Bacillus subtilis* LrpC is a sequence-independent DNA-binding and DNA-bending protein which bridges DNA. *Nucleic Acids Res* **28**: 552–559.
- Thaw, P., Sedelnikova, S.E., Muranova, T., Wiese, S., Ayora, S., Alonso, J.C., et al. (2006) Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic Acids Res* **34**: 1439–1449.
- Torarinsson, E., Klenk, H.-P., and Garrett, R.A. (2005) Divergent transcriptional and translational signals in Archaea. *Environ Microbiol* **7**: 47–54.
- Wang, H., Glansdorff, N., and Charlier, D. (1998) The arginine repressor of *Escherichia coli* K-12 makes direct contacts to minor and major groove determinants of the operators. *J Mol Biol* **277**: 805–824.
- Yellaboina, S., Seshadri, J., Senthil Kumar, M., and Ranjan, A. (2004) PredictRegulon: a web server for the prediction of the regulatory protein binding sites and operons in prokaryote genomes. *Nucleic Acids Res* **32**: W318–W320.
- Yokoyama, K., Ishijima, S.A., Clowney, L., Koike, H., Aramaki, H., Tanaka, C., et al. (2006) Feast/famine regulatory proteins (FFRPs): *Escherichia coli* Lrp, AsnC and related archaeal transcription factors. *FEMS Microbiol Rev* **30**: 89–106.
- Yokoyama, K., Ishijima, S.A., Koike, H., Kurihara, C., Shimowasa, A., Kabasawa, M., et al. (2007) Feast/famine regulation by transcription factor FL11 for the survival of the hyperthermophilic archaeon *Pyrococcus OT3*. *Structure* **15**: 1542–1554.
- Zhang, Q., Iwasaki, T., Wakagi, T., and Oshima, T. (1996) 2-oxoacid: ferredoxin oxidoreductase from the thermoacidophilic archaeon, *Sulfolobus* sp. strain 7. *J Biochem* **120**: 587–599.
- Zolghadr, B., Weber, S., Szabo, Z., Driessen, A.J., and Albers, S.-V. (2007) Identification of a system required for the functional surface localization of sugar binding proteins with class III signal peptides in *Sulfolobus solfataricus*. *Mol Microbiol* **64**: 795–806.

Supporting information

Additional supporting information may be found in the online version of this article.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Figure S1. Analysis of the genomic context of *Ss-lrpB*: primer extension analyses, *porDAB* operon structure analysis and analysis of the presence of IS elements.

A. Schematic representation of the genomic environment of *Ss-lrpB*. Bp length is indicated on top and ORFs are represented by large open arrows. From left to right Sso2125, Sso2126, Sso2127, Sso2128, Sso2129, Sso2130, Sso2131 and Sso2133 are depicted with indication of the promoter regions (arrows). The hybridization positions of primers used in the *porDAB* operon analysis and in the analysis of presence/absence of IS1078 are also shown. An asterisk indicates the position of IS1078 in the different strains. **B.** Primer extension analysis of the transcripts showing the transcription start sites. The main primer extension products are indicated with an arrow. A+G and C+T represent the corresponding Maxam-Gilbert sequencing ladders. A systematic correction in the alignment of the cDNA product with the sequencing ladders has been performed to take into account the difference in migration velocity of the cDNA and the reference ladders due to different ends generated by the AMV reverse transcriptase and the chemical modification and cleavage reactions. **C.** *porDAB* operon analysis with RT-PCR. Lanes 1 and 2. PCRs with primers DC569(f)/DC570(r); lanes 3 and 4. PCRs with primers DC571(f)/DC570(r); lanes 5 and 6. PCRs with primers DC569(f)/DC572(r). Samples in lanes 2, 4 and 6 have been treated identical as the samples in the respective preceding lanes, but without addition of RT. RT-PCRs with DC569(f)/DC570(r), DC571(f)/DC572(r) and DC569(f)/DC572(r) are expected to form 257 bp-, 237 bp- and 1601 bp-fragments, respectively. **D.** PCRs with gDNA to analyze IS rearrangements. Lane 1. PCR with P2 gDNA and primers DC449(f)/DC448(r). Lane 2. PCR with PBL2025 gDNA and primers DC449(f)/DC448(r). Lane 3. PCR with P2 gDNA and primers DC591(f)/DC592(r). Lane 4. PCR with PBL2025 gDNA and primers DC591(f)/DC592(r).

Primer extension analysis allowed the mapping of the transcription start sites of the three potential *Ss-LrpB* target genes (Figure S1B). The positions of the transcription starts are indicated in Figure 2 of the main text. For Sso2127 and *porD*, a single transcript was detected. The transcripts are leaderless. Transcription is initiated at the A residue of the ATG start codon of *porD* and at the residue immediately preceding the start codon of Sso2127. Sso2126-specific reverse transcription resulted in two

cDNAs with a length differing by 1 nt. The strongest transcript is initiated at an A residue, which is preceded by a T residue. This pyrimidine/purine step at the initiator site is a prerequisite for efficient and correct transcription initiation in archaea (Soppa, 1999) and is also observed for the other genes analyzed here. Transcription of Sso2126 is initiated 8 nt upstream of the ATG start codon. Of this stretch only GAG is complementary to the 3' end of 16S rRNA. Therefore, this short 5' untranslated region (5'-UTR) does not seem to harbour a Shine-Dalgarno (SD) site. *S. solfataricus*, and archaea in general, contain a high fraction of leaderless transcripts (Torarinsson *et al.*, 2005; Brenneis *et al.*, 2007). *Ss-lrpB* itself also produces a leaderless transcript (Peeters *et al.*, 2004). Based on the transcription start sites, the most probable recognized TATA boxes and transcription factor B recognition elements (BREs) can be predicted in the three promoter regions. These predictions indicate moderately strong promoters. In the case of *porD*, the TATA box is quite well conserved but the BRE element is not very purine-rich, although it does contain an A residue at positions -32 and -33. This is in contrast to the Sso2126 and Sso2127 promoters, which contain purine-rich BRE elements and weak TATA boxes with conservation mainly in the downstream half sites.

The *porDAB* genes are assumed to form an operon because of their close packing and their annotations as POR δ , POR α and POR β , respectively (She *et al.*, 2001). We experimentally verified the existence of a *porDAB* polycistronic messenger by performing RT-PCR with total RNA extracted from *S. solfataricus* P2 cells (Figure S1C). The primers that were used are complementary to the ORFs of the genes and the corresponding mRNA molecule (Figure S1A). Amplification with cDNA as a template indicates the existence of a mRNA spanning all three *por* genes (Figure S1C). Equivalent samples to which no reverse transcriptase was added, were not productive, confirming the absence of genomic DNA in the RNA preparations.

The genome of *S. solfataricus* P2 contains a large number of insertion sequence (IS) elements, including 8 copies of the ISC1078 element (She *et al.*, 2001). According to the published genome sequence, one of these copies is located in between *Ss-lrpB* and Sso2133 (Figure S1A). However, analysis of the genomic DNA of strains P2 and PBL2025 used in this work demonstrated the absence of this IS element (Figure S1D). A PCR with primer pair DC449(f)/DC448(r) resulted in the amplification of a 1068 bp-fragment instead of the expected 2148 bp-fragment. Sequencing of the P2 genomic region revealed the absence of a 2 bp-direct repeat at the ISC1078 target site (CA at positions -155 and -156 of the *Ss-lrpB* promoter region). This implies that the P2 strain utilized in this study never carried an ISC1078 copy at that position, as observed previously in other laboratories (Redder and Garrett, 2006). Intriguingly, PCR and sequencing analysis with primers DC591(f) and DC592(r) has shown that a copy of ISC1078 is inserted about 6 kbp further upstream in PBL2025, but not in P2 (Figure S1A and D). This resulted in an amplicon of 1898 bp with PBL2025 genomic DNA as opposed to 825 bp with P2 genomic DNA. The IS element starts 37 nt downstream of the Sso2126 stop codon and ends before position -374 of the Sso2127 control region. It is therefore not assumed to interfere with the expression of Sso2126 or the transcriptional regulation of Sso2127.

Brenneis, M., Hering, O., Lange, C., and Soppa, J. (2007) Experimental characterization of Cis-acting elements important for translation and transcription in halophilic archaea. *PloS Genet* **3**: e229.

Peeters, E., Thia-Toong, T.L., Gigot, D., Maes, D., and Charlier, D. (2004) Ss-LrpB, a novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. *Mol Microbiol* **54**: 321-336.

Redder, P., and Garrett, R.A. (2006) Mutations and rearrangements in the genome of *Sulfolobus solfataricus* P2. *J Bacteriol* **188**: 4198-4206.

She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., *et al.* (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci USA* **98**: 7835-7840.

Soppa, J. (1999) Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol* **31**: 1295-1305.

Torarinsson, E., Klenk, H.-P., and Garrett, R.A. (2005) Divergent transcriptional and translational signals in Archaea. *Environ Microbiol* **7**: 47-54.

Table S1. Real-time qPCR data obtained from a series of triplicate experiments for relative gene expression analysis of Sso2126, Sso2127 and Sso2128 (*porD*).

gene	CP_{WT}	σ	CP_{KO}	σ	ratio (WT/KO)
Sso2126	20.9	0.62	21.4	0.35	0.70
Sso2127	19.7	0.40	21.3	0.21	0.32
Sso2128	18.2	0.15	20.9	0.36	0.17

Table S2. List of sequences of oligonucleotides used in this work.

DC104(f)	5'-AACTGCAGTTAAACTTTCTTCCCCATATCTTGCAC-3'
DC178	5'-CTGGAAGATACCTCTATACACTATC-3'
DC283(f)	5'-CTAGAATCTAATAAA TTGCAAAAACCTTCTATACGTAAGTTAACATAGTTACAAT-3'
DC284(r)	5'-CTAGATTTGTAACTAT GTTAACCTTACGTATAGAAGTTTTTGCAATTTATTAGATT-3'
DC285(f)	5'-CTAGACTTGCACATTT TGCCAAAGATTCCAGAAACGATCTAATAAATTGCAAAT-3'
DC286(r)	5'-CTAGATTTGCAATTTA TTAGATCGTTTCTGGAATCTTTGGCAAAAATGTGCAAGT-3'
DC348(f)	5'-GACGGCTAATTAGCAGAGATACACG-3'
DC349(r)	5'-GTTAACCCTACATTACCAGTTGGG-3'
DC350(f)	5'-CTGATCCACTGAGCTATAAACATTA-3'
DC351(r)	5'-GCACCTACTACTAGCATCCACCTAAACC-3'
DC384(f)	5'-CTAGATTTTAAGAATTTTCAATT-3'
DC385(r)	5'-CTAGAATTGAAAATTCTTAAAAT-3'
DC386(f)	5'-CTAGACTTGCATTTTATGCAAGT-3'
DC387(r)	5'-CTAGACTTGCATAAAAATGCAAGT-3'
DC445(f)	5'-CGGGGTACCGAGGAAAGGAAAAATAAAGCAAGAACC-3'
DC446(r)	5'-CATGCCATGGCTTTAAAGATGATGCAAGCGCACATC-3'
DC448(r)	5'-AAGGAAAAAA GCGGCCGCGCCAAACAATTGCATTGTAAATTGG-3'
DC449(f)	5'-CGCGGATCCCCCATCATTATACATTATACATTATC-3'
DC497(f)	5'-CATATGTACATGTAATCTAGGCCATTATCTAAACCG-3'
DC498(r)	5'-GTAAATATAATAAGACTTGATGATACTGATGAG-3'
DC518(f)	5'-CTTCCCCATATCTTTCACATTTTGCCAAAG-3'
DC519(r)	5'-CTTTGGCAAAAATGTGAAAGATATGGGGAAG-3'
DC520(f)	5'-GCCAAAGATTACAGAAACGATCTAATAAATTGC-3'
DC521(r)	5'-GCAATTTATTAGATCGTTTCTGTAATCTTTGGC-3'
DC522(f)	5'-GATTCCAGAAACGATCGAATAAATTGCAAAAAC-3'
DC523(r)	5'-GTTTTTGCATTTATTCGATCGTTTCTGGAATC-3'
DC524(f)	5'-GCAAAAACCTTATATACGTAAGTTAACATAG-3'
DC525(r)	5'-CTATGTAACTTACGTATATAAGTTTTTGC-3'
DC533(f)	5'-TAAAAAGGCATTA TCTTAACATAGTTACAATAATCCTTTTATGTT-3'
DC534(r)	5'-AACATAAAAGGAT TATTGTAACCTATGTTAAGATAATGCCTTTTTTA-3'
DC535(f)	5'-TAAAAAGGCATTA TCTCTAATAAATTGCAATAATCCTTTTATGTT-3'
DC536(r)	5'-AACATAAAAGGAT TATTGCAATTTATTAGAGATAATGCCTTTTTTA-3'
DC537(f)	5'-TAAAAAGGCATTA TCTTGCAAAAACCTTCTATAATCCTTTTATGTT-3'
DC538(r)	5'-AACATAAAAGGAT TATAGAAGTTTTTGCAAGATAATGCCTTTTTTA-3'
DC539(f)	5'-TAAAAAGGCATTA TCTTGCAATTTTGCCATAATCCTTTTATGTT-3'
DC540(r)	5'-AACATAAAAGGAT

	TATGGCAAAATGTGCAAGATAATGCCTTTTTTA-3'
DC541(f)	5'-TAAAAAGGCATTA
	TCTTGCCAAAGATTCCATAATCCTTTTATGTT-3'
DC542(r)	5'-AACATAAAAGGAT
	TATGGAATCTTTGGCAAGATAATGCCTTTTTTA-3'
DC559(f)	5'-TACTGGACCAAGTGGTGGCTAAG-3'
DC560(r)	5'-CCCTACGGGAAAGTTCAAAG-3'
DC569(f)	5'-GCTGAAGTGTGCCCAGTTAAGG-3'
DC570(r)	5'-TACATCGGCGTACATTGCTGCATGAG-3'
DC571(f)	5'-ACTCCCATAACGAAACCCGTTTATTG-3'
DC572(r)	5'-TATTGTTCTTGGTCCAAGCAGCTTTAG-3'
DC573(f)	5'-AGCGGAATGGCTTACTCATTAGCTTG-3'
DC574(r)	5'-GAATTTAACGCTGGTGGTAATCCTACG-3'
DC575(f)	5'-TGCTGGCGTTCTTCTTTGCTTATC-3'
DC576(r)	5'-TTACCCACCATTGCCATGTCTTTAC-3'
DC577(f)	5'-AGAGGGTACAGTAATTCCAGCAGTTC-3'
DC578(r)	5'-TTCAGCGCATATTCCGCATCCTACAC-3'
DC579(f)	5'-ACGTGCAATTCGTTCCCCTGAC-3'
DC580(r)	5'-ACGCCCTTCAACTGGCACTTAG-3'
DC591(f)	5'-CGGGGTACCTTCAGCGGAATGGCTTAC-3'
DC592(r)	5'-CATGCCATGGATCATTCATTCTATACG-3'
DC703(f)	5'-TAAAAAGGCATTA
	TCTTAACATTTTTTACAATAATCCTTTTATGTT-3'
DC704(r)	5'-AACATAAAAGGAT
	TATTGTAAAAATGTTAAGATAATGCCTTTTTTA-3'
DC707(r)	5'-CTTATAAGGGCGATTCATGGATCC-3'